Ensemble-based Probabilistic Forecasting at Horns Rev

Pinson, Pierre; Madsen, Henrik

Published in:
Wind Energy

Link to article, DOI:
10.1002/we.309

Publication date:
2009

Document Version
Early version, also known as pre-print

submitted to Wind Energy - Special issue on Offshore Wind Energy *(August 6, 2008)*

Ensemble-based probabilistic forecasting at Horns Rev

Pierre Pinson* **and Henrik Madsen**
DTU Informatics, Technical University of Denmark, Kgs. Lyngby, Denmark

Abstract

For management and trading purposes, information on short-term wind generation (from few hours to few days ahead) is crucial at large offshore wind farms, since they concentrate a large capacity at a single location. The most complete information that can be provided today consists of probabilistic forecasts, the resolution of which may be maximized by using meteorological ensemble predictions as input. The paper concentrates on the test case of the Horns Rev wind farm over a period of approximately one year, in order to describe, apply and discuss a complete ensemble-based probabilistic forecasting methodology. In a first stage, ensemble forecasts of meteorological variables are converted to power through a suitable power curve model. This model employs local polynomial regression, and is adaptively estimated with an orthogonal fitting method. The obtained ensemble forecasts of wind power are then converted into predictive distributions with an original adaptive kernel dressing method. The shape of the kernels is driven by a mean-variance model, the parameters of which are recursively estimated in order to maximize the overall skill of obtained predictive distributions. Such a methodology has the benefit of yielding predictive distributions that are of increased reliability (in a probabilistic sense) in comparison with the raw ensemble forecasts, while taking advantage of their high resolution.

Keywords: wind power, offshore, ensemble forecasting, recalibration, probabilistic forecasting, reliability, resolution, skill

* *Corresponding author:*
P. Pinson, DTU Informatics, Technical University of Denmark,
Richard Petersens Plads (bg. 321 - 020), DK-2800 Kgs. Lyngby, Denmark.
Tel: +45 4525 3428, fax: +45 4588 2673, email: pp@imm.dtu.dk, webpage: www.imm.dtu.dk/∼pp

Introduction

Future developments of wind power installations are more likely to take place offshore, due to space availability, less problems with local population acceptance, and more steady winds. This is especially the case for countries that already experience a high wind power penetration onshore, like Germany and Denmark. This latter country hosts the two largest offshore wind farms worldwide: Nysted and Horns Rev, whose nominal capacities are of 165.5 and 160 MW, respectively. Today, each of these wind farms can supply alone 2% of the whole electricity consumption of Denmark [1]. Such large offshore wind farms concentrate a high wind power capacity at a single location. Onshore, the same level of installed capacity is usually spread over an area of significant size. As a consequence, forecast accuracy is even more paramount for such offshore location, since a phase error[1] for instance may easily translate to an energy imbalance of very significant magnitude. Such an imbalance would then call for regulation either at the local or at the grid level, resulting in additional costs for the wind power producer, the Transmission System Operator (TSO) and consequently for the electricity consumers. In the present paper, focus is given to look-ahead times in the range 1-48 hour ahead. For literature on shorter horizons (i.e. for the few-minutes to few-hours ahead range), we refer to e.g. [2, 3, 4, 5].

The potential accuracy of wind power prediction systems in offshore conditions, more specifically for the North Sea, is discussed in [6], based on the accuracy of wind forecasts and the ability to model wind profiles. In parallel, the forecast accuracy to be expected if having a pool of wind farms deployed offshore, following the German plans for 25GW offshore capacities, is studied in [7]. Very few literature exists however about actual performance of wind power forecasting systems at large offshore wind farms. The example of a forecasting method application and performance assessment results for an offshore wind farm of limited size in Denmark may be found in [8]. Conclusions in [6] indicate that significant developments will be necessary prior to appropriately understand, model and forecast offshore meteorological phenomena. In addition to that, it is recognized today that the question of wake effects inside and behind large wind farms comprises a real challenge, both for the resource assessment and the forecasting applications [9]. Therefore, since related wind power forecast accuracy may not be dramatically increased in the short term, and since it is known that forecast accuracy is highly situation-dependent, emphasis has to be put on providing forecast users with information on forecast uncertainty. It appears today that for a large range of decision-making problems, the most appropriate way of estimating and communicating forecast uncertainty is with probabilistic forecasts, i.e. forecasts of the probability distribution (or some of its quantiles) of wind generation for each look-ahead time. Relevant literature on this topic includes e.g. [10, 11, 12, 13].

It is foreseen that probabilistic forecasts obtained from ensemble predictions of meteorological variables would have a higher resolution, i.e. a higher ability to resolve among situations with various uncertainty levels, than those derived from purely statistical methods [14]. The ability of ensemble forecasts of wind power (obtained from different types of meteorological ensembles) to inform on the expected level of forecast uncertainty is discussed in [15]. Even though this comprises a promising approach, it is known there are two remaining issues related to ensemble-based probabilistic forecasting. They are namely the power curve model used for conversion of meteorological variables to power generation, as well as the recalibration of wind power ensembles in order to obtain reliable predictive densities [16]. These two points are the focus of the present paper, with application to the Horns Rev wind farm in Denmark. The meteorological ensemble predictions used as input originate from a Multi-Scheme Ensemble Prediction System (MSEPS) described in e.g. [17]. The nonparametric method employed for modeling the conversion function from meteorological variables to wind power production is based on local polynomial regression. Local coefficients are adaptively estimated with orthogonal fitting [18]. This conversion yields ensemble forecasts of wind generation. Ensemble forecast members are subsequently dressed with Gaussian kernels,

[1] A phase error consists of a timing error in ramps up (or down) in wind power production

which are parameterized with a mean-variance model. The relevant parameters are tracked with a recursive Maximum Likelihood (ML) estimation method. The overall ensemble-based forecasting methodology permits to derive predictive distributions of wind power with a maximized overall skill, being a trade-off between reliability and resolution. The only requirements for its real-world application are the availability of ensemble forecasts of meteorological variables for the wind farm, as well as the availability of online measurements of wind power generation.

The paper is structured as following. In a first part, the methodology for the conversion of ensemble forecasts of meteorological variables to power is developed. Focus is subsequently given to the kernel dressing of wind power ensemble forecasts, in order to obtain predictive distributions of wind power. The parameterization of the Gaussian kernels, as well as the method for adaptive estimation of its parameters are detailed. Application results for the Horns Rev test case permit to illustrate the outputs of the proposed forecasting methodology, in addition to evaluate the resulting probabilistic forecasts of wind power. Results on the case-study show a significant reliability of obtained probabilistic forecasts, with a very high resolution. Conclusions end the paper, along with perspectives related to future developments. Note that through the whole paper all variables considered are normalized by their maximum value over the dataset, and thus comprised between 0 and 1.

Generation of ensemble forecasts of wind power

The first step for obtaining ensemble-based probabilistic forecasts of wind generation relates to the conversion of ensemble forecasts of meteorological variables to wind power. Let y_{t+k} be the measured power value at time $t+k$, while $\hat{y}_{t+k|t}$ denotes a power forecast issued at time t for that same lead time. In parallel, $\hat{\mathbf{x}}_{t+k|t}$ is the corresponding vector of predicted meteorological variables. For the wind power application, relevant meteorological variables commonly include wind speed and direction, plus possibly e.g. air density, temperature, pressure or humidity. Note that more physical expertise on the local wind profiles, as well as on the wind-to-power conversion process, may motivate the consideration of other relevant physical variables.

In a general manner, a model of the power curve for the conversion of meteorological forecasts to power at the level of a wind farm, for a given forecast horizon k, writes

$$y_{t+k} = g_{t,k}(\hat{\mathbf{x}}_{t+k|t}) + \varepsilon_{t+k}, \quad \forall t, k \tag{1}$$

where $\{\varepsilon_{t+k}\}_t$ is a sequence of independent and identically distributed (i.i.d.) random variables, such that $\mathbb{E}[\varepsilon_{t+k}] = 0$ and $\sigma^2(\varepsilon_{t+k}) < \infty$. In parallel, $g_{t,k}$ is a nonlinear function to be estimated from data. A t-index is used in order to express the fact that g may be non-stationary, and hence that the power curve model may be allowed to have slow variations over time. Smooth changes in the power curve model may result from e.g. ageing of the turbines or maintenance-related issues. The k-index indicates that a separate power curve model is defined for each prediction horizon, since the characteristics of the g-function may also depend on k. For instance, it may intuitively be expected that the level of noise in meteorological forecasts — in other words of forecast error — would increase as the lead time gets further.

Ensemble forecasts of meteorological variables consist of a set of m alternative predictions for each look-ahead time. They may be generated by perturbing initial conditions of Numerical Weather Prediction (NWP) models, by employing a stochastic parameterization of such models, or alternatively by using different physical parameterizations of the models involved. An accessible and complete introduction to ensemble forecasting in meteorology is given in [19]. Denote by $\hat{\mathbf{x}}^{(j)}_{t+k|t}$ the forecasts of meteorological variables (at time t for lead time $t+k$) given by the j^{th} member ($j = 1, \ldots, m$) of

the ensemble set. If no ensemble member is found (or expected to be) systematically better than the others, one may consider the mean of all ensemble members,

$$\bar{\mathbf{x}}_{t+k|t} = \frac{1}{m}\sum_{j=1}^{m} \hat{\mathbf{x}}^{(j)}_{t+k|t}, \quad \forall t, k \tag{2}$$

as the best forecast that can be extracted from these sets of alternative predictions, see discussion in [20] for instance. As a consequence the power curve model (1) can be rewritten as

$$y_{t+k} = g_{t,k}(\bar{\mathbf{x}}_{t+k|t}) + \varepsilon_{t+k}, \quad \forall t, k \tag{3}$$

with the noise sequence $\{\varepsilon_{t+k}\}_t$ having properties similar to those of the noise sequence in (1), i.e. centered and with finite variance. The above equation permits to model the function for the conversion of the best available forecast of meteorological variables to measured wind generation, and hopefully comprises the most relevant power curve model for ensemble conversion.

