

Very short-term probabilistic forecasting of wind power with generalised logit-Normal distributions

P. Pinson

Technical University of Denmark, Dpt. of Informatics and Mathematical Modelling 2800 Kgs. Lyngby, Denmark

Abstract. Very short-term probabilistic forecasts, which are essential for an optimal management of wind generation, ought to account for the nonlinear and double-bounded nature of that stochastic process. They take here the form of discrete-continuous mixtures of generalised logit-Normal distributions and probability masses at the bounds. Both autoregressive and conditional parametric autoregressive models are considered for the dynamics of their location and scale parameters. Estimation is performed in a recursive least-squares framework with exponential forgetting. The superiority of this proposal over classical assumptions about the shape of predictive densities e.g. Normal and Beta, is demonstrated based on 10-minute ahead point and probabilistic forecasting at the Horns Rev wind farm in Denmark.

Keywords: bounded time-series; dynamic models; probabilistic forecasting; transformation; wind power

1. Introduction

In different areas of forecasting, both theoretical and practical developments are pointing towards various forms of probabilistic forecasting. This has been the case in economics and finance (Tay and Wallis 2000, Timmermann 2000), earthquake prediction (Kagan and Jackson 2000), meteorology (Leutbecher and Palmer 2008), as well as for weather-related processes like renewable energy production (Bacher *et al.* 2009, Pinson *et al.* 2007b) or floods (Clocke and Pappenberger 2009). When considering continuous variables such as 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 power random variable for a set of lead times. The optimal management and trading of wind power calls for probabilistic forecasts (Pinson *et al.* 2007). This follows from a more general result: for a large class of decision-making problems, optimal decisions directly relate to quantiles of conditional predictive distributions (Gneiting 2011).

Lead times in the range of hours to days correspond to the needs of decision-makers in view of the structure and operation of European electricity markets (Giebel *et al.* 2011, Costa *et al.* 2008). They have thus attracted the most attention so far. Density forecasts for these horizons are commonly based on statistical models for the “dressing” of point predictions, conditional to weather conditions (Bremnes 2006, Møller *et al.* 2008, Pinson and Kariniotakis 2010). Alternatively, they result from the dynamic post-processing of ensemble forecasts of meteorological variables (Nielsen *et al.* 2006, Pinson and Madsen 2009, Taylor *et al.* 2009). When wind power penetration reaches a certain level, and with large-scale wind farms injecting power at a single point of the electricity network, it becomes crucial to also have forecasts for horizons of a few minutes ahead. Characterising and modelling the power

fluctuations at these time scales is recognised as a current challenge. Firstly, this is required by the Transmission System Operator (TSO) in order to optimally operate reserves for the continuous balance of the power system. In that case, the relevant horizons are defined as 1, 10 and 30 minutes (Sørensen *et al.* 2007). More specifically for the case of Denmark, the TSO Energinet.dk has defined the 10-minute lead time as the most important one since power fluctuations at this time scale are those that most seriously affect the balance in the power system (Akhmatov 2007). This aspect directly motivates our choice to focus on 10-minute ahead forecasts only. Secondly, very short-term predictions are needed as input to the (offshore) wind farm controllers themselves with typical time steps between 5 and 15 minutes (Kristoffersen and Christiansen 2003, Sørensen *et al.* 2005). Finally, the operation of wind-storage systems at temporal scales consistent with the temporal resolution at which regulation is carried out in practise in some electricity markets e.g. 15 minutes in the Netherlands, or 5 minutes in New Zealand and Australia, calls for very short-term forecasts. Even though emphasis is placed on the 10-minute lead time only, extension to further horizons can be the topic of future work. This may be based on the same predictive distributions but with different input data and dynamics models.

For such short lead times, statistical models based on historical measurements only, though taking advantage of meteorological and physical expertise on the problem at hand, should be preferred. Running Numerical Weather Prediction (NWP) models for these temporal resolutions with frequent updates would be clearly too expensive if not impossible today. Also statistical approaches are known to outperform NWP-based ones for the forecasting of the power output of a wind farm at lead times less than a few hours (Giebel *et al.* 2011). Examples of recent works on improving such statistical approaches, focusing on the wind (speed and potentially direction) variable, include the regime-switching space-time method introduced by Gneiting *et al.* (2006), and the multivariate wind vector models of Hering and Genton (2009). Some other works have been focusing on capturing changes in the wind field dynamics using e.g. Markov-switching approaches (Ailliot *et al.* 2006). In parallel, Pinson and Madsen (2011) discuss the benefits of Markov-switching autoregressive models for point and probabilistic forecasting of wind power generation at forecast horizons of 10-minutes ahead.

We place ourselves in a parametric probabilistic forecasting framework, for which very few proposals exist in the literature, especially at these lead times. Our core objective is to discuss how wind power characteristics motivate the choice for specific types of predictive distributions. A first crucial feature of the wind power variable is that it is double-bounded between a minimum production of zero and a maximum one being equal to the nominal capacity (denoted P_n) of the wind turbine, wind farm, or wind energy portfolio considered. It is in parallel a nonlinear function of wind speed in the form of a sigmoid. Wind power measurements and forecasts are normalised by P_n in the following. They then take values in the unit interval $[0, 1]$. The time variations of P_n , if any, are commonly recorded for recent wind farms. They can be used for an adaptive scaling of measurements and forecasts. A consequence of the above is that predictive densities cannot be Gaussian. This is valid whatever the temporal resolution of wind power time-series or the forecast horizon considered. Their higher-order moments are also directly related to their mean, and potentially to some external signal. These two points have already been discussed in a number of publications (Lange 2005, Pinson 2006, Bludszweit *et al.* 2008). To our knowledge however, no work exists in the literature where these aspects have been appropriately taken into consideration in a probabilistic forecasting framework. 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.

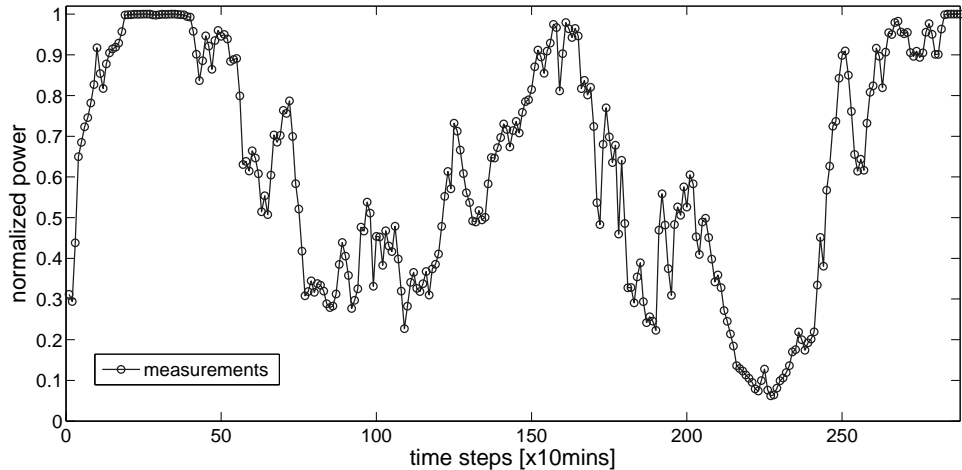


Figure 1. Episode of 2 days (288 time-steps, with a temporal resolution of 10 minutes) with wind power measurements at the Horns Rev wind farm in Denmark.

We propose here a methodology for the probabilistic forecasting of wind power time-series that account for these characteristics. Predictive densities take the form of discrete-continuous mixtures consisting of a generalised logit-Normal (GL-Normal) distribution with potential concentration of probability mass at the bounds of the unit interval $[0, 1]$. Given a shape parameter, these predictive densities are fully characterised by a location and a scale parameter only. By employing the related generalised logit transform for variance stabilisation, the obtained variable can be modelled with censored Gaussian distributions. After preliminary definitions in Section 2, the forecasting methodology is described in Section 3. Both model building and estimation aspects are subsequently dealt with in Section 4, 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 only and could be further improved in the future. In the case of further lead times, say up to a few days ahead, it should be envisaged to build models additionally accounting for the weather dynamics as given by meteorological forecasts. The methodology is applied in Section 5 for the probabilistic forecasting of wind power generation at the Horns Rev wind farm in Denmark, at lead times of 10 minutes. It is evaluated against some simple benchmarks, as well as against other assumptions that could be made regarding the shape of predictive densities e.g. Normal or Beta. The paper ends with concluding remarks in Section 6.

