Downloaded from orbit.dtu.dk on: Nov 07, 2026

DTU Library

Probabilistic forecasts of wind power generation accounting for geographically dispersed information

Tastu, Julija; Pinson, Pierre; Trombe, Pierre-Julien ; Madsen, Henrik

Published in:
IEEE Transactions on Smart Grid

Link to article, DOI:
10.1109/TSG.2013.2277585

Publication date:
2014

Link back to DTU Orbit

Citation (APA):
Tastu, J., Pinson, P., Trombe, P.-J., & Madsen, H. (2014). Probabilistic forecasts of wind power generation accounting for geographically dispersed information. *IEEE Transactions on Smart Grid*, *5*(1), 480-489. https://doi.org/10.1109/TSG.2013.2277585

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

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

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

Probabilistic forecasts of wind power generation accounting for geographically dispersed information

Julija Tastu, Pierre Pinson, *Senior Member, IEEE*, Pierre-Julien Trombe, and Henrik Madsen

***Abstract*—Forecasts of wind power generation in their probabilistic form are a necessary input to decision-making problems for reliable and economic power systems operations in a smart grid context. Thanks to the wealth of spatially distributed data, also of high temporal resolution, such forecasts may be optimized by accounting for spatio-temporal effects that are so far merely considered. The way these effects may be included in relevant models is described for the case of both parametric and non-parametric approaches to generating probabilistic forecasts. The resulting predictions are evaluated on the real-world test case of a large offshore wind farm in Denmark (Nysted, 165 MW), where a portfolio of 19 other wind farms is seen as a set of geographically distributed sensors, for lead times between 15 minutes and 8 hours. Forecast improvements are shown to mainly come from the spatio-temporal correction of the first order moments of predictive densities. The best performing approach, based on adaptive quantile regression, using spatially corrected point forecasts as input, consistently outperforms the state-of-the-art benchmark based on local information only, by 1.5%-4.6%, depending upon the lead time.**

***Index Terms*—renewable energy, prediction, decision-making, power systems operations, offshore.**

I. Introduction

ONE OF THE underlying challenges in implementing smart grid concepts is the efficient integration of renewable energy sources, especially wind energy [1]. Indeed, the stochastic nature of wind power, with its variability and limited predictability, induces difficulties in operating and managing power grids, particularly for balancing electricity consumption and production [2]. Today, the development of advanced wind power prediction systems is considered as one of the most cost-effective solutions for mitigating the impact of the uncertainty stemming from the integration of wind power into power grids. In a recent survey of grid operators' views on wind power integration, 94% of the respondents indicated that the integration of a significant amount of wind power will ultimately depend on the accuracy of wind power forecasts [3]. A history of the short term wind power forecasting and an overview of the state-of-the-art methodology are given in [4] and [5], respectively.

Today the main interest turns from point to probabilistic forecasting [6]. This is driven by the complexity of the related decision making tasks which calls for the forecast uncertainty quantification. For example, when trading wind power on the Danish electricity market, one deals with a non-symmetric penalty function as overproduction and underproduction are not penalized in the same way, when settled through the balancing market. Therefore, in order to bid optimally it is not sufficient to know the expected power generation only. One should also be informed about the possibilities of actual energy production to exceed or to be less than the expected value, hence allowing minimizing expected balancing costs [7]. Other applications of probabilistic forecasts to power grid operations include economic load dispatch and stochastic unit commitment [8]–[10], optimal operation of storage [11], reserve quantification [12] and assessment of operating costs [13].

More generally speaking, the benefits of probabilistic forecasts can be justified by the fact that for a large class of decision making problems the optimal solution is directly linked to a specific quantile rather than the expectation of the future outcome [14].

For continuous variables like wind power generation, probabilistic forecasts take the form of predictive density functions, fully describing that random variable for a set of lead times.

Up to now the number of studies on probabilistic wind power forecasting is relatively small compared with point forecasting. A part of the available studies focus on indirect wind power forecasts, i.e. when firstly wind speed predictive densities are obtained and secondly they are transformed to wind power forecasts. Most often idealized deterministic power curve is used for such transformation [15], [16]. However in practice power curves are stochastic and of a rather complex nature [17]. To account for it, stochastic power curve models can be built, as for example in [18]. Another possible approach (which is also considered in this work) is to construct predictive densities for wind power directly, without the intermediate step of modelling the uncertainty of the wind. Advantages of this approach are (i) no need to directly account for the complexity of the stochastic power curve, (ii) owing to the geographical distribution of wind farms, the corresponding wind power data contains substantially more information than numerical weather predictions or 3 hourly data coming from the few available meteorological stations.

For constructing direct wind power density forecasts, one could follow two different families of approaches: a *parametric* or a *non-parametric* one. By the parametric approach we refer to a distribution-based methodology, which requires an assumption on the shape of the predictive densities. An example can be found in Ref. [19]. By the non-parametric one we refer to the distribution-free techniques, i.e. to the ones that are based on estimating the predictive densities directly from the data, without any constraints on the shape of the resulting distribution. As an example, adaptive resampling [20], time-

J. Tastu (juvl@dtu.dk), P.-J. Trombe (pmtr@dtu.dk) and H. Madsen (hmad@dtu.dk) are with the Department of Applied Mathematics and Computer Science at Technical University of Denmark, Kgs. Lyngby, Denmark. P. Pinson (ppin@dtu.dk) is with the Department of Electrical Engineering at Technical University of Denmark, Kgs. Lyngby, Denmark.

adaptive quantile regression [21] and time-adaptive quantile-copula [22] techniques were recently described, their evaluation suggesting that they have a similar performance level [22], [23].

Our objective here is to introduce and evaluate a methodology allowing to issue probabilistic wind power forecasts optimally accounting for geographically dispersed information. The methods are tailored to targeting a single site of interest while using a number of neighbouring sites as explanatory variables. Focus is placed on time-adaptivity in order to reduce the computational load, and also to allow for smooth variations in the process dynamics, as induced by seasonal effects for instance [24]. The forecasting methodology is evaluated on the test case of a portfolio with 20 wind farms in Denmark, where the offshore wind farm at Nysted (165MW) is the target one, while the others are used as sensors. Predictions have a temporal resolution of 15 minutes with lead times up to 8 hours ahead.

The remainder of the paper is structured as follows. Section II introduces the new challenges and opportunities related to wind power forecasting in a smart grid context. Section III describes the two-step procedures for generating probabilistic forecasts with high temporal resolutions and optimally accounting for geographically distributed information, in both parametric and non-parametric frameworks. The data used for the empirical study is presented in Section IV, while the results obtained are subsequently discussed in Section V. The paper ends with concluding remarks in Section VI.

II. Wind power forecasting in a smart grid context

The evolution from traditional power systems operation to smart grid concepts has two major implications for wind power forecasting applications.

First, it is expected to enhance the way information collected by utilities is used in operational practice [1], with a transition towards higher frequencies for power generation scheduling, from hours to minutes, potentially reducing reserve requirements [25]. This translates to new challenges for the prediction of wind power generation at high temporal resolutions (in the order of few minutes), as recently addressed by [19] (and references therein). These challenges stem from *(i)* the concentration of wind turbines within relatively small geographical areas, hence magnifying power fluctuations, and *(ii)* the lack of intra-hour information in traditional forecasts based on Numerical Weather Prediction (NWP) models. This has for instance led to the consideration of new meteorological observations from remote sensing (e.g., satellite or weather radar images), available at high spatio-temporal resolutions and to be integrated into prediction systems [26].