An appealing approach to the modeling of the power curve defined above consists of local polynomial regression, for which the $g_{t,k}$ function is locally approximated with polynomials at a certain number of fitting points, defined to span the range of potential values of $\bar{\mathrm{x}}_{t+k|t}$ [21]. Indeed, an advantage of local polynomial regression is its nonparametric nature, i.e. no assumption is made on the shape of the power curve. Such power curve model may then be fitted with the aim of minimizing either a classical Least Squares (LS, see [22]) or a Total Least Squares (TLS, see [18]) criterion. The latter method is more generally referred to as orthogonal fitting. The difference in concept between these two alternative fitting approaches is illustrated in Figure 1. In the LS case, the distance to be minimized between observations and the model is defined along the power axis, while for the orthogonal fitting case, this distance is defined as that between observations and their orthogonal projections on the power curve model. It is shown in [18] that the choice of the best fitting method actually depends on the intended application. If one aims at minimizing a Root Mean Square Error (RMSE) criterion for the case of point prediction, then one should prefer LS-fitting of the power curve model. However, if one instead wants to obtain the most faithful description of the true power curve (as it is the case here), one should rather apply an orthogonal fitting method. Conversion of meteorological ensembles to power via a LS-fitted power curve would dampen the uncertainty present in meteorological forecasts and would result in severe underestimation of uncertainties in wind power forecasts.

This estimated power curve is used as a model of the true power curve for the wind farm, and is then employed for the conversion of each meteorological ensemble member to wind power production, every time a new set of meteorological forecasts are provided. If denoting by $\hat{g}_{t,k}$ the power curve model estimated according to equation (3), at time t for k-hour ahead forecasting, the j^{th} ensemble member of wind power production $\hat{y}^{(j)}_{t+k|t}$ for lead time $t+k$ is given by

$$\hat{y}^{(j)}_{t+k|t} = \hat{g}_{t,k}(\hat{\mathbf{x}}^{(j)}_{t+k|t}) \tag{4}$$

while the single point forecast that is extracted from the set of ensemble members is chosen to be its mean, i.e.

$$\hat{y}_{t+k|t} = \frac{1}{m}\sum_{j=1}^{m} \hat{y}^{(j)}_{t+k|t} \tag{5}$$

One could argue that a different power curve model should be estimated for each ensemble member. Indeed, in the case of multi-model meteorological forecasts, for instance for which each model has a different physical parameterization, each ensemble member has its own characteristics. This would not be the case for meteorological ensemble forecasts obtained with initial value perturba-

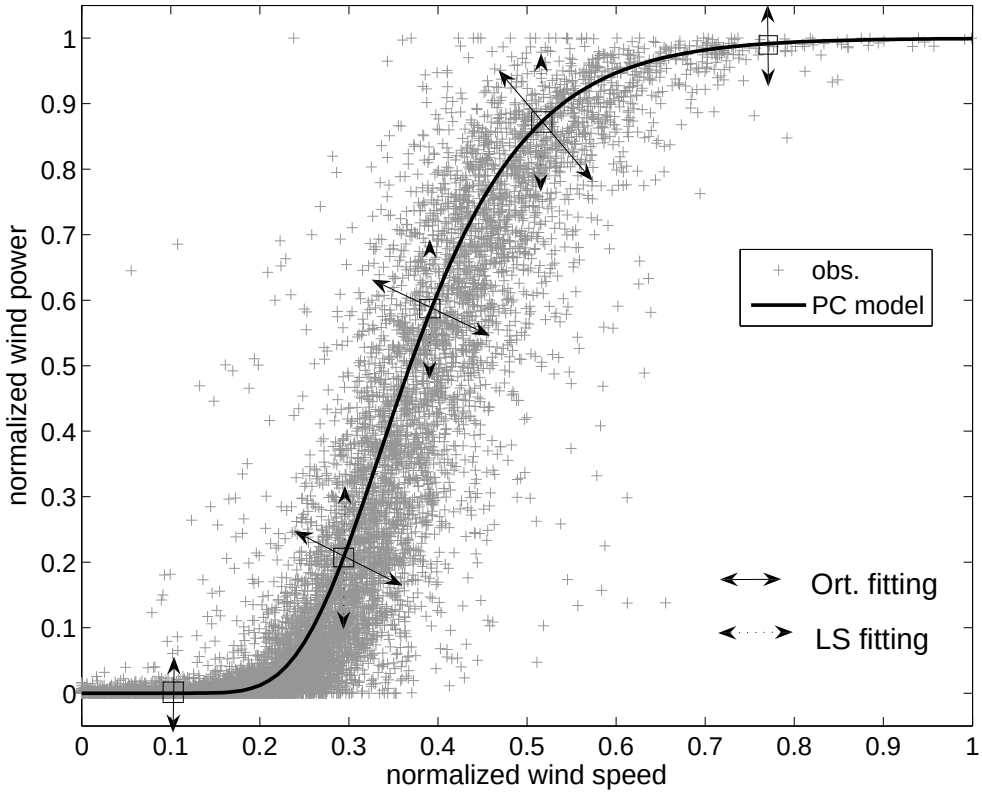


FIGURE 1: Fitting of the power curve model with classical least-squares and with orthogonal fitting methods. The power curve model is nonparametric and uses local polynomial regression.

tions e.g. with singular vectors or bred modes, since in contrast ensemble members would not be distinguishable [23]. The approach proposed here can be used as a generic approach whatever the type of meteorological ensemble forecasts considered as input, and it has the advantage of lowering computational costs, since only one power curve model for each forecast horizon has to be estimated.

Adaptive kernel dressing of wind power ensemble forecasts

The methodology described above permits to obtain ensemble forecasts of wind power. It is generally accepted that ensemble forecasts of wind power are not reliable from a probabilistic point of view, i.e. that the observed probabilities significantly deviate from the nominal ones [16]. In practice, if probabilistic information is derived from such ensembles, one then witnesses an underestimation of potential extreme events, both in terms of range and probabilities. Regarding the shape of predictive distributions, too much probability is concentrated in their central part, while the tails are not thick (and long) enough. This will be illustrated further in the paper. In order to correct for this, it is necessary to recalibrate ensemble forecasts. This is what is performed here, by adaptive kernel dressing of the ensemble predictions of wind power obtained above.

Remember that at a given time t and for each look-ahead time k, ensemble forecasts consist of a number m of alternative predictions $\hat{y}^{(j)}_{t+k|t}$ ($j = 1, \ldots, m$). The methodology introduced in the following is described for a specific horizon k. It can then be independently applied for every look-ahead

times up to the forecast length of the ensemble predictions considered. The basis of the method is to define a predictive density of wind generation as the weighted sum of kernels associated to each ensemble member. Such proposal is extensively described in e.g. [23, 24]. However, since it is known that the characteristics of wind power forecasting uncertainty are strongly dependent on the level of predicted power [15], it is proposed here to parameterize kernels with a mean-variance model. This hence permits to control the shape of kernels depending on the level of predicted power for each of the ensemble members. And, because it is known that the uncertainty of wind power forecast may be nonstationary, it is also proposed to recursively estimate the parameters of the kernel mean-variance model. Long-term variations in forecast uncertainty characteristics can then be accommodated.

Nonparametric predictive densities from kernel dressing

Let us denote by $\hat{f}_{t+k|t}(y)$ the predictive distribution of wind power issued at time t for lead time $t+k$. The idea of kernel dressing of ensemble members consists in saying that $\hat{f}_{t+k|t}(y)$ can be written as a weighted combination of kernels associated to each of the ensemble members. If writing $\hat{f}^{(j)}_{t+k|t}(y)$ the kernel associated to the j^{th} ensemble member $\hat{y}^{(j)}_{t+k|t}$, $j = 1, \dots, m$, this yields

$$\hat{f}_{t+k|t}(y) \;=\; \sum_{j=1}^{m} w_j \hat{f}^{(j)}_{t+k|t}(y) \tag{6}$$

where the sum of the weights w_j is required to sum to 1,

$$\sum_{j=1}^{m} w_j \;=\; 1 \tag{7}$$

From a conceptual point of view, such weights represent the contribution of each ensemble member and its associated kernel to the final predictive distribution. It may be envisaged that some of them have a higher ability to explain uncertainty in wind power forecasts, while some other may have a more limited one. However, in order to simplify the estimation problem that will be formulated in the following (since it would be necessary to recursively estimate m weights for each forecast horizon), the weights are here set to $w_j = 1/m, \; \forall j$. This hence translates to giving the same importance to the information provided by each of the ensemble members. Such simplification actually makes perfect sense when considering the pure ensemble predictions discussed by [23], since it would not be possible to distinguish any ensemble member from the others. This would be the case for ensemble forecasts of wind power produced from those of meteorological variables from the European Center for Medium-Range Weather Forecasts (ECMWF) for instance. This equal weighting also appears relevant for the case of meteorological ensemble forecasts from a multi-model or multi-parameterization approach, if one assumes it is not possible to deem such or such prediction as more relevant in terms of information about forecast uncertainty.