2. Preliminary definitions

Transformations are commonly employed for variance stabilisation, and to work in a Gaussian framework. Besides the general class of power or Box-Cox transformations (Box and Cox 1964), Johnson (1949) originally introduced the so-called logistic (or logit) transform,

subsequently used for e.g. bounded economic variables (Wallis 1987). It directly relates to using logit-Normal distributions for modelling the original effect (Aitchison and Shen 1980). A generalisation of both the logit transform and of logit-Normal distributions is given below, being appropriate for the case of wind power modelling and forecasting applications.

Wind power generation is a double-bounded process, thus making the logistic transformation appealing. Owing to the different inflection in power curves for low and high power values, a suitable transformation should allow for asymmetry. The generalised logit (GL) transformation inspired by Mead (1965) hence comprises an ideal candidate. For an original time-series $\{x_t\}$, the generalised logit transform y_t is given by

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

while the inverse transformation, referred to as inverse generalised logit (IGL), is defined as

$$x_t = \gamma^{-1}(y_t; \nu) = \left\{ 1 + \frac{1}{\exp(y_t)} \right\}^{-1/\nu}, \quad \nu > 0, \quad y_t \in \mathbb{R} \quad (2)$$

The GL transformation generalises the logistic one with an additional shape parameter ν only. Its purpose is to influence the evolution of variance and skewness of these distributions as a function of their mean. For $\nu = 1$, one retrieves the more classical logit transformation, recently used by Lau and McSharry (2010) for wind power prediction. The aim of employing such transformation is to work with a time-series $\{y_t\}$ for which the assumption that conditional densities may be Gaussian is more appropriate. In parallel, in view of the known relationship between conditional expectation and variance of wind power densities (Pinson 2006, Bludszweit *et al.* 2008), it is expected that the conditional variance of the predictive densities for the transformed variable will be independent of their mean.

We introduce the generalised logit-Normal distribution (GL-Normal) based on the aforementioned GL transform. It is denoted by $L_\nu(\mu, \sigma^2)$, where μ and σ are the location and scale parameters, respectively. For a Gaussian variable $Y \sim \mathcal{N}(\mu, \sigma^2)$, the transformed variable $X = \gamma(Y; \nu)$ is such that $X \sim L_\nu(\mu, \sigma^2)$. By generalising the probability density function of a univariate logit-Normal variable given by Aitchison and Shen (1980) and Frederic and Lad (2008), we obtain that for the GL-Normal variable,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \left(\frac{\nu}{x(1-x^\nu)} \right) \exp \left\{ -\frac{1}{2} \left[\frac{\gamma(x; \nu) - \mu}{\sigma} \right]^2 \right\}, \quad x \in (0, 1) \quad (3)$$

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 (1980) 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 and Lad (2008) in order to visualize the moments of GL-Normal distributions as a function of their location and scale parameters.

3. Forecasting methodology based on generalised logit-Normal distributions

Several authors e.g. Lange (2005) and Pinson (2006), have observed that the standard deviation of the distributions of forecast errors was directly linked to its conditional expectation, whatever the time of the year, horizon and forecast model. Appropriate predictive

densities of wind power generation should account for and reproduce this observed feature. Recently, some models for this relationship have been proposed (Pinson and Madsen 2009, Bludszuweit *et al.* 2008, Usaola 2009). More precisely, Bludszuweit *et al.* (2008) looked at second-order polynomial mean-variance models for obtaining the parameters of Beta distributions when analysing the uncertainty of wind power forecasts, a posteriori. The models they obtained had the general form of asymmetric logistic models. Based on a similar observation, Usaola (2009) modeled moments and cumulants of forecast error distributions as a function of the level of predicted power, as input to power flow models. Finally Pinson and Madsen (2009) introduced a logistic mean-variance model for the Gaussian kernels used in the dressing of ensemble forecasts of wind power generation. Applying an appropriate GL-transform to wind power data allows dampening the effect: the variance of the predictive densities for the transformed variable can then be assumed to be independent of its mean.

Write $\{X_t\}$ and $\{Y_t\}$ the stochastic processes whose realisations define the time-series $\{x_t\}$ and $\{y_t\}$, following the preliminary definitions of Section 2. Wind power generation is null for wind speed values below cut-in speed, while being equal to nominal capacity when wind speed is above the rated speed. The values for cut-in and rated wind speeds may well vary depending upon the turbines and wind power portfolio considered. We consequently propose to write the predictive density for the wind power generation X_{t+k} at time $t+k$ as

$$X_{t+k} \sim \omega_{t+k}^0 \delta_0 + (1 - \omega_{t+k}^0 - \omega_{t+k}^1) L_\nu(\mu_{t+k}, \sigma_{t+k}^2) + \omega_{t+k}^1 \delta_1 \quad (4)$$

with δ_0 and δ_1 Dirac delta 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

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

The main reason for the introduction of this concentration of probability mass at the bounds is that the original variable to be GL-transformed must take values in $(0, 1)$ only (as expressed in (1)). For the case of wind power, these bounds are obviously reached on a regular basis. Consider for instance issuing 10-minute ahead point forecasts of wind power generation at Horns Rev (the test case site of Section 5). Point forecasts correspond to the 1-step ahead conditional expectation of wind power production based on an AR model. The distributions of observed wind power generation conditioned upon the prediction level are depicted in Figure 2, for various power classes close to the natural generation bounds for that wind farm. The non-negligible concentration of probability mass at these bounds is visible from this Figure. In a general manner, other forms of distributions e.g. exponential ones, could be used in (4) instead of probability masses at the bounds. This would reflect that even if predicting zero (or nominal) power, there still may be some uncertainty.

Based on (4), but now considering Y_{t+k} the GL-transform of X_{t+k} , the form of its predictive density is given by

$$Y_{t+k} \sim \omega_{t+k}^0 \delta_{-\infty} + (1 - \omega_{t+k}^0 - \omega_{t+k}^1) \mathcal{N}(\mu_{t+k}, \sigma_{t+k}^2) + \omega_{t+k}^1 \delta_{+\infty} \quad (5)$$

Similarly to the coarsening approach of Lesaffre *et al.* (2007) for the modelling of outcome scores in $[0, 1]$ using classical logit-Normal distributions, we define a threshold value ϵ , being in the order of measurement precision, say $\epsilon \leq 10^{-2}$. Wind power values below or equal to ϵ (resp. above or equal to $1 - \epsilon$) are considered to be 0 (resp. 1). This

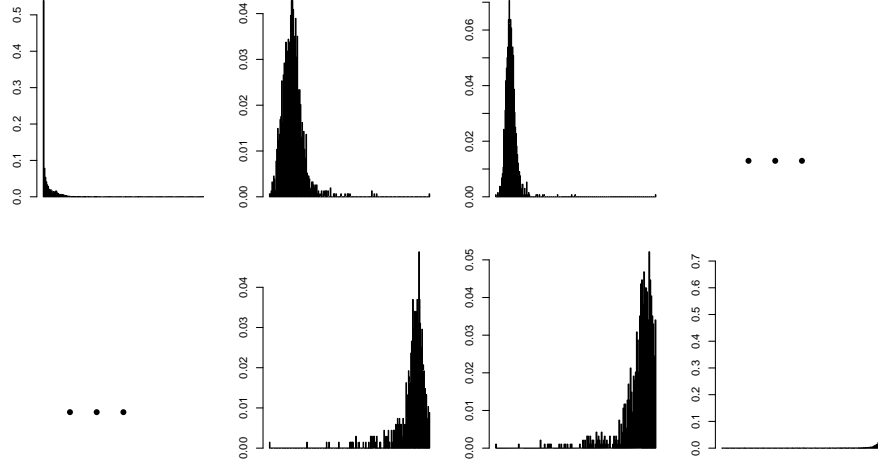


Figure 2. Distributions of observed power conditional to the predicted expected power (within $[0,2]$ % of P_n , $[2,4]$, $[4,6]$, $[94,96]$, $[96,98]$, $[98,100]$). The same scale is used for the x axis in all bar plots.