Second, the evolution towards smart grids is likely to result in an increased volume of available information in space and in time, on both generation and consumption sides [1]. Owing to their geographical dispersion, wind farms comprise a dense network of atmospheric sensors capable of capturing valuable information on the spatio-temporal propagation of meteorological systems, and thereby on the propagation of wind power forecast errors [27], [28]. Historically, most state-of-the art wind power prediction systems are optimized and run locally, for a single region, site or wind farm of interest, using on-site information only (e.g., meteorological forecasts, historical measurements of wind power) [24]. Yet, a new trend in wind power forecasting consists in exploiting spatio-temporal correlations in wind (power) data collected from neighbouring sites and integrating off-site information into prediction systems. The potential gains in terms of forecast accuracy have been underlined and quantified in a number of recent studies.

These studies can be divided into two groups. The first group considers cases for which dominant meteorological conditions are known a priori and the models are designed accordingly. To this group one can assign Ref. [29], [30] where the authors consider situations with one dominant wind direction, and Ref. [15], [31] designed for the situations with a strong channelling effect (considering two dominant directions). The second group does not rely on any dominant, known in advance weather patterns. Intead, the considered models are designed to capture the corresponding effects directly from the data. To this group one can assign Ref. [27], [32]–[35].

In view of this evolving context and of the limitations with existing approaches and input data in wind power forecasting, a path towards improving short-term predictability in wind power generation, and with high temporal resolution, may come from combining information from different sources, different locations and potentially different temporal resolutions. The methodology described in the following can be seen as part of this ongoing effort for making better use of available data and information in a smart grid context.

III. Methodology

The objective of the methodology introduced here is to generate probabilistic forecasts of wind power generation accounting for geographically dispersed information, which are to be of higher quality than forecasts produced based on local information only.

The main idea is based on answering the following question: if one has a snapshot of forecast errors currently (or previously) observed at the number of reference sites, then how does this information translate to the situation at the target location at time $t + k$. Thus, the approach is tailored to a situation with one target location and a number (possibly very small) of reference sites. We do not intend to describe a full space-time covariance structure for the error propagation. Instead, we focus on a rather pragmatic approach, which is focused on capturing as much of information available at present time t as possible and translating it to the situation at time $t + k$.

Depending on the layout of considered wind farms and specificities of the motion of weather systems over the considered territory, the optimal amount of information for explaining a situation at the target location at future time $t + k$ can be obtained from the errors observed at present time t or some past time $t - h$ at the reference sites. In other words,

if the reference sites are rather remote and it takes longer than k for the information to propagate from the reference to the target point, then a snapshot of the past errors $(t-h)$ should be used as explanatory variables. If on average it takes less than k for the information to propagate, one should use the corresponding snapshot taken at time t. Preliminary data analysis (for example, cross-correlation analysis of the forecast errors) can be used to get a hint on the average speed of error propagation over the territory [28]. Further in this work we focus on the case with $h=0$, i.e we use the latest available information as explanatory variables. This is motivated by our will to ease the notation and the fact that this setup was optimal for the considered test case.

The proposed procedure follows two main steps. First, the original single-valued predictions (also referred to as point forecasts) are corrected by integrating off-site information. Subsequently, these are upgraded to full probabilistic forecasts in the form of predictive densities, also allowing for off-site information to shape these predictive densities. Both parametric and non-parametric approaches are described: the former one is based on censored Gaussian distributions, while latter one relyies on time-adaptive quantile regression.

A. Parametric predictive densities

Some initial considerations are to direct our choice for relevant predictive densities. Indeed when normalized by the nominal capacity of the turbine, farm or portfolio of interest, wind power generation is double-bounded between 0 and 1. Also, the non-linear and sigmoid-shaped conversion from wind speed to power results in conditional heteroskedasticity, i.e., in a non-constant variability of wind power generation [36]. Finally, when considering single wind farms rather than aggregated territories, wind power generation may equal 0 and 1 with a non-zero probability. This results in a non-negligible concentration of probability mass at the bounds. These aspects can be also seen in Fig. 1: when the expected power is far from the natural generation bounds (Fig. 1(b)), the conditional histogram resembles that of a Gaussian distribution (a characteristic bell-shape around the expected value can be seen). The closer to the bounds, the less dispersed distributions become and the higher the probability concentration at the closest bound can be noted. Predictive densities must be able to account for these specificities.

Various proposals for density functions were made in the literature, including the generalized Logit-Normal, Censored Normal and Beta distributions compared in [19], also considered in the present work. Since the best results were obtained with Censored Normal distributions, only these are introduced and discussed in the following.

1) Censored Normal distribution: Wind power generation as a Censored Normal (CN) variable follows an ordinary Normal law within the open unit interval. However, since the values outside $[0,1]$ cannot be taken, the tails of the Normal distribution are cut and converted to probability masses at the corresponding bounds (0 and 1, respectively). Formally, a CN predictive density for wind power generation p_t at time t is defined as

$$f_p(y;\mu_t,\sigma_t^2) = w_t^0\delta_0(y) + f^{(0,1)}(y;\mu_t,\sigma_t^2) + w_t^1\delta_1(y), \quad (1)$$

where $y \in [0,1]$, δ_0 and δ_1 are Dirac functions at 0 and 1, respectively, while w_t^0 and w_t^1 are the weights representing probability mass concentration at the bounds. These are given by

$$w_t^0 = \Phi(\frac{-\mu_t}{\sqrt{\sigma_t^2}}); \quad w_t^1 = 1-\Phi(\frac{1-\mu_t}{\sqrt{\sigma_t^2}}), \quad (2)$$

with $\Phi(.)$ denoting the standard Normal distribution function. In parallel, $f^{(0,1)}(x;\mu_t,\sigma_t^2)$ follows a Gaussian density function within the open unit interval (0,1) and equals 0 outside this interval,

$$f^{(0,1)}(y;\mu_t,\sigma_t^2) = \begin{cases} (2\pi\sigma_t^2)^{-\frac{1}{2}}\exp\{-\frac{(y-\mu_t)^2}{2\sigma_t^2}\}, & y \in (0,1) \\ 0, & \text{otherwise} \end{cases}. \quad (3)$$

CN predictive densities as in Eq. (1) can be fully characterized by their location, μ, and scale, σ^2, parameters which correspond to the mean and variance of the latent Gaussian process. These may be well approximated by the mean and variance of the censored process in practice, since the corresponding bias is found to be of a limited magnitude [19]. Both parameters are here predicted employing conditional parametric models. Therefore, a short presentation of a generic conditional parametric models is given below, followed by the specifics of the models for μ and σ^2.

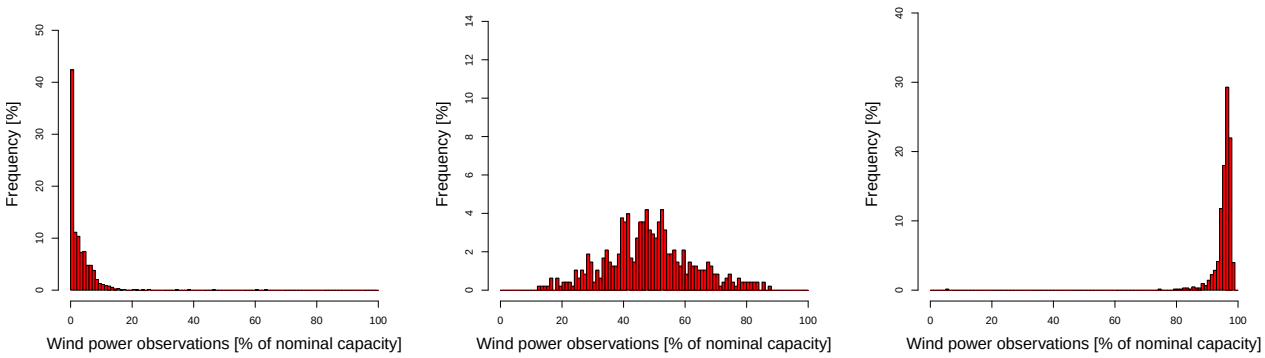
2) Generic conditional parametric models: A generic conditional parametric model reads