In a second stage, it is necessary to propose a shape for the kernel functions in equation (6). The most straightforward choice is to employ Gaussian kernels. This choice is motivated by the fact that in theory, any probabilistic density may be approximated by a sum of Gaussian kernels [25], while Gaussian kernels have nice properties that will ease the derivation of recursive formulas for the adaptive estimation of its parameters. A Gaussian kernel for the j^{th} ensemble member $\hat{y}^{(j)}_{t+k|t}$

has the following form

$$\hat{f}^{(j)}_{t+k|t}(y) \;=\; \frac{1}{{\sigma^{(j)}_{t,k}}^2\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y-\hat{y}^{(j)}_{t+k|t}}{\sigma^{(j)}_{t,k}}\right)^2\right) \tag{8}$$

where $\sigma^{(j)}_{t,k}$ corresponds to its standard deviation. Employing such a formulation for Gaussian kernels translates to making the assumption that the various ensemble members are unbiased, since each of them is centered on its corresponding ensemble member. A k-index is used for the kernel parameters in order to reflect the fact that they will be different for each forecast horizon. In parallel, the t-index reflects the fact that the kernel parameters may evolve with time. By definition, a Gaussian kernel is symmetric and centered on the forecast itself. However, a predictive distribution formulated as a weighted sum of kernels will certainly take the form of a non-symmetric distribution (and possibly multimodal), thus being consistent with the know characteristics of wind power forecast uncertainty [15, 26].

Actually, a crucial issue related to wind power forecast uncertainty is that the standard deviation of prediction errors is directly related to the level of predicted power. Such standard deviation is of lower magnitude for predicted power values in the low and high ranges, while it reaches much higher magnitude in the middle range of the power curve [26]. This would hence imply that the standard deviation of a Gaussian kernel has to be related to its mean, i.e. to the forecast given by the related ensemble member. In practice here, this is performed by defining a mean-variance model that expresses $\sigma^{(j)}_{t,k}$ as a function of $\hat{y}^{(j)}_{t+k|t}$. From the characteristics of forecast uncertainty observed and discussed in [15], it appears appropriate to define this mean-variance model as a logistic function, i.e.

$$\sigma^{(j)}_{t,k} \;=\; \tau^0_{t,k} + \tau^1_{t,k}\left(1-\hat{y}^{(j)}_{t+k|t}\right)\hat{y}^{(j)}_{t+k|t} \tag{9}$$

where $\tau^1_{t,k}$ controls the shape of the logistic function, while $\tau^0_{t,k}$ is a level term, since Gaussian kernels would have a minimum width for wind power forecasts at the 0 or nominal power levels. Note that both parameters of the mean-variance model are not specific to any particular member. Indeed, being consistent with the equal weighting of ensemble members in equation (6) and with the core idea of them not being distinguishable, it is assumed that the same mean-variance model would be valid for all ensemble members. As a consequence, whatever the number of ensemble members, there will always be only two parameters to estimate for each forecast horizon, which are namely $\tau^0_{t,k}$ and $\tau^1_{t,k}$.

Some restrictions have to be set on the range of potential values for the mean-variance model parameters. Since the standard deviation of the kernels must always be positive, this implies that $\tau^0_{t,k} > 0$. In parallel, for the mean-variance model to have a logistic function shape, one must also have $\tau^1_{t,k} > 0$. Finally, it may be envisaged to define a maximum value for each of these parameters, that we will denote by τ^0_s and τ^1_s. Instead of defining strict constraints on the range of $\tau^0_{t,k}$ and $\tau^1_{t,k}$, it appears preferable to employ a suitable transformation, i.e.

$$\nu^i_{t,k} \;=\; \ln\left(\frac{\tau^i_{t,k}}{\tau^i_s-\tau^i_{t,k}}\right), \quad i=0,1 \tag{10}$$

with the corresponding inverse transform as

$$\tau^i_{t,k} \;=\; \frac{\exp(\nu^i_{t,k})}{1+\exp(\nu^i_{t,k})}, \quad i=0,1 \tag{11}$$

Such transformation makes that $\tau_{t,k}^i$ $(i = 0, 1)$ is restricted to the range $(0, \tau_s^i)$, while the transformed variable $\nu_{t,k}^i$ is free to take any value in $\mathbb{R}$. In the following, $\boldsymbol{\tau}_{t,k}$ and $\boldsymbol{\nu}_{t,k}$ will be used as vector notations for the mean-variance model parameters and their transformed counterparts

$$\boldsymbol{\tau}_{t,k}{}^\top = [\tau_{t,k}^0 \;\; \tau_{t,k}^1], \qquad \boldsymbol{\nu}_{t,k}{}^\top = [\nu_{t,k}^0 \;\; \nu_{t,k}^1] \tag{12}$$

Adaptive estimation with recursive Maximum Likelihood estimation

Bayesian Model Averaging has been proposed in [24] for recalibration of meteorological ensemble forecasts (of sea-level pressure and surface temperature). The core idea of the method proposed here is similar, except that the weights are chosen to be fixed, the estimation focuses on the mean-variance model for the kernel parameters (instead of the weight to be assigned to each ensemble member), and it has an adaptive nature. This then translates to a Maximum Likelihood (ML) estimation method, where it is aimed at adaptively maximizing the likelihood of the wind power measurements, given the predictive densities resulting from the model. In mathematical terms, the objective function to be minimized at time t can be written as

$$S_{t,k}(\boldsymbol{\nu}) = -\frac{1}{n_\lambda} \sum_{i=1}^{t-k} \lambda^{t-k-i} \ln\left(u_i(\boldsymbol{\nu})\right) \tag{13}$$

where λ, $\lambda \in (0, 1)$, is the forgetting factor allowing for adaptivity in time (by giving less weight to old observations), and n_λ is the effective number of observations,

$$n_\lambda = \frac{1}{1-\lambda} \tag{14}$$

used for normalizing the objective function. We restrict ourselves to the case for which $\lambda < 1$ in order for n_λ to be finite. In parallel, the term $u_i(\boldsymbol{\nu})$ denotes the likelihood of the observation y_i from the predictive density issued at time $i - k$ for lead time i, and given the model parameters $\boldsymbol{\nu}$,

$$u_i(\boldsymbol{\nu}) = P[y_i|\boldsymbol{\nu}] = \hat{f}_{i|i-k}(y_i) \tag{15}$$

The estimated parameters for the mean-variance model related to the Gaussian kernels, at time t and for horizon k, are then given as those which minimize the objective function of equation (13), that is,

$$\hat{\boldsymbol{\nu}}_{t,k} = \arg\min_{\boldsymbol{\nu}} S_{t,k}(\boldsymbol{\nu}) \tag{16}$$

The interest of this ML estimation method is that minimizing the objective function of (13) is equivalent to minimizing the logarithmic scoring rule known as ignorance, introduced and discussed by [27]. Ignorance considers a trade-off between reliability and resolution, which are the two properties wanted for probabilistic forecasts. Reliability corresponds to the probabilistic correctness of predictive distributions, while resolution stands for the ability of predictive distributions to resolve among situations with different levels of forecast uncertainty. Such a trade-off between reliability and resolution is commonly referred to as overall skill of probabilistic forecasts. For a thorough discussion on these aspects, and more specifically for the case of probabilistic forecasting of wind generation, see [13, 28]. Ignorance is a proper scoring rule which insures that a lower value of the score indeed corresponds to a higher skill of the probabilistic forecasts [29]. As a consequence, recursively minimizing the objective function in (13) will permit to obtain predictive distributions with maximized skill, given the ensemble forecasts used as input and the chosen model for conversion of ensemble forecasts into predictive densities.

From the formulation of the ML estimation problem given above, a corresponding recursive estimation procedure can be derived by applying the method described in [30]. Indeed, the basis for derivation of such recursive procedure is to employ a Newton-Raphson step for expressing the estimate $\hat{\boldsymbol{\nu}}_{t,k}$ as a function of the previous estimate $\hat{\boldsymbol{\nu}}_{t-1,k}$,

$$\hat{\boldsymbol{\nu}}_{t,k} \;=\; \hat{\boldsymbol{\nu}}_{t-1,k} - \frac{\nabla_{\boldsymbol{\nu}} S_{t,k}(\hat{\boldsymbol{\nu}}_{t-1,k})}{\nabla^2_{\boldsymbol{\nu}} S_{t,k}(\hat{\boldsymbol{\nu}}_{t-1,k})} \tag{17}$$