translates to constraining the range of potential variations of the GL-transformed variable, i.e. $y_t \in \mathcal{D}_y = [\gamma(\epsilon; \nu), \gamma(1 - \epsilon; \nu)]$. One can consequently rewrite (5) as

$$Y_{t+k} \sim \omega_{t+k}^0 \delta_{\gamma(\epsilon; \nu)} + \mathcal{N}(\mu_{t+k}, \sigma_{t+k}^2) \mathbf{1}_{\bar{\mathcal{D}}_y} + \omega_{t+k}^1 \delta_{\gamma(1-\epsilon; \nu)} \quad (6)$$

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

$$\mathbf{1}_{\bar{\mathcal{D}}_y} = \begin{cases} 1, & y \in \bar{\mathcal{D}}_y \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

The weight $(1 - \omega_{t+k}^0 - \omega_{t+k}^1)$ does not appear anymore in (6). This is because the predictive density for Y_{t+k} takes the form of a censored Normal instead. This permits to straightforwardly derive the expressions for the weights ω_{t+k}^0 and ω_{t+k}^1 ,

$$\omega_{t+k}^0 = \Phi\left(\frac{\gamma(\epsilon; \nu) - \mu_{t+k}}{\sigma_{t+k}}\right), \quad \omega_{t+k}^1 = 1 - \Phi\left(\frac{\gamma(1 - \epsilon; \nu) - \mu_{t+k}}{\sigma_{t+k}}\right) \quad (8)$$

with Φ the cumulative distribution function of a standard Normal variable. Predictive densities are then fully characterised by their location (μ_{t+k}) and scale (σ_{t+k}) parameters only.

4. Dynamic models for the location and scale parameters

We aim to show the advantages of modelling and forecasting wind power time-series with GL-Normal variables, not to discuss the superiority of certain dynamic models. Therefore 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. It would clearly call for modification of the estimation

methods described. GL-Normal densities could also be envisaged for further lead times (i.e. from a few hours to a few days ahead), for which appropriate statistical approaches should account for the nonlinear transformation of the dynamic information provided by meteorological predictions.

To give a brief overview of our approach, AR and CP-AR models are proposed for the location parameter. In parallel, the scale parameter is viewed as constant 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 suggested that such dynamics for the mean are appropriate when considering short lead times (Pinson and Madsen 2011). Extensive data analysis reported by Vincent *et al.* (2010) and Vincent *et al.* (2011) showed that seasonalities were not observed at the Horns Rev wind farm considered as a case study in the following. They are thus not accounted for in the various models below. Model coefficients are adaptively and recursively estimated based on an exponential forgetting of past observations. This proposal follows from various results reported in the literature (Gneiting *et al.* 2006, Pinson and Madsen 2011, Sanchez 2006, Vincent *et al.* 2010, e.g.) showing that parameters in dynamic models should be seen as slowly changing over time. This is because the dynamics of the processes involved may well vary at time scales of months, seasons and years, with features that cannot easily be accounted for in the models themselves in a parametric framework.

4.1. Autoregressive dynamics

In the case of AR dynamics, the expression relating the location parameter μ_{t+k} to previous observations writes

$$\mu_{t+k} = \boldsymbol{\theta}^\top \tilde{\mathbf{y}}_t \quad (9)$$

where $\boldsymbol{\theta} = [\theta_0 \ \theta_1 \ \dots \ \theta_l]^\top$ and $\tilde{\mathbf{y}}_t = [1 \ y_t \ \dots \ y_{t-l+1}]^\top$. This is while the scale parameter is defined as a constant, i.e. $\sigma_{t+k}^2 = \beta$.

The maximum lag l can be determined after examining the autocorrelation and partial autocorrelation functions of the $\{y_t\}$ time-series, by minimisation of either Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), or finally through a cross-validation exercise. Note that in the case of censored Normal variables, the location parameter may take values outside of the range within which y_t is constrained.

The $\{y_t\}$ time-series corresponds to realisations of censored Normal variables. It may be argued that the vector of AR parameters $\boldsymbol{\theta}$ can be estimated with either Least-Squares (LS) or Maximum-Likelihood (ML) 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 showed 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.

The parameters $\boldsymbol{\theta}$ are estimated within a Recursive Least Squares (RLS) framework (Ljung and Söderström 1983, Madsen 2007). The scale parameter σ_{t+k} can similarly be tracked accordingly. 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.* (2006) and Hering and Genton (2009), for which estimation is performed on a bulk of data at each time step.

The estimate of the parameters $\boldsymbol{\theta}_t$ at time t is defined as

$$\hat{\boldsymbol{\theta}}_t = \arg \min_{\boldsymbol{\theta}} \sum_{i=1}^t \lambda^{t-i} \left(y_i - \boldsymbol{\theta}^\top \tilde{\mathbf{y}}_{i-k} \right)^2 \quad (10)$$

where λ is the forgetting factor, $\lambda \in (0, 1)$, allowing for the 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.

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

$$\hat{\mathbf{R}}_t = \lambda \hat{\mathbf{R}}_{t-1} + \tilde{\mathbf{y}}_{t-k} \tilde{\mathbf{y}}_{t-k}^\top \quad (11)$$

$$\hat{\boldsymbol{\theta}}_t = \hat{\boldsymbol{\theta}}_{t-1} + \hat{\mathbf{R}}_t^{-1} \tilde{\mathbf{y}}_{t-k} \varepsilon_t \quad (12)$$

where ε_t is defined as $\varepsilon_t = y_t - \boldsymbol{\theta}_{t-1}^\top \tilde{\mathbf{y}}_{t-k}$. Note that there is a direct correspondence of the RLS estimator in the above with what would be the Kalman filter estimator of $\hat{\boldsymbol{\theta}}_t$ in a state-space framework. These aspects are extensively discussed by e.g. Madsen (2007) and Li (2008). Consequently, it is interesting to note that the posterior distribution of the true parameter $\boldsymbol{\theta}_t$ given the available information up to time $t - 1$ is Gaussian with mean $\hat{\boldsymbol{\theta}}_{t-1}$ and covariance $\hat{\mathbf{R}}_{t-1}^{-1}$ (Ljung and Söderström 1983, pp. 55-56).

The update (12) of the parameter estimates does not ensure that they will remain within the stationary region. This does not appear to be a problem for the 1-step ahead forecasting application. It would be more problematic if aiming at simulating the process over a period of time or at issuing multi-step ahead forecasts. With time-varying parameters, it may also well be that the parameter estimates cross into the nonstationary region and then get back into the stationary one after few time steps.

Since the scale parameter is given by a constant only, its adaptive estimation at time t can be performed thanks to an exponential smoothing scheme, being consistent with the way we update the parameters of the dynamic model for μ_{t+k} . Owing to the sensitivity of the GL transform close to the bounds of the unit interval, we have noticed that it would be beneficial to robustify the exponential smoothing of the scale parameter. This can be carried out by down-weighting observations when the location parameter gets closer to the bounds. Our proposal for this weight is given by

$$w_t = 4\boldsymbol{\theta}_t^\top \tilde{\mathbf{y}}_{t-k} \left(1 - \boldsymbol{\theta}_t^\top \tilde{\mathbf{y}}_{t-k} \right), \quad w_t \in [0, 1] \quad (13)$$

And the resulting exponential smoothing scheme writes

$$\hat{\beta}_t = \lambda_t^* \hat{\beta}_{t-1} + (1 - \lambda_t^*) \left(y_t - \boldsymbol{\theta}_t^\top \tilde{\mathbf{y}}_{t-k} \right)^2 \quad (14)$$

with $\lambda_t^* = 1 - (1 - \lambda)w_t$.