$$y_t = \boldsymbol{\theta}^\top(z_t)\mathbf{x}_t + \epsilon_t, \quad (4)$$

where y_t is the value for the response variable at time t, $\mathbf{x}_t = [x_{1,t}, x_{2,t}, \cdots, x_{l,t}]^\top$ and z_t are two groups of explanatory variables. $\boldsymbol{\theta}(.) = [\theta_1(.), \theta_2(.), \cdots, \theta_l(.)]^\top$ is a vector of coefficient functions to be estimated and ϵ_t is a noise term. The estimation of $\boldsymbol{\theta}(.)$ can be performed in an adaptive recursive manner as presented in Ref. [37], to which the reader is referred to for more details. Adaptivity in parameter estimation reduces computational costs significantly, hence comprising an essential element for operational implementation.

Briefly, the estimation of $\boldsymbol{\theta}(.)$ is carried out in a semi-parametric way, i.e., without imposing any particular shape for the coefficient functions. The only assumption is that these are smooth enough to be locally approximated by constants (or polynomials in a more general setting). The estimation problem then boils down to estimating those local constants (polynomials) at a number m of fitting points $z_{(j)}, j=1,\ldots,m$. This is done by estimating linear models at each of these fitting points, hence yielding local estimates of $\boldsymbol{\theta}(z_{(j)})$. After the local coefficients are estimated, the values of coefficient functions at any given point z_t can be obtained by interpolation techniques. In this work we considered linear interpolation. From our experience if a sufficient number of fitting points is considered, linear interpolation is sufficient and the increase of model complexity by considering splines is unnecessary.

The number of fitting points, the forgetting factor and the bandwidth can be chosen empirically as the values optimizing the performance of the resulting model.



(a) Expected power in 2 - 4 % of nominal capacity (b) Expected power in 48 - 50 % of nominal capacity (c) Expected power in 96 - 98 % of nominal capacity

Fig. 1. Distribution of the observed power conditional to different levels of the expected power generation given by the state-of-the-art forecasting system, WPPT. 1 hour ahead forecasts are considered. Note that the range of frequencies on the y-axis varies from one plot to the next.

3) Models for the location parameter: The location parameter of CN predictive densities is given by wind power point forecasts $\hat{p}_{t|t-k}$ issued at time $t-k$ for time t, k being the lead time,

$$\hat{\mu}_t = \hat{p}_{t|t-k}. \tag{5}$$

Two alternatives for obtaining $\hat{p}_{t|t-k}$ are considered, which are:

1) *Local* forecasts, $\hat{p}_{t|t-k} = \tilde{p}_{t|t-k}$, are obtained using a traditional wind power forecasting tool. Thus, they are optimized with respect to local information, only, and do not account for information available at neighbouring sites;
2) *Spatio-temporal* forecasts, $\hat{p}_{t|t-k} = \check{p}_{t|t-k}$, are obtained by adjusting the local forecasts based on the geographically distributed information from other wind farms. This is here carried out with the method of Ref. [27], using a conditional parametric model for tracking spatio-temporal dependencies, as rapidly described hereafter.

We suppose that the local forecast error made at time t at the target location, $\xi_t = p_t - \tilde{p}_{t|t-k}$, depends on the errors previously recorded at a set I of neighbouring sites, $\xi^{(i)}_{t-k}$, $i \in I$. This dependency is assumed to be governed by the forecasted wind direction as

$$\xi_t = \boldsymbol{\theta}^\top(\hat{w}_{t|t-k})\mathbf{x}_t + \epsilon_t, \tag{6}$$

with $\hat{w}_{t|t-k}$ denoting the wind direction predicted at time $t-k$ for time t. It is in practice given by the global average of the wind field forecast over the considered territory. In parallel, $\mathbf{x}_t$ is a column vector of the lagged local forecast errors at the set of neighbouring locations, $\xi^{(i)}_{t-k}$, $i \in I$.

Estimation of coefficient functions $\boldsymbol{\theta}$ in model (6) is performed as for the generic model (4). We have used 9 fitting points in the estimation procedure: from 0° to 320° with increments of 40°. Once the estimates of $\boldsymbol{\theta}$ are obtained, one can then correct the local wind power forecasts $\tilde{p}_{t|t-k}$ with

$$\check{p}_{t|t-k} = \tilde{p}_{t|t-k} + \hat{\boldsymbol{\theta}}^\top_{t-k}(\hat{w}_{t|t-k})\mathbf{x}_t. \tag{7}$$

4) Models for the scale parameter: The scale parameter σ^2_t is approximated by the conditional variance of wind power generation. It is estimated by modelling squared residuals ε^2_t:

$$\varepsilon^2_t = (p_t - \hat{p}_{t|t-k})^2, \tag{8}$$

where, as explained in the above, $\hat{p}_{t|t-k} = \tilde{p}_{t|t-k}$ or $\hat{p}_{t|t-k} = \check{p}_{t|t-k}$, for the case of local and spatio-temporal forecasts, respectively. The variance is consequently given by the expectation of these squared residuals, for which relevant models are to be proposed.

The volatility of wind power generation is not constant in time, owing to evolving dynamics in the wind itself, but also owing to the power curve that amplifies or dampens wind fluctuations in a nonlinear manner. The former aspect could for instance be accounted for with regime-switching models [33]. This was not done here since they were not found to improve the skill of the resulting probabilistic forecasts. In parallel, the effect of the power curve can be accommodated by letting model parameters vary with the level of expected power generation, translating to conditioning wind power volatility on the slope of the power curve. For that purpose, the so-called Conditional Parametric ARCH (CP-ARCH) model is introduced, as well as its CP-ARCHX extension when adding geographically distributed information. An extensive comparison of these various approaches and models (ARCH, GARCH and regime-switching models) can be found in Ref. [38].

a) CP-ARCH – using local information only: The CP-ARCH model can be formulated as

$$\begin{aligned} \varepsilon_t &= \sigma_t r_t, \\ \sigma^2_t &= \alpha_0(\hat{p}_{t|t-k}) + \alpha_1(\hat{p}_{t|t-k})\varepsilon^2_{t-k}, \end{aligned} \tag{9}$$

where both point forecasts $\hat{p}_{t|t-k}$ and squared residuals ε^2_{t-k} can again relate to local or spatio-temporal forecasts, r_t is a noise term and α_0, α_1 are coefficient functions to be estimated. This model somewhat states that the conditional variance of wind power generation σ^2_t, at the target location and at time t, is a function of the previously observed forecast errors at that location, ε^2_{t-k}, only. The model coefficients α_0 and α_1 are made a function of the expected power generation, as given by

the point forecasts. This is where the impact of the non-linear shape of the underlying power curve is accounted for.

The model of Eq. (9) is essentially a conditional parametric AR model for ε_t^2, which can be rewritten as

$$\varepsilon_t^2 = \alpha_0(\hat{p}_{t|t-k}) + \alpha_1(\hat{p}_{t|t-k})\varepsilon_{t-k}^2 + r_t. \quad (10)$$

Consequently, the estimation of α_0 and α_1 is performed as for the generic conditional parametric model (4), more precisely by setting $y_t = \varepsilon_t^2$, $\boldsymbol{\theta} = [\alpha_0 \ \alpha_1]^\top$, $\mathbf{x}_t = \left[1 \ \varepsilon_{t-k}^2\right]^\top$ and $z_t = \hat{p}_{t|t-k}$. We have used 20 fitting points in the estimation procedure: from 0% to 100% quantile of wind power generation with increments of 5% (excluding median). The initial selection of fitting points was done arbitrary. Further attempts to increase the number of discretization points did not result in model improvements. Thus, provided that the computational cost was acceptable (see SectionV-B), we used the initial setup.