From equation (13), one can deduce that

$$S_{t,k}(\boldsymbol{\nu}_{t-1,k}) \;=\; \lambda S_{t,k}(\boldsymbol{\nu}_{t-1,k}) - \frac{1}{n_\lambda} \ln\left(u_t(\boldsymbol{\nu}_{t-1,k})\right) \tag{18}$$

which then can be used for deriving recursive formulas for the calculation of $\nabla_{\boldsymbol{\nu}} S_{t,k}$ and $\nabla^2_{\boldsymbol{\nu}} S_{t,k}$. Indeed, that for $\nabla_{\boldsymbol{\nu}} S_{t,k}$ writes

$$\nabla_{\boldsymbol{\nu}} S_{t,k}(\boldsymbol{\nu}_{t-1,k}) \;=\; -\frac{1}{n_\lambda} \frac{\nabla_{\boldsymbol{\nu}} u_t(\boldsymbol{\nu}_{t-1,k})}{u_t(\boldsymbol{\nu}_{t-1,k})} \tag{19}$$

since $\boldsymbol{\nu}_{t-1,k}$ is assumed to be the optimal estimate at time $t-1$, thus minimizing the objective function $S_{t-1,k}$, and yielding $\nabla_{\boldsymbol{\nu}} S_{t-1,k}(\boldsymbol{\nu}_{t-1,k}) = 0$. In a similar manner, by assuming that u_t is almost linear around the optimal estimate, a recursive formula for the Hessian of the objective function can be written as

$$\nabla^2_{\boldsymbol{\nu}} S_{t,k}(\boldsymbol{\nu}_{t-1,k}) \;=\; \lambda \nabla^2_{\boldsymbol{\nu}} S_{t-1,k}(\boldsymbol{\nu}_{t-1,k}) + \frac{1}{n_\lambda} \frac{\nabla_{\boldsymbol{\nu}} u_t(\boldsymbol{\nu}_{t-1,k}) \left(\nabla_{\boldsymbol{\nu}} u_t(\boldsymbol{\nu}_{t-1,k})\right)^\top}{u_t^2(\boldsymbol{\nu}_{t-1,k})} \tag{20}$$

Then, by defining the information vector

$$\mathbf{h}_{t,k} \;=\; \frac{\nabla_{\boldsymbol{\nu}} u_t(\hat{\boldsymbol{\nu}}_{t-1,k})}{u_t(\hat{\boldsymbol{\nu}}_{t-1,k})} \tag{21}$$

and the estimate of its inverse covariance matrix

$$\mathbf{R}_{t,k} \;=\; \nabla^2_{\boldsymbol{\nu}} S_{t,k}(\hat{\boldsymbol{\nu}}_{t-1,k}) \tag{22}$$

one deduces from equations (17)- (20) the two-step scheme for the updating of the $\boldsymbol{\nu}$-estimate at time t, i.e.

$$\hat{\boldsymbol{\nu}}_{t,k} \;=\; \hat{\boldsymbol{\nu}}_{t-1,k} + \frac{1}{n_\lambda} {\mathbf{R}_{t,k}}^{-1} \mathbf{h}_{t,k} \tag{23}$$

$$\mathbf{R}_{t,k} \;=\; \lambda \mathbf{R}_{t-1,k} + \frac{1}{n_\lambda} \mathbf{h}_{t,k} {\mathbf{h}_{t,k}}^\top \tag{24}$$

Note that for the application of this updating scheme, it is also assumed that the objective function can be seen as locally quadratic around the estimate $\hat{\boldsymbol{\nu}}_{t,k}$, and thus that

$$\nabla^2_{\boldsymbol{\nu}} S_{t,k}(\hat{\boldsymbol{\nu}}_{t,k}) \;\simeq\; \nabla^2_{\boldsymbol{\nu}} S_{t,k}(\hat{\boldsymbol{\nu}}_{t-1,k}) \;=\; \mathbf{R}_{t,k} \tag{25}$$

In practice, for determining the information vector $\mathbf{h}_{t,k}$ when the new wind power observation y_t is

made available, few quantities need to be calculated. Since we have

$$\nabla_{\boldsymbol{\nu}} u_t(\hat{\boldsymbol{\nu}}_{t-1,k}) \;=\; \left[\frac{\partial u_t}{\partial \nu^0}(\hat{\boldsymbol{\nu}}_{t-1,k}) \;\; \frac{\partial u_t}{\partial \nu^1}(\hat{\boldsymbol{\nu}}_{t-1,k})\right]^\top \tag{26}$$

it is necessary to compute two derivatives at each time step. From the definitions given in equations (6), (10) and (15), remembering that the weights are set to $1/m$, and after few mathematical developments, one obtains

$$\frac{\partial u_t}{\partial \nu^0}(\hat{\boldsymbol{\nu}}_{t-1,k}) \;=\; \frac{\hat{\nu}^0_{t-1,k}(1-\hat{\nu}^0_{t-1,k})}{m} \sum_{j=1}^{m} \frac{\partial \hat{f}^{(j)}_{t|t-k}}{\partial \sigma^{(j)}}(y_t) \tag{27}$$

and

$$\frac{\partial u_t}{\partial \nu^1}(\hat{\boldsymbol{\nu}}_{t-1,k}) \;=\; \frac{\hat{\nu}^1_{t-1,k}(1-\hat{\nu}^1_{t-1,k})}{m} \sum_{j=1}^{m} \hat{y}^{(j)}_{t|t-k}(1-\hat{y}^{(j)}_{t|t-k}) \frac{\partial \hat{f}^{(j)}_{t|t-k}}{\partial \sigma^{(j)}}(y_t) \tag{28}$$

with

$$\frac{\partial \hat{f}^{(j)}_{t|t-k}}{\partial \sigma^{(j)}}(y_t) \;=\; \left(\frac{(y_t - \hat{y}^{(j)}_{t|t-k})^2}{{\hat{\sigma}^{(j)}_{t-1,k}}^2} - 1\right) \frac{\hat{f}^{(j)}_{t|t-k}}{\hat{\sigma}^{(j)}_{t-1,k}}(y_t) \tag{29}$$

and finally where the standard deviation value $\hat{\sigma}^{(j)}_{t-1,k}$ is directly given by inverse transformation of $\hat{\boldsymbol{\nu}}_{t-1,k}$ through (11), then plugged into equation (9).

From the various formulae derived above, one clearly sees the interest of the recursive estimation scheme, which is at time t to use the last wind power measurement only for updating the model parameters. In order to initialize this recursive estimation scheme, one has to define some initial value for $\boldsymbol{\tau}_{0,k}$ (thus allowing to set $\boldsymbol{\nu}_{0,k}$), ideally from an expert guess on the shape of the function used for modeling the mean-variance relationship. In parallel, the initial inverse covariance matrix $\mathbf{R}_{0,k}$ can be filled in with zero values. Obviously, such a matrix cannot be inverted as would be necessary for updating model parameters with (23). The approach to be taken then consists in using (24) for updating $\mathbf{R}_{t,k}$ only as long as $\mathbf{R}_{t,k}$ is non-invertible, and start using (23) when this stage is reached eventually.

Obtaining predictive densities and some of their quantiles

At each time step t, the forecast information available consists of set of ensemble forecasts $\hat{y}^{(j)}_{t+k|t}$ ($j = 1, \ldots, m$) of wind power for look-ahead times up to the forecast length. In parallel for each horizon k, the parameters $\hat{\boldsymbol{\nu}}_{t,k}$ of the mean-variance model related to the Gaussian kernel parameters are updated by using the newly available measurements of wind power production y_t. From $\hat{\boldsymbol{\nu}}_{t,k}$, the parameters $\hat{\boldsymbol{\tau}}_{t,k}$ to be used in equation (9) are obtained from inverse transformation through equation (11).

For a given horizon k the mean-variance model is fully specified by $\hat{\boldsymbol{\tau}}_{t,k}$. The standard deviation of each of the m kernels composing the final predictive distributions can then be calculated as a function of the forecast values of their related ensemble members. Figure 2 provides the example of such a set of kernels (dotted lines), where the ensemble forecast values are given by the various circles. The closer the ensemble prediction values are to the nominal power of the wind farm, the sharper the kernels, thus reflecting lower level of forecast uncertainty. The final predictive distribution is obtained as an equally weighted combination of the m kernels, and is represented with a bold solid

line in Figure 2. This predictive distribution reflects the spreading of ensemble prediction values: there is here a higher density of probability of power production values being in the high power range. However, the tail expanding towards lower values indicates that there is still some probability of observing lower power production. Note that the equally weighted combination of Gaussian kernels is also consistent with the plain averaging of ensemble members used for deriving point forecasts as in equation (5), since the expectation of $\hat{f}_{t+k|t}(y)$ indeed corresponds to $\hat{y}_{t+k|t}$.

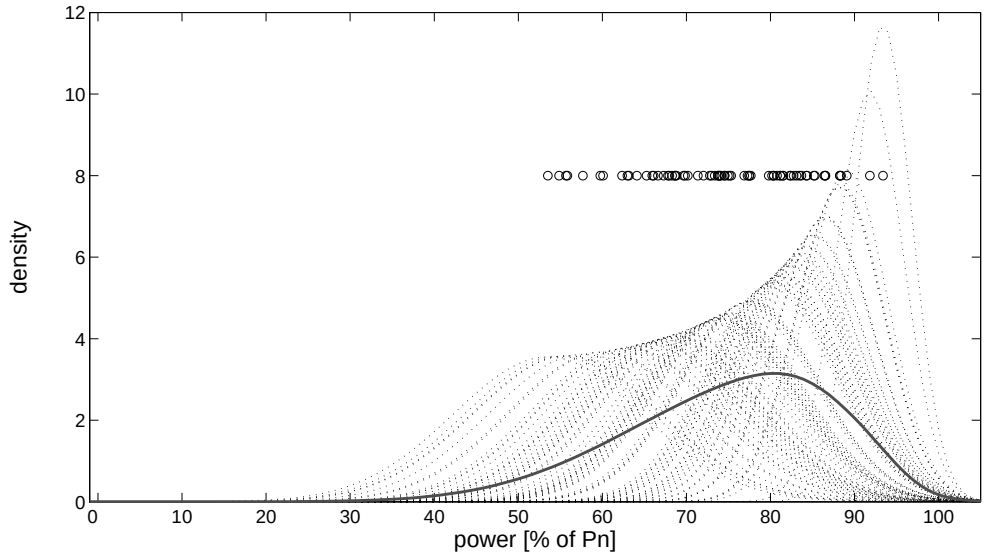


FIGURE 2: Example of a predictive distribution (for a given lead time) obtained as a weighted combination of Gaussian kernels. Dotted lines correspond to the m individual kernels, while the bold solid line gives the weighted combination. Circles are for the values of each ensemble member, thus related to the mean of each kernel.