It appears appropriate to initialise $\boldsymbol{\theta}_t$ with a vector of zeros, the resulting model hence corresponding to white noise. β_t may be initialised with a small value or alternatively with an expert guess. In parallel, the inverse covariance matrix $\hat{\mathbf{R}}_t$ can be initialised as an identity matrix times a small constant, yielding a small load on its diagonal. Only the updating formula (11) is used for the first steps of the recursive estimation procedure in order to gather sufficient information on the process characteristics. When that point

is reached (say, for $t > 50$), both (12) and (14) are also employed at each time step for updating the models for the location and scale parameters. Note that $\hat{\mathbf{R}}_t$ is necessarily invertible in view of the updating performed in (11), as long as the underlying model is not over-parametrized.

4.2. Conditional parametric autoregressive dynamics

If considering CP-AR dynamics instead, the constant parameters in (9) are replaced by coefficient functions of some covariate, these functions being of unknown form,

$$\mu_{t+k} = \boldsymbol{\theta}(\omega)^\top \tilde{\mathbf{y}}_t \quad (15)$$

where $\boldsymbol{\theta}(\omega) = [\theta_0(\omega) \theta_1(\omega) \dots \theta_l(\omega)]^\top$, while $\tilde{\mathbf{y}}_t$ is defined as in (9). In parallel the model for the scale parameter becomes an unknown function of this same covariate, i.e. $\sigma_{t+k}^2 = \beta(\omega)$. The maximum lag l in the dynamic model above can be chosen similarly to that for the more simple AR dynamics. Recent developments on conditional parametric and the more general varying-coefficient models are reported in Fan and Zhang (2008).

The most obvious choice for ω is wind direction, since it directly induces shadowing effects within a wind park, thus affecting the dynamic response of that wind park to meteorological conditions. Wind direction might also influence the wind speed dynamics themselves, as is the case at Horns Rev for instance (Vincent *et al.* 2010). We denote by ω_t the wind direction measurement at time t , $\omega_t \in [0, 2\pi]$. In view of the short lead times considered in the test case application, conditioning the AR dynamics upon the wind direction measurement available at time t can be seen as reasonable. If focusing on further lead times, it would certainly be beneficial to use wind direction forecasts instead.

The coefficient functions $\boldsymbol{\theta}(\omega)$ and $\beta(\omega)$ are estimated 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.* (2000) and Pinson *et al.* (2008), to which the reader is referred to for more details. The only assumption about 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 ω .

Let us focus on a single fitting point $\omega_{(j)}$ only. The objective function to be minimised 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

$$\hat{\boldsymbol{\theta}}_{(j),t} = \arg \min_{\boldsymbol{\theta}_{(j)}} \sum_{i=1}^t \Lambda_{(j),t}(i) c_{(j),i} \left(y_i - \boldsymbol{\theta}_{(j)}^\top \tilde{\mathbf{y}}_{i-k} \right)^2 \quad (16)$$

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

$$c_{(j),i} = T \left(\frac{|\omega_i - \omega_{(j)}|}{\hat{h}_{(j)}} \right) \quad (17)$$

In the above, $|\cdot|$ denotes a polar distance, being consistent with ω being a circular variable. $\hat{h}_{(j)}$ is the bandwidth of the kernel assigned to $\omega_{(j)}$. T may for instance be chosen as a

tricube function,

$$T(v) = \begin{cases} (1 - v^3)^3, & v \in [0, 1] \\ 0, & v > 1 \end{cases} \quad (18)$$

In parallel in (16), $\Lambda_{(j),t}$ is the function that permits the exponential forgetting of past observations, i.e.

$$\Lambda_{(j),t}(i) = \begin{cases} \lambda_{(j),t}^{\text{eff}} \Lambda_{(j),t-1}(i), & 1 \leq i \leq t-1 \\ 1, & i = t \end{cases} \quad (19)$$

$\lambda_{(j),t}^{\text{eff}}$ is the effective forgetting factor for the fitting point $\omega_{(j)}$. It follows the definition given by Nielsen *et al.* (2000), which links $\lambda_{(j),t}^{\text{eff}}$ and $c_{(j),t}$ so that

$$\lambda_{(j),t}^{\text{eff}} = 1 - (1 - \lambda)c_{(j),t} \quad (20)$$

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

A little algebra, which is skipped here since extensively covered by Nielsen *et al.* (2000) and Pinson *et al.* (2008), yields the set of recursive formulae for the adaptive estimation of the local coefficients $\hat{\boldsymbol{\theta}}_{(j),t}$,

$$\varepsilon_{(j),t} = y_t - \hat{\boldsymbol{\theta}}_{(j),t-1}^\top \tilde{\mathbf{y}}_{t-k} \quad (21)$$

$$\mathbf{R}_{(j),t} = \lambda_{(j),t}^{\text{eff}} \mathbf{R}_{(j),t-1} + c_{(j),t} \tilde{\mathbf{y}}_{t-k} \tilde{\mathbf{y}}_{t-k}^\top \quad (22)$$

$$\hat{\boldsymbol{\theta}}_{(j),t} = \hat{\boldsymbol{\theta}}_{(j),t-1} + c_{(j),t} \mathbf{R}_{(j),t}^{-1} \tilde{\mathbf{y}}_{t-k} \varepsilon_{(j),t} \quad (23)$$

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 $\hat{\boldsymbol{\theta}}_{(j),t} = \hat{\boldsymbol{\theta}}_{(j),t-1}$ and $\mathbf{R}_{(j),t} = \mathbf{R}_{(j),t-1}$. This confirms the role of the effective forgetting factor, i.e. to down-weight 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 (14), it is given by

$$\hat{\beta}_{(j),t} = \lambda_{(j),t}^* \hat{\beta}_{(j),t-1} + (1 - \lambda_{(j),t}^*) \left(y_t - \boldsymbol{\theta}_{(j),t}^\top \tilde{\mathbf{y}}_{t-k} \right)^2 \quad (24)$$

with $\lambda_{(j),t}^* = 1 - (1 - \lambda)w_t c_{(j),t}$, w_t still being defined by (13).

When starting the recursive process, the local inverse covariance matrices $\mathbf{R}_{(j),t}$, as well as the local parameter estimates $\hat{\boldsymbol{\theta}}_{(j),t}$ and $\hat{\beta}_{(j),t}$, can be initialised for all fitting points as in Section 4.1 for the case of AR dynamics. The value of the parameter estimates $\hat{\boldsymbol{\theta}}_t(\omega)$ and $\hat{\beta}_t(\omega)$ for any ω can then be obtained by interpolation through the estimated local parameters at the various fitting points $\omega_{(j)}$ ($j = 1, \dots, m$).

5. 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.* (2006) for instance, and build models for direct 2-step ahead forecasting. Statistical models based only on historical measurements are in order for forecast horizons up to 6-8 hours ahead. However for horizons further than one hour ahead, dynamic information from meteorological forecasts permits to significantly improve forecast accuracy (Giebel *et al.* 2011, Costa *et al.* 2008).

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

5.1. 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, normalised by P_n , and to the average wind direction for the wind farm. An averaging procedure has been developed in order to obtain time-series of 10-minute wind direction and power averages. This averaging rate makes 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 averaging procedure has a threshold parameter τ_v , which corresponds to the minimum percentage of data needed to be considered as valid within a given time interval, so that the related power (or wind direction) average is also considered as valid. Choosing a sufficiently high value for τ_v permits to avoid sampling effects in the averaging procedure that would impact the representativeness of the average values obtained. The threshold chosen is $\tau_v = 75\%$. The available raw data is from 16th February 2005 to 25th 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.

5.2. 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. 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 of the 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_λ . It allows one to better appraise the size of the equivalent ‘sliding window’ in the adaptive estimation of the dynamic model parameters, such as that

considered by Gneiting *et al.* (2006) and Hering and Genton (2009). In parallel, the censoring parameter ϵ 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_λ and ν 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 (1974). The criterion to be minimised 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 (Gneiting *et al.* 2007). For the CP-AR dynamics, the bandwidth values are selected similarly.