A forecast $\hat{\sigma}_t^2$ of the scale parameter σ_t^2, issued at time $t-k$ for time t, is finally given by

$$\hat{\sigma}_t^2 = \hat{\alpha}_{0,t-k}(\hat{p}_{t|t-k}) + \hat{\alpha}_{1,t-k}(\hat{p}_{t|t-k})\varepsilon_{t-k}^2. \quad (11)$$

b) CP-ARCHX – accounting for geographically distributed information: Additional spatio-temporal effects can be included into the CP-ARCH model by extending it to a CP-ARCHX one, yielding

$$\begin{aligned} &\varepsilon_t = \sigma_t r_t, \\ &\sigma_t^2 = \alpha_0(\hat{p}_{t|t-k}) + \alpha_1(\hat{p}_{t|t-k})\varepsilon_{t-k}^2 + \gamma(\hat{p}_{t|t-k})\xi_{t-k}^{(j)2}, \end{aligned} \quad (12)$$

where, in addition to the variables and parameters of the CP-ARCH model (9), $\xi_{t-k}^{(j)2}$ denotes squared errors from the local point forecasts at the j^{th} neighbouring site. The estimation of α_0, α_1 and γ is similar to the case of generic conditional parametric models, by setting $y_t = \varepsilon_t^2$, $\boldsymbol{\theta} = [\alpha_0 \ \alpha_1 \ \gamma]^\top$, $\mathbf{x}_t = \left[1 \ \varepsilon_{t-k}^2 \ \xi_{t-k}^{(j)2}\right]^\top$ and $z_t = \hat{p}_{t|t-k}$.

Finally, a forecast for the scale parameter of CN predictive densities, issued at time $t-k$ for time t, is given by

$$\begin{aligned} \hat{\sigma}_t^2 \quad = \quad & \hat{\alpha}_{0,t-k}(\hat{p}_{t|t-k}) + \hat{\alpha}_{1,t-k}(\hat{p}_{t|t-k})\varepsilon_{t-k}^2 \\ & + \hat{\gamma}_{t-k}(\hat{p}_{t|t-k})\xi_{t-k}^{(j)2}. \end{aligned} \quad (13)$$

B. Non-parametric predictive densities

A non-parametric approach to issuing predictive densities of wind power generation does not rely on any assumption of a known distribution for the data. Instead, it suggests predicting a set of m quantiles $q_t^{(\tau)}$ and reconstructing full cumulative distribution function based on interpolation techniques. This gives more flexibility, though at a cost, since requiring to setup and estimate m models, resulting in a larger number of parameters, while still only partly describing densities. The quantiles defining the predictive densities are quantile forecasts for pre-defined nominal proportions, here obtained in a quantile regression framework.

1) Generic quantile regression models: The most basic form of a quantile regression model for a response y_t and a nominal proportion τ, as introduced in Ref. [39], is

$$q_t^{(\tau)} = \mathcal{F}_{y_t}^{-1}(\tau|\mathbf{x}_t) = \boldsymbol{\beta}^\top \mathbf{x_t} + r_t, \quad (14)$$

where $\mathcal{F}_{y_t}^{-1}$ is the inverse distribution function of y_t, $\mathbf{x_t}$ is a column vector of explanatory variables, $\boldsymbol{\beta}$ is a vector of parameters to be estimated and r_t is a noise term. Given a set of N observations on which the model is to be fitted, estimates of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$ are obtained by solving the following linear programming problem:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{t=1}^{N} \left(y_t - \boldsymbol{\beta}^\top \mathbf{x}_t\right)\left(\tau - \mathbf{1}(y_t < \boldsymbol{\beta}^\top \mathbf{x}_t)\right), \quad (15)$$

where $\mathbf{1}(.)$ is an indicator function, equal to 1 if the condition between brackets realizes, and to 0 otherwise.

The optimization problem (15) can be solved in a time-adaptive fashion by applying the method described in Ref. [21]. Briefly, it consists in updating the dataset used for estimation in a sensible way each time new data points become available, so as to minimize computational costs while allowing for smooth time variations in the model parameters and still covering the whole range of variations for the various explanatory variables.

Finally based on the parameters for the quantile regression model estimated based on past observations, a forecast for $q_t^{(\tau)}$, issued at time $t-k$, can be obtained with

$$\hat{q}_{t|t-k}^{(\tau)} = \hat{\boldsymbol{\beta}}^\top \mathbf{x}_t \quad (16)$$

By having a bank of m quantile regression models with, say, $\tau = j/(m+1)$, $j = 1, \cdots, m$, and then issuing quantile forecasts for these various nominal proportions, full predictive distributions are constructed.

2) Quantile regression models for wind power generation: Building non-parametric predictive densities for wind power generation by using quantile regression is performed in two steps. First, wind power point predictions, $\hat{p}_{t|t-k}$ are used to determine the mean of the corresponding predictive distributions (see Section III-A3). Second, uncertainty around the mean is shaped, thus upgrading point forecasts to full predictive densities. This is done by estimating a conditional distribution of the point forecast errors at the target location (ε_t) and adding it to the estimate of the expected power generation, i.e.,

$$\mathcal{F}_{p_t} = \hat{p}_{t|t-k} + \mathcal{F}_{\epsilon_t}, \quad (17)$$

where $\mathcal{F}_{p_t}$ and $\mathcal{F}_{\epsilon_t}$ stand for the cumulative density functions of p_t and ε_t, respectively.

When defining quantile models for the distribution of the forecast errors, it is essential to account for some of the important characteristics of the process. Here namely, the uncertainty is known to be shaped by the power curve [36]. This dependence is known to be non-linear and thus, the quantile models we study here are given by

$$q_t^{(\tau)} = \mathcal{F}_{\varepsilon_t}^{-1}(\tau|\hat{p}_{t|t-k}) = g(\hat{p}_{t|t-k}) + r_t. \quad (18)$$

No particular shape is imposed for the function g. Instead, we estimate it in a non-parametric way, using a spline representation. In other words, it is assumed that g can be viewed as a piecewise cubic function of $\hat{p}_{t|t-k}$ and, thus, can be

expressed as a linear combination of the known basis functions b_j, resulting in the following quantile models

$$q_t^{(\tau)} = \beta_0 + \sum_{j=1}^{K-1} b_j(\hat{p}_{t|t-k})\beta_j + r_t, \quad (19)$$

where b_j are natural cubic B-spline basis functions, K is the number of knots used for the spline construction and $\beta_i(\tau)$ are coefficients to be estimated. Such spline representation of g permits to use the estimation techniques valid for linear models such as that in Eq. (15), by setting $y_t = \varepsilon_t$, $\mathbf{x}_t = \left[1 \;\; b_1(\hat{p}_{t|t-k}) \;\; \ldots \;\; b_{K-1}(\hat{p}_{t|t-k})\right]^\top$ and $\boldsymbol{\beta} = [\beta_0 \;\; \beta_1 \;\; \ldots \;\; \beta_{K-1}]^\top$. Model (19) is referred to as QR in the following.