For communication of the probabilistic forecasts, an optimal solution would be the communication of the mean and standard deviation of each of the m Gaussian kernels composing the predictive distributions for every look-ahead time. This is because predictive densities are fully defined by this set of variables. However, as this may not be optimal in terms of data storage and in terms of ready-to-use information for the forecast users, it may be preferred to communicate a number of quantiles of predictive distributions, for a set of nominal proportions. It often seems appropriate in practice to define these nominal proportions as uniformly distributed over the interval [0,1]. Quantiles of a finite mixture of Gaussian kernels as given by equation (6) can be straightforwardly computed with numerical techniques e.g. those described in [31].

From Figure 2, one may visually notice that the proposed kernel dressing of ensemble members does not constrain predictive densities in the interval [0,1] (1 corresponding to nominal capacity ofthe wind farm), even though this is the physical range of potential power production. In order to respect this physical range, one might envisage to truncate predictive distributions or simply discard parts of predictive distributions being outside [0,1]. However, it has been noticed that predicted probabilities of wind power production below 0 or above 1 were seldom, and if the case, very low. Therefore, no truncation (or discarding) has been performed, even though probabilities of event outside physical range may be counter-intuitive. It will be seen in the results Section below that this does not seem to affect the reliability of probabilistic forecasts significantly.

Application results

Below are presented the results related to the ensemble-based probabilistic forecasts of wind power at Horns Rev. A brief description of the case-study is first given, along with a description of the methodology employed for the selection of necessary parameters of the forecasting methodology described in the paper. Emphasis is then on the evaluation of the resulting probabilistic forecasts, in terms of reliability and overall skill. Complementary results related to the orthogonal fitting of the power curve model may be found in [32].

Case-study description

The Horns Rev wind farm is located off the west coast of Jutland in Denmark. It has a nominal power of 160 MW, and an annual energy yield of around 600 GWh. It represents one of the largest offshore wind farms worldwide. The original power measurement data consist of one-second measurements for each wind turbine. Focus is given to the total power output. Time series of power production are normalized by the wind farm rated capacity P_n. Power measurements or values of the different error criteria are hence all expressed in percentage of P_n. An averaging procedure has been developed in order to obtain time-series of hourly measured power averages. This hourly temporal resolution corresponds to the needs of energy actors for management and trading of wind power in Denmark, when horizons considered range from 6- to 48-hour ahead. Because there may be some erroneous or suspicious data in the raw measurements, the averaging procedure has a threshold parameter τ, which corresponds to the minimum percentage of data that need to be considered as valid in a given time interval, so that the related power average is considered as valid too. The threshold chosen is $\tau = 75\%$. The available raw data are from 16$^{\text{th}}$ February 2005 to 25$^{\text{th}}$ January 2006. From this raw data, the dataset of hourly power averages contains 73.4% of values considered as valid over this period.

The MSEPS ensemble forecasts of meteorological variables used as input have $m = 75$ members that all may be considered as relevant forecasts for the coming future. They are obtained from different, and coherent, physical parameterizations of the meteorological model employed [17]. They cover the period for which wind power production data are available. These meteorological forecasts have a forecast resolution of one hour for look-ahead times up to 48-hour ahead. They are issued every 6 hours starting from midnight. In order to increase the size of the dataset, and also in order to mimic online applications for which forecasts would be updated every hour, meteorological forecasts are iteratively slided, as if new forecasts were issued every hour. Consider for instance a meteorological forecast series issued at midnight for the following 48 hours, and with a temporal resolution of 1 hour. Then at 1:00 am, there is no new forecast provided by the meteorological service, but one may still use the previously delivered forecast series for simulating the delivery of a 'new' meteorological forecast series. In practice, it then consists of that issued at midnight, but for which the first value (i.e. that for 1:00 am) has been removed. This operation is repeated each hour until the following meteorological forecast series is really issued, that is, 6 hours after. Such artificial increase of the size of the dataset certainly has an effect on forecast quality, in the form of a small reduction of average forecast accuracy. However, it does not change the fundamental properties of the distributions of wind power generation that are aimed to be estimated. A consequence of such operation is a reduction of the forecast length to 43-hour ahead. The dataset obtained includes 8200 forecast series. The main meteorological variables that are of interest for modeling the power curve at Horns Rev are wind speed and direction. As output of the MSEPS ensemble prediction system, forecasts of these variables are available at several heights. It has been chosen to concentrate on upper heights — that is, heights closer to hub height, thus preferring to consider as input meteorological forecasts for 105 meters above sea level.

Several tries have been performed in order to integrate wind direction in polar coordinates in the model. However, since no significant difference has been noticed regarding both the shape of obtained power curves and the quality of resulting predictions, the results presented here rely on a function that model the conversion of wind speed to power only. Such an aspect should be further investigated in the future, since the wake effects inside the wind farm should certainly influence the wind farm power curve as a function of the prevailing wind direction.

Selection of the parameters of the forecasting approach

For the modeling of the power curve at Horns Rev, the orthogonal fitting approach is employed. The power variable is locally approximated with local linear regression, i.e. with polynomials of degree 1. This is owing to inherent restrictions of the method introduced in [18]. A number of 25 fitting points is chosen, spanning the whole range of potential wind speed values, and in regard of their distributions. In other words, each bin formed by two consecutive fitting points contains 5% of the wind speed values contained in the available dataset. It is then necessary to decide on the quantity of data to be used locally for fitting the model. The method employed relies on a nearest-neighbor bandwidth, for which a percentage α^{pc} of available data around each fitting point should be used for updating model coefficients, see e.g. [22]. In addition, the orthogonal fitting method is time-adaptive thanks to the use of exponential forgetting. The level of forgetting is controlled by a user-defined parameter λ^{pc}, $\lambda^{\text{pc}} \in (0, 1]$. In order to decide on optimal values of the 2 parameters λ^{pc} and α^{pc}, the available dataset is split into two parts: the first 1000 data series are considered as a learning part, on which decision is made on the optimal $(\lambda^{\text{pc}}, \alpha^{\text{pc}})$ combination, while the remaining 7200 forecast series are used for out-of sample evaluation of forecast accuracy, and for application of the adaptive kernel dressing method. Over the learning period, the first 400 data points are disregarded since considered as a batch initialization period, while the following 600 ones are used for one-fold cross-validation. The $(\lambda^{\text{pc}}, \alpha^{\text{pc}})$ combination that permits to minimize a Normalized Root Mean Square Error (NRMSE) criterion of wind power point forecasts over the cross-validation period is then employed for the whole dataset. For more information on cross-validation and its interest in statistical parameter selection, we refer to [33]. The $(\lambda^{\text{pc}}, \alpha^{\text{pc}})$ combinations obtained is (0.995,0.37). The significantly high value of the parameter α^{pc} in comparison with values commonly found for LS fitting of power curve models is due to inherent differences in the fitting method themselves, see discussion in [18].

The method for transformation of ensemble forecasts of wind power into predictive distributions is then applied to the dataset resulting from conversion of meteorological ensembles to ensemble forecasts of wind power. Out of the 7200 available series of ensemble forecasts of wind power, the first 1500 are used as a batch learning period and for optimal decision on the forgetting factor λ. Even though a different mean-variance model is defined for each look-ahead, only one forgetting factor is considered, with the underlying assumption that the slow variations in the process characteristics are of similar pace for all look-ahead times. The decision on an optimal value for λ is made in a similar fashion than for the ensemble forecasts dealt with above, that is, by employing a cross-validation technique. The last 1000 forecast series of the learning period comprise the cross-validation set, and are thus considered for evaluating which value of λ yields the maximum overall skill of the predictive distributions. Overall skill is measured with the ignorance score described in [27]: the lower the ignorance score value, the higher the skill of probabilistic forecasts. For a given predictive distribution $\hat{f}_{t+k|t}(y)$ and corresponding wind power measurement y_{t+k}, the ignorance score value $\beta_{t,k}$ is given by

$$\beta_{t,k} \;=\; -\ln(\hat{f}_{t+k|t}(y_{t+k})) \tag{30}$$

so that ignorance can be averaged over the evaluation set, and calculated separately for each hori-

zon,

$$\bar{\beta}_k \;=\; -\frac{1}{N}\sum_{t=1}^{N}\ln(\hat{f}_{t+k|t}(y_t)) \tag{31}$$

or additionally averaged over all horizons,

$$\bar{\bar{\beta}} \;=\; -\frac{1}{Nk_{\max}}\sum_{t=1}^{N}\sum_{k=1}^{k_{\max}}\ln(\hat{f}_{t+k|t}(y_t)) \tag{32}$$

where N and $k_{\max}$ stands for the number of forecast series and their number of look-ahead times, respectively. The optimal value of the forgetting factor is chosen as that which minimizes $\bar{\bar{\beta}}$ over the validation set. It is found as $\lambda = 0.995$. This value is then employed for the whole dataset and for the models defined for all look-ahead times. The probabilistic forecasts obtained over the last 5700 forecast series (the evaluation set), are then evaluated as would be done for genuine operational predictions. Regarding the mean-variance model parameters, they are initialized to

$$\boldsymbol{\tau}_{0,k}{}^{\top} \;=\; [0.1 \;\; 0.7], \quad \forall k$$

from expert knowledge on the expected level of forecast uncertainty depending on the level of predicted wind power. In parallel, the upper bounds for the mean-variance model parameters are set to