Focusing on the order of the various models first, results from Pinson and Madsen (2011) 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 verified 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 initialised 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_λ and ν to be performed simultaneously or iteratively. It has been observed for the data of the present case-study that both ways yielded similar results. We thus present them below in Figure 3 after n_λ and ν were optimised one after the other, iteratively. More precisely, Figure 3(a) depicts the variations in 1-step-ahead CRPS for ν fixed to its optimal value, and with n_λ 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_λ fixed to its optimal value, and with ν 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 $n_\lambda = 2500$ (corresponding to approximately 17.5 days) and $\nu = 3.2$. This optimal value for n_λ is smaller than the 45 days deemed optimal for the sliding estimation windows of Gneiting *et al.* (2006) and Hering and Genton (2009) when considering lead times of 2 hours. The speed of change in the characteristics of wind and power dynamics certainly depends upon the temporal resolution of the time-series. It may also be related to the local wind climate: the offshore wind farm at Horns Rev is known for particular characteristics of its wind and power dynamics (Pinson and Madsen 2011, Vincent *et al.* 2011).

Figure 3(b) demonstrates the interest of introducing the generalised form of the logit transformation and hence to generalise the use of the more classical logistic transform by Lau and McSharry (2010): a significant decrease in CRPS is seen for $\nu > 1$, i.e. when considering negatively-skewed mean-variance relationships for the GL-Normal predictive densities. Our proposal probabilistic forecasting methodology seems to be fairly robust with respect to the choice of n_λ and ν 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_λ and ν , i.e. $n_\lambda = 2500$ and $\nu = 3.2$. In addition, 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, as such a 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

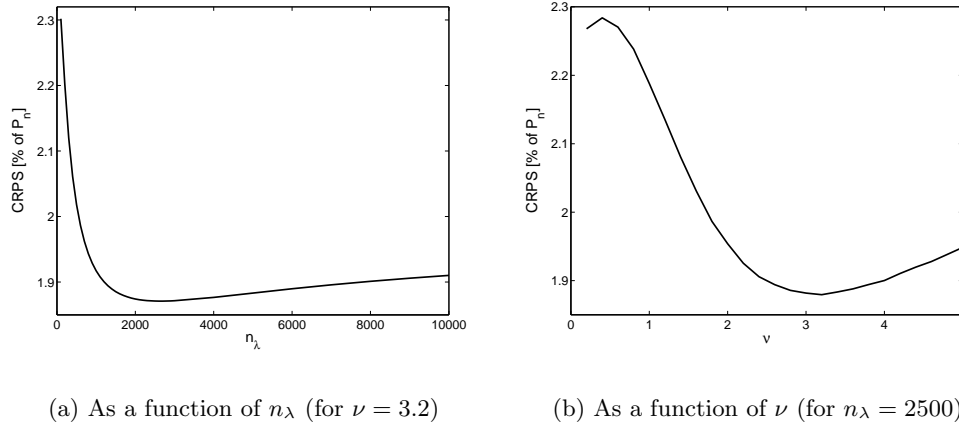


Figure 3. Evolution of the 1-step-ahead CRPS of GL-Normal density forecasts of wind power generation based on simple AR-dynamics. These results are obtained on the cross-validation set, as a function of the parameters n_λ and ν .

successive fitting points. Regarding the Kernel bandwidth in (17), both fixed and nearest-neighbour bandwidths have been envisaged. The latter concept was preferred, in order to be consistent with the idea introduced above, so it would be more consistent to respect the distribution of wind direction data when defining the fitting points. Following Nielsen *et al.* (2000), the nearest-neighbour 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 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 recognise that less fitting points and higher bandwidth values are present in areas with less data available.

As a basis for the evaluation of the forecast performance of the methodology introduced in the paper, different types of benchmark methods are considered, some of them being reviewed by Giebel *et al.* (2011), Costa *et al.* (2008) and Pinson and Madsen (2011). First of all for point forecasting the persistence forecast method, for which the forecast at time $t+k$ is given by the measurement available at time t , is known to be difficult to outperform for short lead times. Another basic benchmark is an extension of persistence based on the average of the last few measurements. To be consistent with the autoregressive orders of the models employed here, this moving average benchmark consists of the average of the last 3 measurements. In parallel the most common benchmark for 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. Since climatology is fairly easy to outperform for short lead times, we also consider the

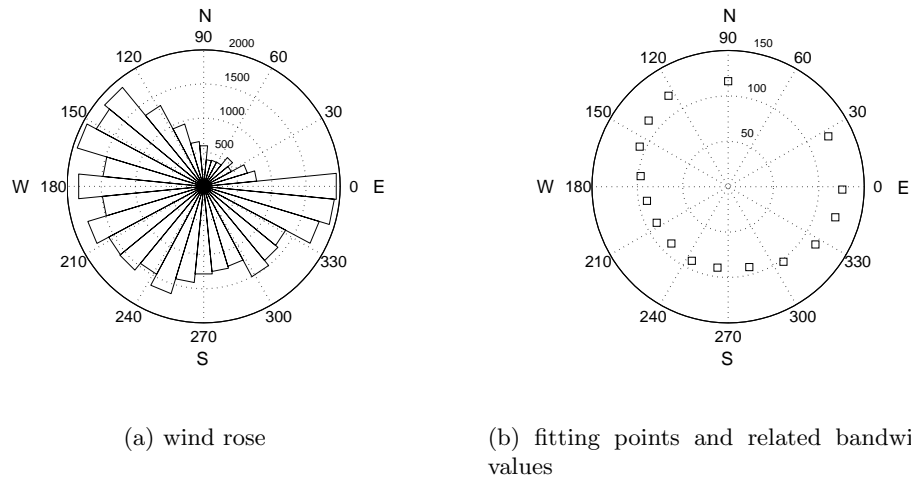


Figure 4. Wind rose, fitting points and related bandwidths values based on the distribution of wind direction values over the learning set. All values are in degrees, following the trigonometric convention.

probabilistic generalization of persistence based on a random walk model as a benchmark. At a given time this probabilistic forecast takes the form of a Gaussian predictive density with its mean given by the last available power measurement and its variance resulting from exponential smoothing of past squared residuals. For consistency, this exponential smoothing uses the same forgetting factor as for the more advanced approaches considered, i.e. with $n_\lambda = 2500$.

The main objective of the present paper is to demonstrate the interest of considering GL-Normal distributions for wind power prediction. As a basis for comparison, we therefore consider similar forecasting methodologies based on more classical assumptions about the shape of predictive densities, but with the same dynamical models for their location and shape parameters. For point forecasting the common assumption is that of a Normal distribution, while in a probabilistic forecasting framework, the most common ones are (censored) Normal or Beta distributions (Bludszuweit *et al.* 2008, Pinson and Madsen 2011). Additional point and probabilistic forecasts based on these assumptions will be seen as a benchmark. Results will be given for AR models only, and labelled as ‘Normal AR’ and ‘Beta AR’. The optimal value found for the effective number of observations n_λ is $n_\lambda = 2000$ (corresponding to approximately 14 days) for the case of both assumptions on predictive densities.

5.3. Results and comments

A fully probabilistic approach to forecast evaluation is employed. This means 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 (2011) among others, one should extract the optimal point forecasts from predictive densities based on the target evaluation

Table 1. Monthly results for the evaluation of point forecasts extracted from density forecasts: focus on the NMAE for the median (in % of the nominal capacity P_n). Best score values are highlighted using bold fonts.

Month	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	All
Moving average	3.44	3.28	3.47	3.62	3.47	3.57	3.60	3.62	3.50
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

Table 2. Monthly results for the evaluation of point forecasts extracted from density forecasts: focus on the NRMSE for the expectation (in % of the nominal capacity P_n). Best score values are highlighted using bold fonts.