In that setup, additional spatio-temporal effects captured by point forecast errors previously recorded at the j^{th} sensor location may be accounted for. This translates to having $\xi_{t-k}^{(j)}$ as a supplementary explanatory variable in the QR model, as a linear or nonlinear term (represented by splines). Since the spline-based representation did not result in any improvement, we focus on the simpler linear case. The resulting model is denoted by QR-X and is written as

$$\mathcal{F}_{\varepsilon_t}^{-1}(\tau|\hat{p}_{t|t-k}) = g(\hat{p}_{t|t-k}) + \gamma\xi_{t-k}^{(j)} + r_t. \quad (20)$$

Estimation is similar to the case of model (15), by setting $y_t = \varepsilon_t$, $\mathbf{x}_t = \left[1 \;\; b_1(\hat{p}_{t|t-k}) \;\; \ldots \;\; b_{K-1}(\hat{p}_{t|t-k}) \;\; \xi_{t-k}^{(j)}\right]^\top$ and $\boldsymbol{\beta} = [\beta_0 \;\; \beta_1 \;\; \ldots \beta_{K-1} \;\; \gamma]^\top$.

Quantile regression is used to provide 18 quantile forecasts with nominal proportions going from 5% to 95% by 5% increments, except for the median. 0% and 100% quantiles are set to 0 and 1, respectively. The setup corresponds to the one used in [23]. Linear interpolation is used to reconstruct full distribution functions from the set of quantiles. From our experience if a sufficient number of quantiles is considered, the linear interpolation is sufficient and the increase of complexity by considering splines is unnecessary. Empirical study could be performed in order to check whether an increase in the number of fitting points improves the performance of the models.

IV. Data

The data used in this study were provided for 20 wind farms located in Denmark. All wind farms are owned and operated by the same power company. The respective locations of these wind farms are shown in Fig. 2. For each wind farm, the following information is available:

- Wind power measurements at a temporal resolution of 15 minutes. They are normalized by the respective nominal capacities P_n of the various wind farms;
- Point forecasts of wind power generation, with lead times from 0 to 48 hours, and temporal resolution of 15 minutes. These predictions were generated with the Wind Power Prediction Tool (WPPT) [24], which is one of the state-of-the-art prediction models for the short-term wind power forecasting as discussed in Ref. [5];
- Meteorological forecasts of wind speed and wind direction at 10 meters above ground level, with lead times from 1 to 48 hours, and temporal resolution of 1 hour. These forecasts were generated by the HIRLAM model operated by the Danish Meteorological Institute (DMI) [40].

The data covers a period from May 1, 2008 to December 31, 2009. A first part of the data from May 1, 2008 to December 31, 2008 was used as a burn-in period in order to allow time-adaptive parameters not to be influenced anymore by their initial values. Forecast evaluation was carried out over the remainder of the dataset, from January 1, 2009 to December 31, 2009. Due to the large number of missing values, the effective evaluation period was eventually consisting of approximately 8.5 months (more than 25.000 forecast series for each of the lead times considered).

The Nysted wind farm was chosen as the target wind farm in this study. Nysted is located offshore on the Rødsand sand bank, near Lolland, Denmark, and is the southernmost of all wind farms shown in Fig. 2. There are two main reasons behind this choice. Firstly, Nysted was the largest wind farm in Eastern Denmark until 2010, with a rated capacity of 165MW, and therefore was one of the main contributors to the aggregated amplitude of wind power fluctuations in that region. It also accounted for about 36% of the installed capacity owned by the company operating it. And secondly, Nysted has an appealing location with many wind farms located "upwind" in view of the prevalence of westerly flows over Denmark [41]. Indeed, improvements in forecast accuracy resulting from the use of off-site information are expected to be larger for wind farms located "downwind", as shown by Ref. [27] for instance.

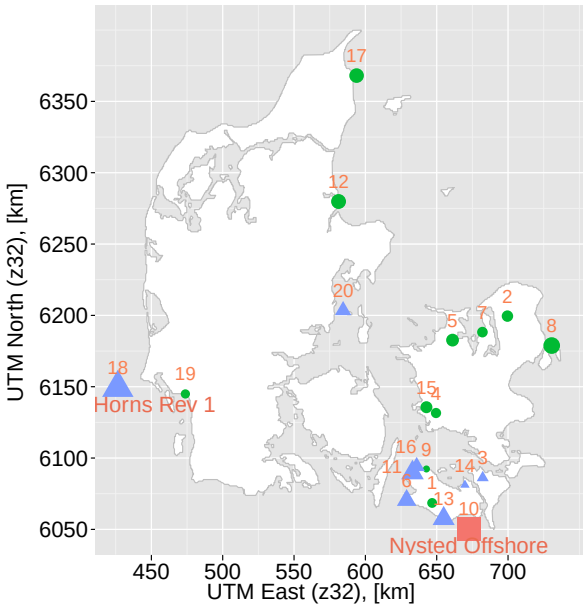


Fig. 2. Map of the 20 wind farms included in the analysis. The Nysted wind farm is marked as number 10 and with a red square. Information from the wind farms with blue triangles contributed to improve the predictability of wind power generation at Nysted. Information from the wind farms with green bullet points was found not to improve the predictability of wind power generation at Nysted. The size of the points is proportional to the rated capacity of the wind farms, on a logarithmic scale.

Out of the 19 nearby wind farms, only 8 wind farms (number 3, 6, 11, 13, 14, 16, 18 and 20) were used as explanatory

variables. The selection was performed empirically, based on the stepwise selection.

In Ref. [28] it was found that correlations in wind power forecast errors between two wind farms were very small for distances larger than 50 km, over the Western Denmark area. Our findings are consistent with these results since most wind farms that contribute to improve wind power predictability at Nysted are located within such range (see Fig. 2). More surprisingly, despite the large distances separating Nysted from wind farms 18 and 20 (approximately 265 km and 176 km, respectively), integrating information from these last two wind farms also led to substantial gains in forecast accuracy. Both wind farms 18 and 20 are located out in the open sea as Nysted, while the remaining wind farms in this analysis are located onshore or near-shore. This difference with the results in Ref. [28] could be explained by a higher spatial persistence and homogeneity of wind field dynamics over waters than over lands, where the terrain roughness is known to be a very influential factor.

V. Empirical results

A. Model notation

The following notations are used for the model names:

- CN:CP/CP-ARCH refers to CN predictive densities with μ given by the spatio-temporal point forecasts (7) and σ^2 estimated using the CP-ARCH model (9);
- CN:CP/CP-ARCHX refers to CN predictive densities with μ given by the spatio-temporal point forecasts (7) and σ^2 estimated using the CP-ARCHX model (12);
- WPPT/QR stands for the non-parametric predictive densities based on the time-adaptive quantile regression (QR) (19). Local forecasts (WPPT) are used as input;
- CP/QR stands for the non-parametric predictive densities based on the time-adaptive quantile regression (QR) (19). Spatio-temporal forecasts given by (7) are used as input;
- CP/QR-X stands for non-parametric predictive densities with additional consideration of the spatio-temporal effects in the uncertainty modelling step as in (20). Spatio-temporal forecasts given by (7) are used as input;

B. Computational details

The parameters for the parametric densities were updated every 15 min. A single update step took less than a second of computing time. The quantile regression models were updated daily. A single update step took approximately 3 seconds when evaluating 20 different quantile models. The computations were performed on a laptop, having a processor Intel i 7-2620M CPU 2.70 GHz and the installed RAM of 8 GB

C. Overall evaluation

The evaluation and comparison of probabilistic forecasting approaches follows the guidelines, scores and diagnostic tools described in Ref. [42]. The lead score is the Continuous Ranked Probability Score (CRPS), which is a proper score for density forecasts. This score is negatively oriented: the smaller it is, the better the forecasts are.