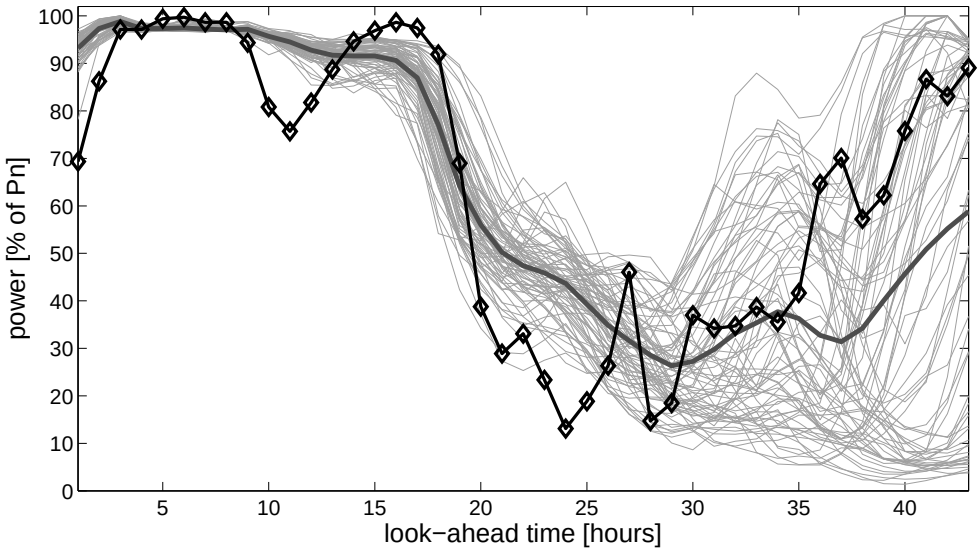
$$\tau^s_{0,k} \;=\; 0.5, \quad \tau^s_{1,k} \;=\; 2, \quad \forall k$$

thus reflecting the fact that the minimum standard deviation of the kernels may not be superior to 0.5, and that it appears unlikely that their maximum standard deviation (in the steep part of the power curve) would be above 1.

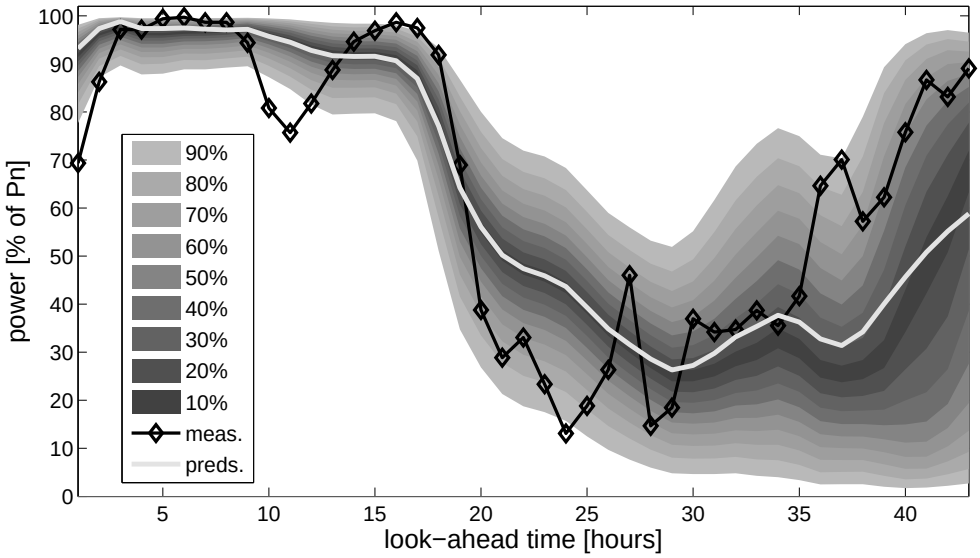
Results and discussion

As an illustration of the output of the forecasting methodology introduced in the paper, Figure 3 depicts an episode with both the raw ensemble forecasts of wind power (Figure 3(a)), and the corresponding probabilistic forecasts obtained after adaptive kernel dressing of the ensemble members (Figure 3(b)). The date and time of the day for these forecasts and related measurements are not indicated, for confidentiality reasons. For this episode, as it is case for the remainder of the evaluation set, predictive distributions are summarized by 19 quantiles with nominal proportions ranging from 0.05 to 0.95 with a 0.05 increment. Predictive distributions in Figure 3(b) are represented as a fan chart, i.e. as a set of prediction intervals with increasing nominal coverage and fading color. Prediction intervals have their bounds defined by quantiles with nominal proportions that are symmetric (in probability) with respect to the median. Figure 3 also gives the mean of ensemble members, which is introduced above as the best point forecast that can be derived from the set of ensemble members if not having information on potential superiority of certain ensemble members over the others. One may notice from both Figures that predictive distributions for each look-ahead time are not symmetric, and also not centered on the derived point predictions. This comprises a simple illustration of the non-Gaussianity of predictive distributions even though defined as a weighted combination of Gaussian kernels. A result of this non-Gaussianity and skewness of predictive distributions is that the point predictions, corresponding to the mean of predictive distributions for each look-ahead time, differ from the median of such distributions. Aspects related to non-Gaussianity of probabilistic forecasts of wind power are extensively discussed in [13].

As explained above, a first objective of the kernel dressing of ensemble predictions is to increase



(a) Ensemble forecasts of wind power, compared with measurements (diamond markers). The bold line is the mean of ensemble members, which corresponds to the point forecasts that could be derived from the ensembles.



(b) Probabilistic forecasts obtained after adaptive kernel dressing of the ensemble members, in the form of a fan chart. The point forecasts (light-colored bold line) and measurements (diamond markers) for this period are also given (diamond markers).

FIGURE 3: Episode with ensemble forecasts of wind power and the corresponding probabilistic forecasts obtained after adaptive kernel dressing of the ensemble members, along with measurements over the period.

the reliability of the probabilistic information that may be extracted from the raw ensembles. For verifying that this objective is met, the reliability of both raw ensemble forecasts of wind power and of the predictive distributions obtained after kernel dressing is evaluated with reliability diagrams, as presented in [13]. Since reliability evaluation results have been found to be qualitatively similar for all look-ahead times, they are presented and commented on for two look-ahead times only. The corresponding results are depicted in Figure 4, for the 12- and 24-ahead forecast horizons. In both cases, reliability is depicted as a deviation from the perfect reliability case, as a function of the

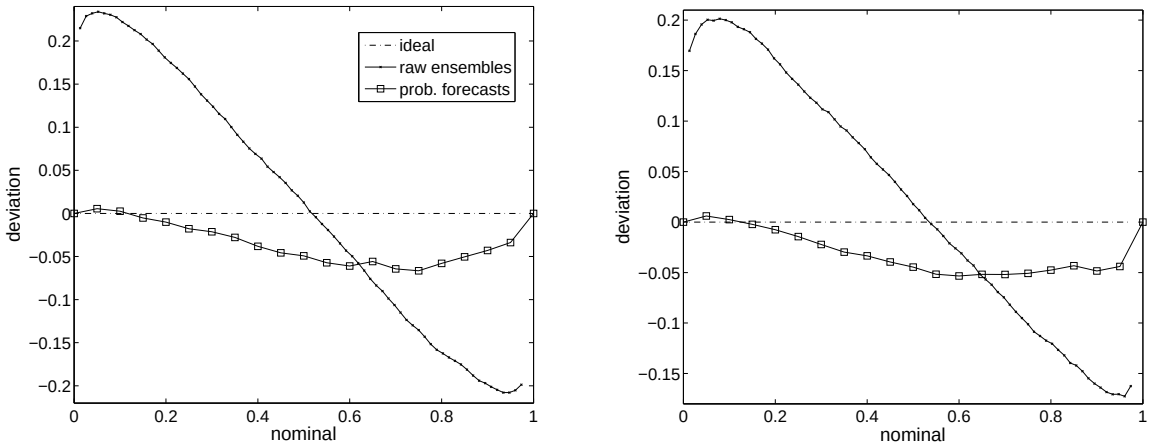
nominal proportions of quantile forecasts. The deviation from perfect reliability can be straightforwardly calculated as the difference between observed and nominal proportions of the quantile forecasts that define predictive distributions. For the case of raw ensemble forecasts, since it has been considered that ensemble members cannot be distinguished, predictive distributions are defined by sorting ensemble members in ascending order, and to consider them as quantile forecasts with increasing nominal proportions $j/(m+1)$ ($j = 1, \ldots, m$). Note that the perfect reliability case in Figure 4 can be assimilated to the case of having climatology-based probabilistic forecasts, based on all available wind power measurements in the dataset. Climatology-based probabilistic forecasts consist of a single predictive distribution obtained from the density of wind power measurements at the wind farm, and that would be used whatever the look-ahead time, time of the year or meteorological conditions. Even though this probabilistic forecasting method would not have a significant operational value, it has the advantage of being perfectly reliable and of having no resolution, thus offering an ideal benchmark.