Month	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	All
Moving average	5.70	6.06	5.93	6.35	5.79	8.06	6.35	5.88	6.31
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

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 normalised 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, we 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 in the wind power forecasting literature (Giebel *et al.* 2011, Costa *et al.* 2008, Madsen *et al.* 2005). Owing to the length of the evaluation set (approximately 8 months), it has been chosen to study these scores overall and on a monthly basis. The evaluation results for point forecasts based on the persistence and moving-average benchmarks, as well as for the Normal-AR, and the proposed GL-Normal AR and CP-AR predictive densities, are collated in Tables 1 and 2. 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. Persistence has a clear advantage over the other moving-average benchmark, for which all score values are much higher. Even though monthly fluctuations can be observed in the NMAE and NRMSE scores, it appears that the Normal AR benchmark almost consistently outperforms the persistence and moving-average benchmarks, 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. The recent work of Gallego *et al.* (2011), where it was shown that for short lead times wind direction influences power dynamics in a smooth manner and that this should be accounted for in statistical approaches to forecasting, confirms our results.

In the second stage, we look at the quality of predictive densities, concentrating first on

Table 3. Monthly results for the evaluation of density forecasts with a CRPS criterion (in % of the nominal capacity P_n). Best score values are highlighted using bold fonts.

Month	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	All
Climatology	18.67	19.75	20.07	18.61	19.79	24.65	21.81	21.63	20.47
Persistence	2.30	2.06	2.03	2.06	1.96	2.35	2.29	2.02	2.12
Normal AR	1.85	2.01	1.99	1.98	1.87	2.31	2.23	1.97	2.01
Beta AR	1.81	1.91	1.96	1.93	1.78	2.08	2.08	1.90	1.92
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

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.* (2007). Similar to the above, the CRPS criterion is calculated on a monthly basis as well as globally. Evaluation results are gathered in Table 3, for the two basic benchmarks (climatology and the probabilistic extension of persistence), Normal-AR and Beta-AR predictive densities, and finally for the proposed GL-Normal AR and CP-AR probabilistic forecasts.

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. The fact that Beta predictive distributions have a varying shape consistent with the double-bounded nature of wind power is reflected by the CRPS values for the Beta-AR predictive densities. The additional feature of GL-Normal distributions consisting of their natural mean-variance relationship still gives them an advantage though. 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%. The climatology benchmark is dramatically outperformed in all cases, while the probabilistic extension of persistence still appears as a competitive benchmark. Overall, non-negligible benefits may be achieved by proposing adequate models for the wind power dynamics combined with an improved assumption about the shape of predictive densities.

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.* (2006), 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 quantile composing predictive densities of wind power generation against the nominal ones (Pinson *et al.* 2007b). 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 α should cover the actual observation $\alpha\%$ of the time. Climatology is not represented since it has perfect marginal calibration.

Normal-AR predictive densities introduce a systematic and significant probabilistic bias in the probabilistic forecasts they provide. Predictive densities seem to be too wide on average, since observed proportions are larger than the nominal ones for nominal proportions

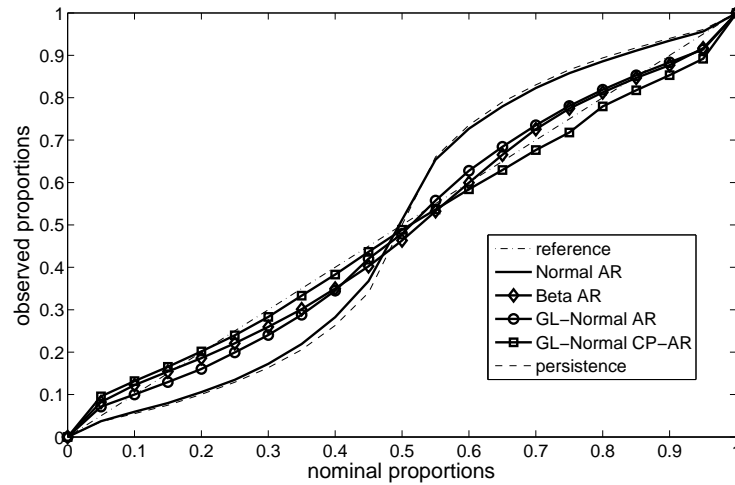


Figure 5. Evaluation of the reliability of predictive densities with reliability diagrams in the form of quantile-quantile plots.

above 0.5, and inversely for nominal proportions below 0.5. A simple explanation 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 over-predicted when wind generation is at a low and high level. The calibration results for the persistence benchmark are very close to that of Normal-AR predictive densities owing to the similarity in their definition. In contrast when employing GL-Normal distributions for probabilistic forecasting, predictive densities do not tend to be systematically over- or under-dispersive due to their natural mean-variance relationship. This appears to be valid for both types of dynamics considered for the mean of predictive densities. Similarly to GL-Normal predictive densities, the Beta-AR densities have an acceptable level of reliability due to their natural shape variations as a function of their location parameter.

It was argued that the doubled-bounded nature of wind power generation causes the fact that the shape of predictive densities ought to vary depending upon the power level. This motivated the proposal of GL-Normal predictive densities. In parallel following Gneiting *et al.* (2007) and other authors, the correspondence of predicted probabilities and observed frequencies should be verified for the evaluation set as a whole but potentially also for any subset. The calibration of predictive densities is hence further investigated as a function of the level of power, in a spirit similar to that of Pinson *et al.* (2007b). Mutually exclusive subsets of the evaluation set are defined based on measurements, so that they are the same for all probabilistic forecasting approaches, following the proposal for calibration assessment based on forecast strata introduced and extensively discussed by Bröcker (2008). The criterion employed for quantifying the lack of calibration of predictive densities is the Mean Absolute Deviation from Calibration (MADC), defined as the integrated absolute difference between the reference case of perfectly calibrated forecast and the observed proportions of the quantiles of predictive densities. It then represents the area between the reliability

Table 4. Results for the evaluation of the calibration of predictive densities conditional to the level of measured power and overall, with the MADC criterion (in %).

Level	[0,0.2]	[0.2,0.4]	[0.4,0.6]	[0.6,0.8]	[0.8,1]	Overall
Climatology	38.86	26.64	20.03	23.33	34.91	0
Persistence	7.85	6.92	5.99	5.78	5.99	5.82
Normal AR	19.38	10.71	2.99	2.70	11.61	7.91
Beta AR	21.71	10.56	3.28	2.26	16.75	2.41
GL-Normal AR	6.50	3.43	6.34	5.80	6.85	2.85
GL-Normal CP-AR	9.19	1.78	3.45	3.17	11.15	2.22

curves and the diagonal as in the case of the reliability diagram of Figure 5. Lower MADC values indicate more reliable predictive densities. It is calculated in the case of the whole evaluation set and over subsets defined by various ranges of measured wind power, e.g. when normalized wind power generation is between 0 and 0.2. The corresponding results are summarised in Table 4, with the “overall” column giving the MADC values for the reliability assessment presented in Figure 5.

Climatology has perfect reliability when evaluated over the whole evaluation set since predictive densities are unconditional and directly derived from the collected measurements. This unconditional nature is clearly affecting its reliability when assessed over the various subsets defined by the level of observed wind power owing to the level-varying characteristics of the stochastic process. A similar result can be observed for the case of Normal-AR and Beta-AR predictive densities for low and high power values. For medium power values, the low MADC values indicate that both Normal and Beta assumptions, as well as the related variance estimation, are fairly acceptable. When looking across the various levels of measured power and at overall results, both persistence and GL-Normal predictive densities seem to be better calibrated. One remembers, however, that calibration is only one aspect of probabilistic forecasting skill (Gneiting *et al.* 2007). Appropriate ranking of competing approaches can only be carried out based on the skill score values of Table 3. Based on all the results presented, GL-Normal predictive densities appear to be best, with a fair advantage to the CP-AR dynamics for their location and scale parameters. And even if persistence seems to have an acceptable calibration, this does not translate to these benchmark probabilistic forecasts having a competitive skill level (as seen from Table 3).