Adaptive quantile regression with original WPPT point forecasts as input (WPPT/QR) is considered as the base benchmark. Other predictive densities are compared to the benchmark approach and the relative improvements in CRPS (skill scores) are calculated. The summary of the results is given in Fig. 3.

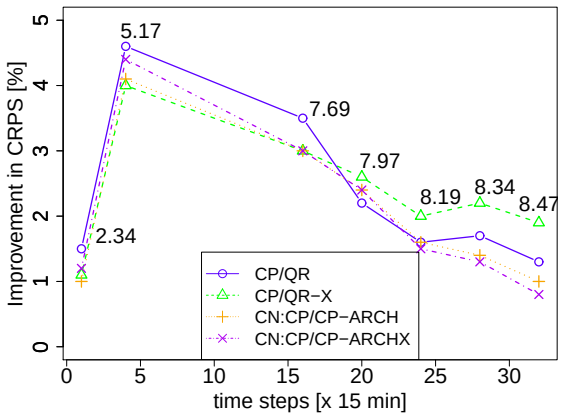


Fig. 3. Evaluation of predictive densities in terms of relative CRPS improvement, as a function of lead time. Point labels indicate the CRPS values [% of nominal capacity] for the benchmark model WPPT/QR.

The models have similar levels of CRPS improvement for all the lead times considered. This was also confirmed by using the Diebold-Mariano test statistics t_n [43], which may be more generally used to test for equal performance of probabilistic forecasts. t_n is asymptotically standard Normal under the null hypothesis of vanishing score differentials. This hints at the fact that forecast improvements brought in by the proposed methodologies mainly come from the space-time correction of the point forecasts defining the conditional mean of predictive densities. Further consideration of space-time dynamics in the modelling of uncertainty around the mean does not seem to bring additional benefit. This is confirmed by the fact that CP/QRX does not outperform CP/QR, and similarly CN:CP/CP-ARCHX does not outperform CN:CP/CP-ARCH.

Consequently, we further focus on the evaluation of the CP/QR and CN:CP/CP-ARCH predictive densities, since comprising the best performing non-parametric and parametric probabilistic forecasts, respectively. Both types of probabilistic forecasts outperform the benchmark approach given by WPPT/QR (see Fig. 3). The statistical significance of those improvements has been verified using the Diebold-Mariano test statistics t_n. The corresponding results are given in Table I.

The non-parametric densities accounting for the space-time dynamics (CP/QR) show statistically significant improvements for lead time up to 8 hours ahead. Such an observation is consistent with the spatio-temporal scales for the inertia of weather systems passing over Denmark. The improvements for the parametric alternative based on CN predictive densities can be considered statistically significant up to 5 hours ahead. At the same time, it is not possible to reject the null hypothesis such that CP/QR and CN:CP/CP-ARCH perform similarly for

TABLE I
TESTING FOR EQUAL FORECAST PERFORMANCE WITH THE DIEBOLD-MARIANO TEST STATISTICS t_n. p IS THAT THE PROBABILITY THAT THE CORRESPONDING VALUE OF t_n IS ACHIEVED WITH THE NULL HYPOTHESIS BEING TRUE. '*' SYMBOLS MARK STATISTICALLY SIGNIFICANT DIFFERENCES IN CRPS.

Horizon		15 min	1 h	4 h	5 h	6 h	7 h	8 h
CP/QR vs WPPT/QR	t_n	-9.29	-12.32	-5.63	-3.43	-2.34	-2.60	-3.19
	p	0.00*	0.00*	0.00*	0.00*	0.02*	0.01*	0.00*
CN:CP/CP-ARCH vs WPPT/QR	t_n	-6.06	-10.45	-3.22	-2.28	-1.34	-1.07	-0.65
	p	0.00*	0.00*	0.00*	0.02*	0.18	0.28	0.51
CP/QR vs CN:CP/CP-ARCH	t_n	-1.60	-0.90	-0.41	-0.25	-0.00	-0.39	-1.08
	p	0.11	0.37	0.68	0.80	1.00	0.70	0.28

all lead times up to 8 hours ahead. From the 5-hour lead time, the benefits from space-time considerations start fading away. Mainly, this is caused by the fact that the benchmark approach is chosen non-parametric. Thus, CP/QR and WPPT/QR are of the same model family, while CN:CP/CP-ARCH differs from WPPT/QR in its nature, since being parametric. This results in higher variance of score differentials once comparing CN:CP/CP-ARCH and WPPT/QR, and subsequently lower t_n values.

Another interesting point to mention is that the peak in the improvements is observed for the prediction horizons of 1 hour ahead. This is in line with the layout of the considered wind farms. From the map in Fig. 2 one can see that almost all the considered reference sites are within 50 km from Nysted. According to [28], an average speed of error propagation over Denmark is 30-50 km/h (depending on the prevailing wind direction). This result is consistent with the fact that the peak of cross-correlations between Nysted and almost all the reference sites comes at lags of approximately 1 hour ahead and correspondingly results in the highest improvements.

D. Conditional evaluation

Emphasis is then placed on the situation-dependent performance of probabilistic forecasts, through a conditional forecast evaluation exercise. Results are shown and discussed for 1-hour ahead forecasts only, since these are qualitatively similar for the other lead times from 15 min to 8 hours ahead.

Firstly, since predictive densities evolve with the level of expected power generation, the overall skill of these densities is assessed based on the CRPS score, as a function of the point forecast values. The corresponding results are depicted in Fig. 4.

The CRPS for all the predictive densities considered increases for expected power levels in the medium range. This can be explained by the fact that higher uncertainties in the wind power generation are faced in periods with medium power generation, i.e., in the steep slope part of the power curve. The performances of the parametric and the non-parametric densities are similar in this predicted power range. Some differences can be noted, however, close to the generation bounds. Closer to the generation bounds, the censoring effect in the parametric densities is more present (see Figure 1), hence leading to a higher bias in parameter estimates. For the particular case of the the upper bound, that is, when expected power is close to nominal capacity, it is also that power down-regulation actions were not always

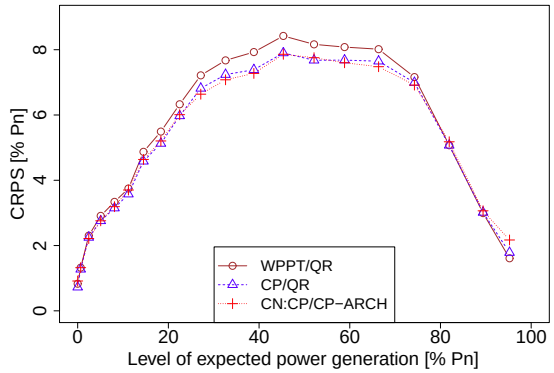


Fig. 4. CRPS conditional to the expected wind power generation (given by point forecasts) for the 1-hour lead time, with WPPT/QR considered as the benchmark. Power levels are given by 20 equally populated classes deduced from the distribution of point forecasts.

flagged and discarded from the dataset, as much as they should be, then affecting the parameter estimation and evaluation of the forecasts. Better results are therefore expected to be seen in case down-regulation actions are better dealt with when gathering wind power generation datasets in the future.

Similar differences in conditional forecast skill can be observed when assessing the skill of predictive densities conditional to actual power measurements, see Fig. 5. Parametric densities perform better during periods when observed power is not close to the generation bounds. In contrast during periods with low and high power generation, the non-parametric densities show superior results. This suggests that CN:CP/CP-ARCH forecasts have a better ability to discriminate among the observations when the power is in its medium range and a worse ability to discriminate as the power generation gets closer to its natural bounds.