For the case of both look-ahead times in Figures 4(a) and 4(b), the reliability of raw ensemble forecasts of wind power is far from perfect. Quantiles with low nominal proportions are clearly overestimated, while those with high nominal proportions are in contrast underestimated. Corresponding deviations from perfect reliability are here up to $\pm 25\%$. This translates to saying that ensemble forecasts of wind power are underdispersive, i.e. that their spread does not reflect the whole range of potential outcomes. And this, even though the orthogonally-fitted power curve model is less compressed and a better approximation of the power curve of the wind farm.

Underdispersivity of ensemble forecasts of wind power has also been observed by Nielsen *et al.* [34] for the case of using meteorological ensembles from ECMWF or NCEP (National Center for Environmental Prediction, in the United States) as input, with a similar range of deviations from perfect reliability. MSEPS ensembles and meteorological ensembles from ECMWF and NCEP are generated from different methods, but it seems that deviations for perfect reliability of resulting wind power ensemble forecasts are common to these various methods. This then confirms that on a general basis one should expect a significant lack of reliability of ensemble forecasts of wind power if no statistical recalibration is used. Reasons for this lack of dispersion may include the fact that the power curve model may still not be ideal, and more certainly the lack of dispersion of ensemble forecasts of meteorological variables. Underdispersivity of meteorological ensemble forecasts may in turn originates from misrepresentation of the uncertainty in the initial state of the atmosphere, or in the error growth induced by inappropriate representation of the atmosphere dynamics in the model employed. The contribution of both of these aspects in underdispersivity of meteorological ensemble forecasts is still subject to discussion, see for instance [35] and references therein.

Adaptive kernel dressing permits here to get closer to perfect reliability. Note that due to sampling effects, even if ensemble-based probabilistic forecasts were reliable, the curve resulting from their reliability evaluation would not lie along the line $y = 0$ [36, 37]. One notices a small (but apparently systematic) underestimation of the upper quantiles of predictive distributions. This may be due to the shape chosen for kernels, i.e. Gaussian, which may be not the most appropriate one for describing uncertainty information given by each ensemble member. Future works should focus on comparing different types of kernels, in order to see how reliability results may be affected by such a choice. From a more general perspective on the results obtained, it appears that deviations from perfect reliability observed here for the Horns Rev test case over a year are comparable to that observed for ensemble-based probabilistic forecasts discussed in [16] for the case of the Nysted offshore wind farm in Denmark, which has a nominal capacity slightly larger than that of the Horns Rev wind farm, and which is located off the south cost of Zealand in the Baltic Sea. Ensemble-based probabilistic forecasts in [16] are obtained by conversion via a logistic-type power curve, and subsequent recalibration of the quantiles of predictive distributions with an appropriate conditional parametric model. An advantage of the method proposed here though is its flexibility and potential for future developments e.g. by releasing the constraint on equal weighting of ensemble members,

extension of the mean-variance models to other moments, changes in the shape of the Kernels, etc.



(a) $k = 12$ (b) $k = 24$

FIGURE 4: Evaluation of the reliability of both raw ensemble forecasts of wind power and obtained predictive densities, for two look-ahead times. Reliability is assessed in terms of deviation from perfect reliability, which could be for instance met if producing probabilistic forecasts based on climatology.

Even though the aim of adaptive kernel dressing is to increase the reliability of ensemble forecasts of wind power when converting them into predictive distributions, one remembers that this is done by aiming at optimizing the overall skill of predictive distributions. Overall skill consists of a compromise between reliability and resolution. Therefore, not observing perfect reliability as it is the case here is actually not a surprise. Over the evaluation set, the skill of predictive distribution is assessed by calculating ignorance score values. A different value of the ignorance score is calculated for each forecast horizon (i.e. with $\bar{\beta}_k$, defined by (31)), as a different model has been set up for each of them, and because it is expected that overall skill would decrease as the lead time increases. Results from this overall skill assessment are depicted in Figure 5.

When using the ignorance score, an interesting reference value is 0. This is because if having no knowledge at all on past and potential future power production at the wind farm, the safest probabilistic forecast would simplify to a uniform distribution, i.e. to assuming that any power production is equally probable. Whatever the outcome, the ignorance score value of a uniform predictive distribution is 0. Then, if a predictive distribution is more informative than a uniform distribution, the ignorance score value is negative, and decreases as the content of information increases. Inversely, if the ignorance score value is above zero, this means that issued predictive distributions are actually worse than a simple uniform density guess. The second reference value that is of interest when evaluating predictive distributions is that of climatology-based probabilistic forecasts. Indeed, as stated above, such probabilistic forecasts are perfectly reliable, but they have no resolution. As a consequence, a lower ignorance score value for ensemble-based predictive distributions would demonstrate their superior skill originating from their higher resolution (and hopefully their acceptable level of reliability).

In the case of Figure 5, one observes that the ignorance score values of climatology-based predictive distributions are just below 0, while those for ensemble-based ones are significantly lower. Even though climatology based predictive forecasts may be perfectly reliable, they do not contain much more information than uniform predictive distributions, i.e. than purely random guesses on potential wind power production. In addition, the overall skill of climatology-based predictive distributions is constant whatever the look-ahead time, as it is always the same predictive distribution that is used. In contrast, the overall skill of ensemble-based predictive distributions diminishes as the

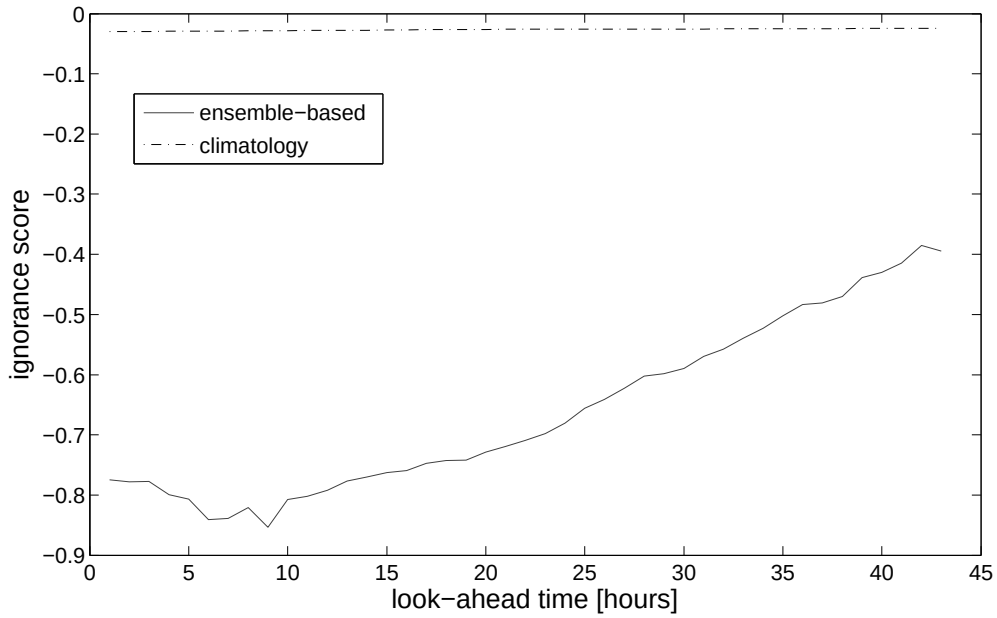


FIGURE 5: Evaluation of the overall skill of predictive densities of wind power, compared to that of predictive densities obtained from climatology. Even though climatology-based probabilistic forecasts are perfectly reliable, their overall skill is dramatically lower than that of ensemble-based probabilistic forecasts, since having no resolution.

look-ahead time increases (since ignorance score values gets closer to 0), while staying at a similar level for look-ahead times between 1- and 10-hour ahead. However over the whole forecast length considered, the overall skill of ensemble-based predictive distributions is dramatically higher, thus showing the additional resolution of probabilistic forecasts originating from the methodology proposed in the present paper.

Conclusions

Ensemble and probabilistic forecasting of wind power production, as well as the development of forecasting methodologies specially dedicated to offshore conditions, are currently two crucial research and development areas. This is due to their potential impacts on easing the integration of large amount of offshore wind power into the existing electricity grids. Even for the case of the classical point forecasting of wind power, very few results exist about the actual or expected operational performance of state-of-the-art forecasting systems if employed for large wind farms, and for offshore conditions. The main objective of the present paper has been to merge these aspects by describing and evaluating an ensemble-based forecasting methodology applied to the test case of the Horns Rev wind farm in Denmark. It is true that the mathematical methods described for the conversion of meteorological ensemble forecasts to ensemble forecasts of wind power, or for obtaining predictive distributions with optimized skill, are not specifically dedicated to offshore conditions. This is owing to the fact that most of the improvements in forecast accuracy (and understanding of forecast uncertainties) for offshore conditions may expectedly come from improvements in the meteorological forecasts themselves, by better accounting for thermal stability of sea-air interaction for instance. In parallel however, the method for power curve modeling may appropriately capture and account for the effects of wakes as a function of certain meteorological variables (mainly wind speed and direction). For the dataset considered, neither significant improvement in terms of forecast accuracy, nor differences in the shape of modeled power curves, have been observed when

considering the power curve model as a function of the forecast wind direction. Further application to a longer dataset may permit to better reveal its potential influence.