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. Let us have a look at some of the parameters of the CP-AR dynamic model, more specifically the coefficient for the first autoregressive lag, as well as the variance of the GL-Normal predictive densities, at the end of the evaluation set. These various quantities are represented in polar coordinates in Figure 7. Variations as a function of the wind direction ω can be observed. The variance of predictive densities appears to be significantly higher for Westerly and North-Westerly winds, and lower for the other sectors. This is in line with the extensive wind data analysis performed by Vincent *et al.* (2010) and Vincent *et al.* (2011) based on meteorological measurements obtained at Horns Rev. It was 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 for winds coming from land.

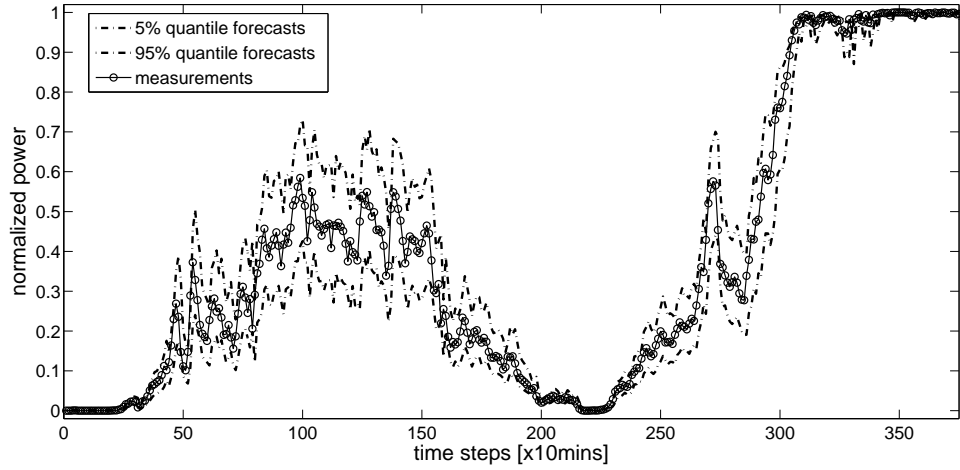


Figure 6. Episode of 375 time-steps (with a time resolution of 10 minutes), with quantile forecasts of nominal proportions 0.05 and 0.95, along with related normalised wind power measurements. These quantile forecasts are extracted from the predictive densities with CP-AR dynamics.

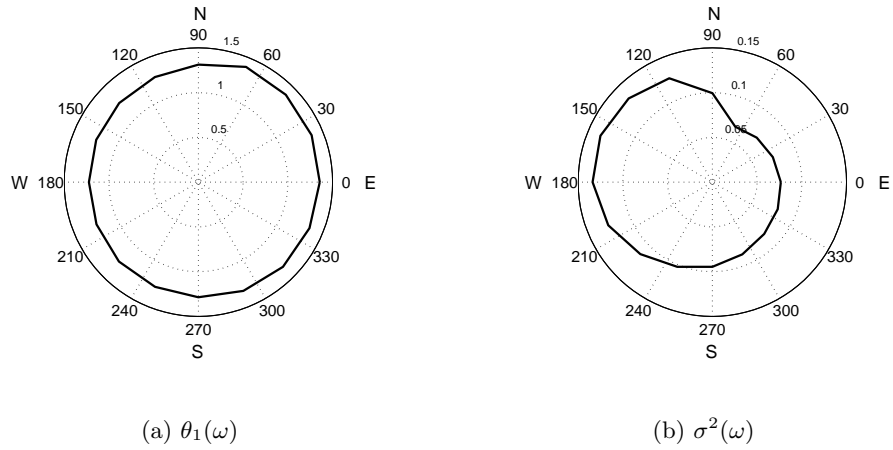


Figure 7. Some of the CP-AR model parameters as a function of the wind direction ω at the end of the evaluation set.

6. Concluding remarks

A generalization of the logit-Normal distribution was introduced in order to properly account for the nonlinear and double-bounded nature of wind power generation in short-term probabilistic forecasting. It was observed that its characteristics make it more appropriate than the more commonly assumed Normal and Beta distributions. We believe that GL-Normal distributions could be viewed as a basic feature of parametric probabilistic forecasting methodologies for lead times between a few minutes and a few days ahead. Being fully characterised by a location and a scale parameter only (assuming that the shape parameter is a meta-parameter estimated once and for all from available data), fairly simple models can be employed for the description of its dynamics.

In our proposal forecasting methodology, predictive densities of wind power take the form of discrete-continuous mixtures consisting of these GL-Normal distributions with potential concentration of probability mass at the bounds of the unit interval $[0, 1]$. These mixtures could be rethought in the future based on further data analysis and if looking at longer lead times. Since forecast uncertainty may not be null when predicting zero or nominal power, probability masses should be replaced by exponential distributions, the parameters of which would then also have to be estimated. For these more general mixtures, it will certainly be beneficial to develop estimation methods in a dynamic (Bayesian) model averaging framework, as in Raftery *et al.* (2010).

The models employed for the location and scale parameters of the predictive densities account for the autoregressive nature of the wind power generation process, as well as the non-linear influence of wind direction. For further lead times and multi-step ahead forecasting some of the numerous models described in the literature could be employed for the dynamics of GL-Normal distribution parameters. They could for instance account for additional effects and covariates, like meteorological measurements which are known to be important for short lead times (Gallego *et al.* 2011), or meteorological forecasts for lead times further than 6 hours ahead. More particularly for very short-term forecasts, the proposal of new dynamic models should be supported by some of the recent and 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 (2011) suggest that regime-switching may be present in the time-series of wind power generation. This is confirmed by the results of Vincent *et al.* (2011) explaining that these switches relate to local thermal effects, atmospheric stability and to the presence of convective cells in the vicinity or over this site. Regime-switching approaches are also relevant for onshore sites especially in complex terrain environments, as investigated by Gneiting *et al.* (2006) when focusing on wind channelling effects.

GL-Normal distributions should also be considered for the modelling and forecasting of other nonlinear and bounded processes where it is expected or known that the effects of the bounds are not symmetrical, as for bounded economic variables (Wallis 1987). For the wind power application, the shape parameter ν of GL-Normal distributions will surely vary from site to site. In a more general manner, this parameter will depend upon the process considered and should be estimated from data. The cross-validation approach employed here yielded acceptable results. It may however be envisaged to develop a more rigorous framework based on e.g. maximum likelihood estimation for selecting ν .

Acknowledgements

The work presented has been partly supported by the European Commission under the SafeWind project (ENK7-CT2008-213740), by the Danish Public Service Obligation (PSO) programme through the project 'Radar@Sea' (PSO-2009-1-0226), and finally by the Danish Research Council for Technology and Production Sciences (FTP-274-08-0573), which are hereby acknowledged. We are thankful to 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, Niels K. Poulsen and Henrik Madsen for inspiring discussions on probabilistic forecasting and bounded time-series. We finally acknowledge two anonymous reviewers, an associate editor, as well as Richard Chandler, for useful comments and suggestions.