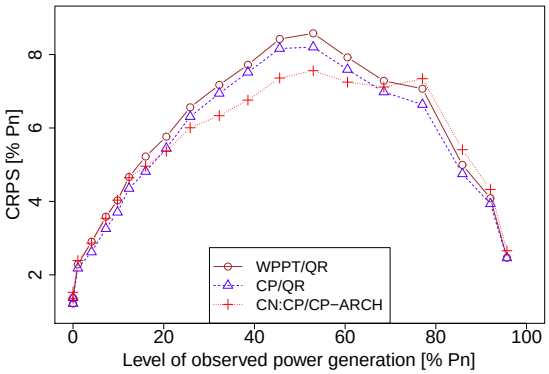


Fig. 5. CRPS conditional to the observed power generation level for the 1-hour lead time, with WPPT/QR being the benchmark. Power levels are given by 20 equally populated classes deduced from the distribution of power measurements.

As an illustration of the type of probabilistic forecasts finally obtained with the various approaches proposed here, Figure 6 gives the example of predictive densities issued by the CP/QR model, issued on the 24th of November, 2009, at 17:15, their shape evolving with the level of power and with the lead time.

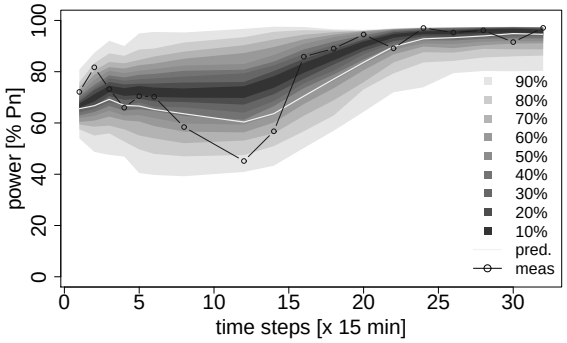


Fig. 6. Predictive densities of wind power generation from the CP/QR model, depicted by a set of prediction intervals with increasing nominal coverage rates. These forecasts were issued on the 24th of November, 2009, at 17:15, for lead times between 15 mins and 8 hours ahead.

VI. Conclusions

Focus was given to probabilistic wind power forecasting with consideration of geographically distributed information, hence permitting to capture additional space-time dynamics. The proposed methodology can be used for issuing density forecasts for a single site of interest, while using information from the other wind farms as explanatory variables.

The methodology is based on discrete formulation of the problem as opposed to proposing a full space-time covariance model which normally would call for a larger amount of reference sites spread throughout the territory. The approach chosen here is based on proposing ways to best summarize a snapshot of forecast errors observed at time t when issuing probabilistic forecasts for time $t + k$.

Two ways of constructing predictive densities were described and analysed. The parametric approach relied on CN distributions while the non-parametric one employed quantile regression techniques. All estimation methods were introduced in a time-adaptive framework in order to reduce computational costs, while allowing for long-term variations in the process dynamics, as induced by meteorological systems for instance.

The empirical results obtained on the test case of a portfolio of wind farms in Denmark show that accounting for spatio-temporal effects improves the quality of probabilistic forecasts for a range of lead times, here up to 8 hours. It was shown to be sufficient to focus on correcting the conditional expectation of wind power generation. Additional inclusion of spatio-temporal effects into the uncertainty modelling step did not significantly further improve the skill of the predictive densities. None of the proposed approaches outperformed the benchmark for lead times further than 8 hours ahead, in line with the scales of motion of weather systems over a small region such as that covered by Denmark. Thus, likely, for a different test case, depending on the geographical layout of the considered wind farms and some meteorological particularities of the area, this maximum lead time would differ.

The performance of parametric and non-parametric approaches were also compared, uncovering that they both performed similar for lead times up to 5 hours ahead, and with an advantage for non-parametric predictive densities for further lead times. Based on overall skill, the highest-quality forecasts were obtained by adaptive quantile regression with spatially corrected point predictions, with CRPS improvements between 1.5% to 4.6% depending upon the lead time. Further research on both parametric and non-parametric approaches to wind power probabilistic forecasting may challenge the comparison carried out here.

Acknowledgement

The authors gratefully acknowledge DONG Energy for the data and financial support provided through the project "Spatio-temporal correction of wind power forecasts". Additional acknowledgements are due to Tryggvi Jónsson for his help with quantile regression, as well as to two reviewers for their comments on previous version of the manuscript.

References

[1] H. Farhangi, "The path of the smart grid," *Power and Energy Magazine, IEEE*, vol. 8, pp. 18–28, 2010.

[2] T. Ackermann *et al.*, *Wind power in power systems*. Wiley Online Library, 2005, vol. 140.

[3] L. Jones and C. Clark, "Wind integration - a survey of global views of grid operators," in *Proceedings of the 10th International Workshop on Large-Scale Integration of Wind Power into Power Systems*, 2011.

[4] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa, "A review on the young history of the wind power short-term prediction," *Renewable and Sustainable Energy Reviews*, vol. 12, no. 6, pp. 1725–1744, 2008.

[5] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl, "The state-of-the-art in short-term prediction of wind power–a literature overview. 2nd edition," Technical University of Denmark, Tech. Rep., 2011, available at: http://orbit.dtu.dk.

[6] T. Gneiting, "Editorial: probabilistic forecasting," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 171, no. 2, pp. 319–321, 2008.

[7] A. Botterud, Z. Zhou, J. Wang, R. Bessa, H. Keko, J. Sumaili, and V. Miranda, "Wind power trading under uncertainty in lmp markets," *Power Systems, IEEE Transactions on*, vol. 27, no. 2, pp. 894–903, 2012.

[8] X. Liu, "Economic load dispatch constrained by wind power availability: A wait-and-see approach," *Smart Grid, IEEE Transactions on*, vol. 1, no. 3, pp. 347–355, 2010.

[9] A. Botterud, Z. Zhou, J. Wang, J. Sumaili, H. Keko, J. Mendes, R. J. Bessa, and V. Miranda, "Demand dispatch and probabilistic wind power forecasting in unit commitment and economic dispatch: A case study of illinois," *Sustainable Energy, IEEE Transactions on*, vol. 4, no. 1, pp. 250–261, 2013.

[10] C. Liu, J. Wang, A. Botterud, Y. Zhou, and A. Vyas, "Assessment of impacts of phev charging patterns on wind-thermal scheduling by stochastic unit commitment," *Smart Grid, IEEE Transactions on*, vol. 3, no. 2, pp. 675–683, 2012.

[11] Á. Duque, E. Castronuovo, I. Sánchez, and J. Usaola, "Optimal operation of a pumped-storage hydro plant that compensates the imbalances of a wind power producer," *Electric Power Systems Research*, vol. 81, no. 9, pp. 1767–1777, 2011.

[12] R. Bessa, M. Matos, I. Costa, L. Bremermann, I. Franchin, R. Pestana, N. Machado, H. Waldl, and C. Wichmann, "Reserve setting and steady-state security assessment using wind power uncertainty forecast: a case study," *IEEE Transactions on Sustainable Energy*, 2012.

[13] M. Ortega-Vazquez and D. Kirschen, "Assessing the impact of wind power generation on operating costs," *Smart Grid, IEEE Transactions on*, vol. 1, no. 3, pp. 295–301, 2010.

[14] T. Gneiting, "Quantiles as optimal point forecasts," *International Journal of Forecasting*, vol. 27, no. 2, pp. 197–207, 2011.

[15] T. Gneiting, K. Larson, K. Westrick, M. Genton, and E. Aldrich, "Calibrated probabilistic forecasting at the stateline wind energy center," *Journal of the American Statistical Association*, vol. 101, no. 475, pp. 968–979, 2006.