Regarding the proposed mathematical methods themselves, it would of particular interest to carry out further evaluation works by applying it to a variety of wind farms, and for longer time periods. Emphasis should be given to the choice of alternative kernels e.g. Beta ones, in order to assess the influence of such a choice on the reliability and resolution of obtained predictive distributions. The concept of employing a mean-variance model for the Kernel parameters may be generalized so that this model is made nonparametric for instance. This may also be extended to higher moments (i.e. skewness and kurtosis) in order to better describe the shape of kernels to be used. In parallel, since it may be possible to distinguish ensemble members, with some of them having a higher ability to explain uncertainty of wind power forecasts, it would be of particular interest in extend and generalize the kernel dressing framework introduced here. Indeed, by relaxing the constraint of equal weighting of every kernel (see equations 6 and 7), one could give more weight to certain ensemble members in the definition of predictive distributions. Each ensemble member could have its own (and specific) mean-variance model, reflecting its peculiar characteristics. Owing to the large number of ensemble members to be dealt with, this would then lead to an even larger number of parameters to be tracked adaptively, resulting in complex estimation problem. This may be performed in a BMA framework or by modeling the superiority of certain ensemble members over the others with a hidden Markov chain. Note that the proposed method should then be consistent with a combination scheme for ensemble members permitting to derive an optimal point forecast, as such an optimal point forecast is to be the expectation of the related predictive distribution. And, whatever the framework chosen, it will only be possible to apply the resulting approach for multi-model or multi-parameterization kind of ensembles, like those considered in the present paper. Ideally, for pure ensembles e.g. those generated by ECMWF, it should not be possible to deem certain members as superior to the others. Finally, another path towards improvement of the method proposed here consists of modifying the objective function to be adaptively minimized. Indeed, minimizing the objective function introduced here translates to aiming at maximizing the overall skill of predictive distributions, without considering the share of reliability and resolution to that overall skill. Defining an objective function in the form of a weighted decomposition into reliability and resolution would then permit to give more or less emphasis to each of these aspects depending on the intended application.

For the evaluation of ensemble-based predictive distributions of wind generation, another benchmark should also be considered in the future, which consists of purely statistical methods for probabilistic forecasting of wind generation. This will permit to discuss and potentially demonstrate the higher resolution of ensemble-based probabilistic forecasts in comparison with those based on purely statistical methods. It is actually still an open question if, for the power application, probabilistic forecasts based on meteorological ensemble predictions may be of higher quality (and potentially have a higher value) than those derived from purely statistical methods that clearly capture today the specificities of wind power forecast uncertainty.

Acknowledgments

The results presented have been generated as a part of the project 'High-resolution ensemble forecasting at Horns Rev' sponsored by the Danish Public Service Obligation (PSO) fund (under contract 2006-1-6387), which is hereby greatly acknowledged. DONG Energy Generation A/S and Vattenfall Danmark A/S are acknowledged for providing the wind power data, and Weprog A/S for providing the ensemble predictions of meteorological variables, for the Horns Rev offshore wind farm. The authors are finally grateful to two anonymous reviewers and to an editor for their helpful comments and suggestions.

References

1. Sweet W. Reap the wild wind. *IEEE Spectrum* 2002; **39**:34–39.
2. Akhmatov V. Influence of wind direction on intense power fluctuations in large offshore wind farms in the North Sea. *Wind Engineering* 2007; **31**:59–64.
3. Pinson P, Christensen LE, Madsen H, Sørensen PE, Donovan MH, Jensen LE. Regime-switching modeling of the fluctuations of offshore wind generation. *Journal of Wind Engineering and Industrial Aerodynamics* 2008, in press.
4. Sørensen PE, Cutululis NA, Vigueras-Rodriguez A, Jensen LE, Hjerrild J, Donovan MH, Madsen H. Power fluctuations from large wind farms. *IEEE Transactions on Power Systems* 2007; **22**:958–965.
5. Sørensen PE, Cutululis NA, Vigueras-Rodriguez A, Madsen H, Pinson P, Jensen LE, Hjerrild J, Donovan MH. Modelling of power fluctuations from large offshore wind farms. *Wind Energy* 2007; **11**:29–43.
6. Tambke J, Lange M, Focken U, Wolff JO, Bye JAT. Forecasting offshore wind speeds above the North Sea. *Wind Energy* 2005; **8**:3–16.
7. Tambke J, von Bremen L, Saleck N, Grwe U, Poppinga C, Claveri L, Lange M, Focken U, Bye JAT, Wolff JO. Accuracy of short-term predictions for 25 GW offshore wind power in Germany. *Proceedings of the OWEMES'06 Conference, 'Offshore Wind and other Marines Renewable Energies in Mediterranean and European Seas'*, Civitavecchia, Italy, 2006.
8. Pinson P, Ranchin R, Kariniotakis G. Short-term wind power prediction for offshore wind farms - Evaluation of fuzzy-neural network based models. *Proceedings of the Global Wind Power Conference 2004, Chicago, Illinois (USA)*, 2004.
9. Henderson AR, Morgan C, Smith B, Sørensen HC, Barthelmie RJ, Boesmans B. Offshore wind energy in Europe - A review of the state-of-the-art. *Wind Energy* 2003; **6**:35–52.
10. Bremnes JB. A comparison of a few statistical models for making quantile wind power forecasts. *Wind Energy* 2006; **9**:3–11.
11. Gneiting T, Larson K, Westrick K, Genton MG, Aldrich E. Calibrated probabilistic forecasting at the stateline wind energy center - The regime-switching space-time method. *Journal of the American Statistical Association* 2006; **101**:968–979.
12. Møller JK, Nielsen HAa, Madsen H. Time-adaptive quantile regression. *Computational Statistics and Data Analysis* 2008; **52**:1292–1303.
13. Pinson P, Nielsen HAa, Møller JK, Madsen H, Kariniotakis G. Nonparametric probabilistic forecasts of wind power: required properties and evaluation. *Wind Energy* 2007; **10**:497–516.
14. Pinson P, Nielsen HAa, Nielsen TS, Madsen H, Kariniotakis G. Properties of interval and quantile forecasts of wind generation and their evaluation. *Proceedings of the EWEC'06, 'European Wind Energy Conference', Scientific Track*, Athens, Greece, 2006.
15. Pinson P. Estimation of the uncertainty in wind power forecasting. Ph.D. Thesis, Ecole des Mines de Paris, Paris, France, 2006.
16. 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. *Proceedings of the IEEE PMAPS'06 Conference, 'Probabilistic Methods Applied to Power Systems' Conference*, Stockholm, Sweden, 2006.
17. Moehrlen, C. Uncertainty in wind energy forecasting. Ph.D. Thesis, University College Cork, Cork, Ireland, 2004.
18. 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.
19. Palmer TN. Predicting uncertainty in forecasts of weather and climate. *Report on Progresses in Physics* 2000; **63**:71–116.
20. Murphy JM. The impact of ensemble forecasts on predictability. *Quarterly Journal of the Royal Meteorological Society* 1998; **114**:463–493.
21. Nielsen TS, Nielsen HAa, Madsen H. Prediction of wind power using time-varying coefficient functions. *Proceedings of the 15th IFAC World Congress, 'International Federation of Automatic Control'*, Barcelona, Spain, 2002.

22. 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.
23. Bröcker J, Smith LA. From ensemble forecasts to predictive distributions. *Tellus* 2008; **60**: 663–678.
24. Raftery AE, Gneiting T, Balabdaoui F, Polakowski M. Using bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review* 2005; **133**:1155–1174.
25. Scott DW. *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley:New York, 1992.
26. Lange M. On the uncertainty of wind power predictions - Analysis of the forecast accuracy and statistical distribution of errors. *Transactions of the ASME - Journal of Solar Energy Engineering* 2005; **127**:177–184.
27. Roulston M, Smith LA. Evaluating probabilistic forecasts using information theory. *Monthly Weather Review* 2002; **130**:1653–1660 (Notes and Correspondence).
28. Gneiting T, Balabdaoui F, Raftery AE. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society Series B* 2007; **69**:243–268.
29. Bröcker J, Smith LA. Scoring probabilistic forecasts: on the importance of being proper. *Weather and Forecasting* 2007; **22**:382–388.
30. Madsen H. *Time Series Analysis*, Chapman & Hall/CRC: London, 2007.
31. Rahman M, Rahman R, Pearson LR. Quantiles for finite mixtures of Normal distributions. *International Journal of Mathematical Education in Science and Technology* 2006; **37**:352–357.
32. Moehrlen C, Jørgensen J, Pinson P, Madsen H, Kristoffersen JR. HRensembleHR - High Resolution Ensemble for Horns Rev. *Proceedings of the EOW'07 Conference, 'European Offshore Wind Conference and Exhibition'*, Berlin, Germany, 2007.
33. Stone M. Cross-validation and assessment of statistical predictions (with discussion). *Journal of the Royal Statistical Society Series B* 1974; **36**:111–147.
34. Nielsen HAa, Madsen H, Nielsen TS, Badger J, Giebel G, Landberg L, Sattler K, Feddersen H. Wind power ensemble forecasting using wind speed and direction ensembles from ECMWF or NCEP. Technical Report, PSO project FU-2101: 'Ensemble forecasts for wind power', Informatics and Mathematical Modeling, Technical University of Denmark, Denmark, 2005.
35. Leutbecher M, Palmer TN. Ensemble forecasting. *Journal of Computational Physics* 2008; **227**:3515–3539.
36. Bröcker J, Smith LA. Increasing the reliability of reliability diagrams. *Weather and Forecasting* 2007; **22**:651–661.
37. Toth Z, Talagrand O, Candille G, Zhu Y. Probability and ensemble forecasts. In *Forecast verification: A Practitioner's Guide in Atmospheric Sciences*. Jolliffe IT, Stephenson DB. (eds). Wiley & Sons: New York, 2003; 137–164 .