References

- Ailliot, P., Monbet, V. and Prevosto M. (2006) An autoregressive model with time-varying coefficients for wind fields. *Environmetrics*, **17**, 107–117.
- Aitchison, J. and Shen, S. (1980) Logistic-normal distributions: some properties and uses. *Biometrika*, **67**, 261–272.
- Akhmatov, V. (2007) Influence of wind direction on intense power fluctuations in large offshore wind farms in the North Sea. *Wind. Eng.*, **31**, 59–64.
- Bacher, P., Madsen, H. and Nielsen, H.Aa. (2009) Online short-term solar power forecasting. *Sol. Energy*, **83**, 1772–1783.
- Bludszuweit, H., Domínguez-Navarro, J.A. and Llombart A. (2008) Statistical analysis of wind power forecast errors. *IEEE T. Power Syst.*, **23**, 983–991.
- Box, G.E.P. and Cox, D.R. (1964) An analysis of transformations. *J. Royal Stat. Soc. B*, **26**, 211–243.
- Bremnes, J.B. (2006) A comparison of a few statistical models for making quantile wind power forecasts. *Wind Energy*, **9**, 3–11.
- Bröcker, J. (2008) On reliability analysis of multi-categorical forecasts. *Nonlin. Proc. Geophys.*, **15**, 661–673.
- Cloke, H.L. and Pappenberger, F. (2009) Ensemble flood forecasting: a review. *J. Hydrol.*, **375**, 613–626.
- Costa, A., Crespo, A., Navarro, J., Lizcano, G., Madsen, H. and Feitosa E (2008) A review on the young history of the wind power short-term prediction. *Renew. Sust. Energ. Rev.*, **12**, 1725–1744.
- Fan, J.Q. and Zhang, W. (2008) Statistical methods with varying-coefficient models. *Stat. Interf.*, **1**, 179–195.
- Frederic, P. and Lad, F. (2008) Two moments of the logit-normal distribution. *Comm. Stat. B-Simul.*, **37**, 1263–1269.
- Gallego, C., Pinson, P., Madsen, H., Costa, A. and Cuerva, A. (2011). Influence of local wind speed and direction on wind power dynamics Application to offshore very short-term forecasting. *Appl. Energ.*, **88**, 4087–4096.
- Giebel, G., Brownsword, R., Kariniotakis, G., Denhard, M. and Draxl C. (2011) The state of the art in short-term prediction of wind power - A literature overview, 2nd edition. Technical report, EU project Anemos.plus. [Available online: <http://orbit.dtu.dk>]
- Gneiting, T. (2011) Quantiles as optimal point predictors. *Int. J. Forecasting*, **27**, 197–207.
- Gneiting, T., Balabdaoui, F. and Raftery A.E. (2007) Probabilistic forecasts, calibration and sharpness. *J. Royal Stat. Soc. B*, **69**, 243–268.
- Gneiting, T., Larson, K., Westrick, K., Genton, M.G. and Aldrich, E. (2006) Calibrated probabilistic forecasting at the stateline wind energy center - The regime-switching space-time method. *J. Am. Stat. Assoc.*, **101**, 968–979.

- Hering, A.S. and Genton, M.G. (2009) Powering up with space-time wind forecasting. *J. Am. Stat. Assoc.*, **105**, 96–104.
- Johnson, N. (1949) Systems of frequency curves generated by methods of translation. *Biometrika*, **36**, 146–176.
- Kagan, Y.Y. and Jackson, D.D. (2000) Probabilistic forecasting of earthquakes. *Geophys. J. Int.*, **143**, 438–453.
- Kristoffersen, J.R. and Christiansen, P. (2003) Horns Rev offshore wind farm: its main controller and remote control system. *Wind. Eng.*, **27**, 351–359.
- Lange, M. (2005) On the uncertainty of wind power predictions - Analysis of the forecast accuracy and statistical distribution of errors. *J. Solar Energy-T. ASME*, **127**, 177–184.
- Lau, A. and McSharry, P. (2010). Approaches for multi-step density forecasts with application to aggregated wind power. *Ann. Appl. Stat.*, **4**, 1311–1341.
- Lesaffre, E., Rizopoulos, D. and Tsonoka, R. (2007) The logistic-transform for bounded outcome scores. *Biostatistics*, **8**, 72–95.
- Leutbecher, M. and Palmer, T.N. (2008) Ensemble forecasting. *J. Comput. Phys.*, **227**, 3515–3539.
- Li, T.H. (2008) On exponentially weighted recursive least squares for estimating time-varying parameters. *J. Stat. Theo. Pract.*, **2**, 339–354.
- Ljung, L. and Söderström, T. (1983) *Theory and Practice of Recursive Estimation*. MIT Press: Boston.
- Madsen, H. (2007) *Time Series Analysis*. Chapman & Hall/CRC: London.
- Madsen, H., Pinson, P., Nielsen, T.S., Nielsen, H.Aa. and Kariniotakis, G. (2005) Standardizing the performance evaluation of short-term wind power prediction models. *Wind Eng.*, **29**, 475–489.
- Mead, R. (1965) A generalized logit-normal distribution. *Biometrics*, **21**, 721–731.
- Møller, J.K., Nielsen, H.Aa. and Madsen, H. (2008) Time-adaptive quantile regression. *Comput. Stat. Data An.*, **52**, 1292–1303.
- Nielsen, H.Aa., Nielsen, T.S., Joensen, A.K. and Madsen, H. and Holst, J. (2000) Tracking time-varying coefficient functions. *Int. J. of Adapt. Control*, **14**, 813–828.
- Nielsen, H.Aa., Nielsen, T.S., Madsen, H., Badger, J., Giebel, G., Landberg, L., Sattler, K., Voulund, L. and Tøfting, J. (2006) From wind ensembles to probabilistic information about future wind power production - Results from an actual application. in *Proc. IEEE PMAPS Conference, Probabilistic Methods Applied to Power Systems*, Stockholm, Sweden.
- Pinson, P. (2006) *Estimation of the uncertainty in wind power forecasting*. Ph.D. thesis, Ecole des Mines de Paris, Paris, France. [Available online: <http://pastel.paristech.org>]
- Pinson, P., Chevallier, C. and Kariniotakis, G. (2007) Trading wind generation from short-term probabilistic forecasts of wind power. *IEEE T. Power Syst.*, **22**, 1148–1156.
- Pinson, P. and Kariniotakis, G. (2010) Conditional prediction intervals of wind power. *IEEE T. Power Syst.*, **25**, 1845–1856.
- Pinson, P. and Madsen, H. (2009) Ensemble-based probabilistic forecasting at Horns Rev. *Wind Energy*, **12**, 137–55.
- Pinson, P. and Madsen, H. (2011) Adaptive modeling and forecasting of wind power fluctuations with Markov-switching autoregressive models. *J. Forecasting*, available online.
- Pinson, P., Nielsen, H.Aa., Madsen, H. and Nielsen, T.S. (2008) Local linear regression with adaptive orthogonal fitting for the wind power application. *Stat. Comput.*, **18**, 59–71.
- Pinson, P., Nielsen, H.Aa., Møller, J.K., Madsen, H. and Kariniotakis, G. (2007) Nonparametric probabilistic forecasts of wind power: required properties and evaluation. *Wind Energy*, **10**, 497–516.
- Raftery, A.E., Karny, M. and Ettlér, P. (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics*, **52**, 52–66.
- Sanchez, I. (2006) Short-term prediction of wind energy production. *Int. J. Forecasting*, **22**, 43–56.
- Stone, M. (1974) Cross-validation and assessment of statistical predictions (with discussion). *J.*

- Royal Stat. Soc. B*, **36**, 111–147.
- Sørensen, P.E., Cutululis, N.A., Viguera-Rodríguez, A., Jensen, L.E., Hjerrild, J., Donovan, M.H. and Madsen, H. (2007) Power fluctuations from large wind farms. *IEEE T. Power Syst.*, **22**, 958–965.
- Sørensen, P. and co-authors (2005) Operation and control of large wind turbines and wind farms - Final report. Technical report, Risø National Laboratories for Sustainable Energy, Technical University of Denmark, Risø-R-1532(EN).
- Tay, A.S. and Wallis, K.F. (2000) Density forecasting: a survey. *J. Forecasting*, **19**, 235–254.
- Taylor, J.W., McSharry, P.E. and Buizza, R. (2009) Wind power density forecasting using wind ensemble predictions and time series models. *IEEE T. Energy Conver.*, **24**, 775–782.
- Timmermann A. (2000) Density forecasting in economics and finance. *J. Forecasting*, **19**, 231–234.
- Usola, J. (2009) Probabilistic load flow with wind production uncertainty using cumulants and Cornish–Fisher expansion. *Int. J. Elec. Power. Syst.*, **31**, 474–481.
- Vincent, C.L., Giebel, G., Pinson, P. and Madsen, H. (2010) Resolving non-stationary spectral signals in wind speed time-series using the Hilbert-Huang transform. *J. App. Meteorol. Clim.*, **49**, 253–267.
- Vincent, C.L., Pinson, P. and Giebel, G. (2011) Wind fluctuations in the North Sea. *Int. J. Climatol.*, available online.
- Wallis, K.F. (1987) Time series analysis and bounded economic variables. *J. Time Ser. Anal.*, **8**, 115–123.