[16] J. W. Taylor, P. E. McSharry, and R. Buizza, "Wind power density forecasting using ensemble predictions and time series models," *Energy Conversion, IEEE Transactions on*, vol. 24, no. 3, pp. 775–782, 2009.

[17] I. Sanchez, "Short-term prediction of wind energy production," *International Journal of Forecasting*, vol. 22, no. 1, pp. 43–56, 2006.
[18] J. Jeon and J. W. Taylor, "Using conditional kernel density estimation for wind power density forecasting," *Journal of the American Statistical Association*, vol. 107, no. 497, pp. 66–79, 2012.
[19] P. Pinson, "Very-short-term probabilistic forecasting of wind power with generalized logit–normal distributions," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 2012.
[20] P. Pinson and G. Kariniotakis, "Conditional prediction intervals of wind power generation," *Power Systems, IEEE Transactions on*, vol. 25, no. 4, pp. 1845–1856, 2010.
[21] J. Møller, H. Nielsen, and H. Madsen, "Time-adaptive quantile regression," *Computational Statistics & Data Analysis*, vol. 52, no. 3, pp. 1292–1303, 2008.
[22] R. Bessa, V. Miranda, A. Botterud, Z. Zhou, and J. Wang, "Time-adaptive quantile-copula for wind power probabilistic forecasting," *Renewable Energy*, vol. 40, no. 1, pp. 29–39, 2012.
[23] P. Pinson, H. Nielsen, J. Møller, H. Madsen, and G. Kariniotakis, "Non-parametric probabilistic forecasts of wind power: required properties and evaluation," *Wind Energy*, vol. 10, no. 6, pp. 497–516, 2007.
[24] H. Nielsen, T. Nielsen, and H. Madsen, "An overview of wind power forecasts types and their use in large-scale integration of wind power," in *Proceedings of the 10th International Workshop on Large-Scale Integration of Wind Power into Power Systems*, 2011.
[25] D. Lew and R. Piwko, "Western wind and solar integration study," National Renewable Energy Laboratories, Tech. Rep., 2010.
[26] P. Trombe, P. Pinson, T. Bøvith, N. Cutululis, C. Draxl, G. Giebel, A. Hahmann, N. Jensen, B. Jensen, N. Le, H. Madsen, L. Pedersen, A. Sommer, and C. Vincent, "Weather radars - The new eyes for offshore wind farms?" *Wind Energy*, 2012.
[27] J. Tastu, P. Pinson, and H. Madsen, "Multivariate conditional parametric models for a spatiotemporal analysis of short-term wind power forecast errors," in *Scientific Proceedings of the European Wind Energy Conference, Warsaw (PL)*, 2010, pp. 77–81.
[28] R. Girard and D. Allard, "Spatio-temporal propagation of wind power prediction errors," *Wind Energy*, 2012.
[29] I. G. Damousis, M. C. Alexiadis, J. B. Theocharis, and P. S. Dokopoulos, "A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation," *Energy Conversion, IEEE Transactions on*, vol. 19, no. 2, pp. 352–361, 2004.
[30] M. Alexiadis, P. Dokopoulos, and H. Sahsamanoglou, "Wind speed and power forecasting based on spatial correlation models," *Energy Conversion, IEEE Transactions on*, vol. 14, no. 3, pp. 836–842, 1999.
[31] K. A. Larson and K. Westrick, "Short-term wind forecasting using off-site observations," *Wind energy*, vol. 9, no. 1-2, pp. 55–62, 2006.
[32] A. Hering and M. Genton, "Powering up with space-time wind forecasting," *Journal of the American Statistical Association*, vol. 105, no. 489, pp. 92–104, 2010.
[33] P.-J. Trombe and P. Pinson, "High-resolution forecasting of wind power generation with regime-switching models and off-site observations," Technical University of Denmark, Dpt. of Applied Mathematics and Computer Science, Tech. Rep., 2012, available at: http://orbit.dtu.dk.
[34] A. Lau, "Probabilistic wind power forecasts: From aggregated approach to spatio-temporal models," Ph.D. dissertation, Mathematical Institute, University of Oxford, 2011.
[35] R. Jursa and K. Rohrig, "Short-term wind power forecasting using evolutionary algorithms for the automated specification of artificial intelligence models," *International Journal of Forecasting*, vol. 24, no. 4, pp. 694–709, 2008.
[36] M. Lange, "On the uncertainty of wind power predictionsanalysis of the forecast accuracy and statistical distribution of errors," *Journal of Solar Energy Engineering*, vol. 127, p. 177, 2005.
[37] H. Nielsen, T. Nielsen, A. Joensen, H. Madsen, and J. Holst, "Tracking time-varying-coefficient functions," *International Journal of Adaptive Control and Signal Processing*, vol. 14, no. 8, pp. 813–828, 2000.
[38] J. Tastu, P. Pinson, P.-J. Trombe, and H. Madsen, "Spatio-temporal correction targeting nysted offshore – probabilistic forecasts," Technical University of Denmark, Tech. Rep., 2012.
[39] R. Koenker and G. Bassett Jr, "Regression quantiles," *Econometrica: journal of the Econometric Society*, pp. 33–50, 1978.
[40] B. Sass, N. Nielsen, J. Jørgensen, B. Amstrup, M. Kmit, and K. Mogensen, "The operational DMI-HIRLAM system 2002-version," DMI - Danmarks Meteorologiske Institut, Tech. Rep., 2002.
[41] J. Cappelen and B. Jørgensen, "Observed wind speed and direction in Denmark - with climatoligical standards normals, 1961-90," DMI - Danmarks Meteorologiske Institut, Tech. Rep., 1999.
[42] T. Gneiting and A. Raftery, "Strictly proper scoring rules, prediction, and estimation," *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 359–378, 2007.
[43] F. X. Diebold and R. S. R. S. Mariano, "Comparing predictive accuracy," *Journal of Business & Economic Statistics*, vol. 13, no. 3, pp. 253–263, 1995.

Julija Tastu received the M.Sc. degree in Applied Mathematics from Kaunas University of Technology, Lithuania, in 2007. She is currently pursuing the Ph.D. degree at the Dpt. of Applied Mathematics and Computer Science of of the Technical University of Denmark. Her research interests include among others wind power forecasting and spatio-temporal modelling.

Pierre Pinson (M'11, SM'13) received the M.Sc. degree in Applied Mathematics from the National Institute for Applied Sciences (INSA Toulouse, France) and the Ph.D. degree in Energy from Ecole des Mines de Paris. He is the Professor in Modelling of Electricity Markets at the Technical University of Denmark, Centre for electric Power and Energy, Dpt. of Electrical Engineering. His research interests include among others forecasting, uncertainty estimation, optimization under uncertainty, decision sciences, and renewable energies. He acts as an Editor for the *IEEE Transactions on Power Systems* and for *Wind Energy*.

Pierre-Julien Trombe received the M.Sc. degree in Applied Mathematics from the National Institute for Applied Sciences (INSA), Toulouse, France, in 2005, and the Ph.D. degreee from the Technical University of Denmark in 2013. He is currently a postdoctoral researcher at the Dpt. of Applied Mathematics and Computer Science of of the Technical University of Denmark. His research interests include among others forecasting, image analysis and renewable energies.

Henrik Madsen received the Ph.D. degree in statistics at the Technical University of Denmark in 1986. He is a Professor in Stochastic Dynamical Systems at the Dpt. of Applied Mathematics and Computer Science of of the Technical University of Denmark. His research interests includes analysis and modelling of stochastic dynamics systems, optimization, and control, with applications mostly related to energy systems, informatics, environmental systems, bioinformatics, biostatistics, process modelling, and finance.