ARTICLE IN PRESS
International Journal of Forecasting ▮ (▮▮▮▮) ▮▮▮–▮▮▮
Contents lists available at ScienceDirect
International Journal of Forecasting
journal homepage: www.elsevier.com/locate/ijforecast

Short-term probabilistic forecasting of wind speed using stochastic differential equations

Emil B. Iversen *, Juan M. Morales, Jan K. Møller, Henrik Madsen
Technical University of Denmark, Matematiktorvet, building 303b, DK-2800 Lyngby, Denmark

ARTICLE INFO

Keywords:
Wind speed
Probabilistic forecasting
Wind power
Stochastic differential equations

ABSTRACT

It is widely accepted today that probabilistic forecasts of wind power production constitute valuable information that can allow both wind power producers and power system operators to exploit this form of renewable energy economically, while mitigating the potential adverse effects relating to its variable and uncertain nature. In order to provide reliable wind power forecasts for periods beyond a couple of hours, forecasts of the wind speed are fundamental. In this paper, we propose a modeling framework for wind speed that is based on stochastic differential equations. We show that stochastic differential equations allow us to capture the time dependence structure of wind speed prediction errors naturally (from 1 to 24 h ahead) and, most importantly, to derive point and quantile forecasts, predictive distributions, and time-path trajectories (also referred to as scenarios or ensemble forecasts), all using one single stochastic differential equation model that is characterized by a few parameters.

1. Introduction

The last few years have witnessed a remarkable increase in the contribution of renewable energy sources to the global electricity supply, with the largest share in many countries coming from wind turbines (The European Wind Energy Association, 2013). However, wind power production is highly variable and uncertain, thus challenging traditional practices for power system operation and the trade of this form of renewable energy in electricity markets. In order to mitigate the adverse effects of the stochastic nature of wind, good forecasts of the power generated by wind farms are a must. Furthermore, to manage and exploit wind energy fully, these forecasts should provide power system operators and wind power producers not only with a single-valued guess of the future wind generation – a so-called *point forecast* – but also with information on possible outcomes and their associated probabilities of occurrence. This enriched form of forecasting is known as *probabilistic forecasting*, and takes on its full meaning in the context of wind power management and trading (Morales, Conejo, Madsen, Pinson, & Zugno, 2014; Zhou et al., 2013).

The literature on wind power forecasting is now vast, but is centered mostly on techniques for producing point predictions. For a comprehensive review of the topic, the interested reader is referred to the work of Costa et al. (2008), Monteiro et al. (2009) and Foley, Leahy, Marvuglia, and McKeogh (2012). On the other hand, methods for wind power probabilistic forecasting are not so well developed. Wind power density forecasts are commonly obtained by superimposing a model for the probability distribution of prediction errors on a point forecast (typically, the average or most likely outcome; see Bremnes, 2006; Møller, Nielsen, & Madsen, 2008; Pinson & Kariniotakis, 2010), or by post-processing ensemble forecasts from meteorological models so that they represent the true predictive density (Nielsen et al., 2006; Pinson & Madsen, 2009; Taylor, McSharry, & Buizza, 2009). In the realm of probabilistic

* Corresponding author.
E-mail address: jebi@dtu.dk (E.B. Iversen).

http://dx.doi.org/10.1016/j.ijforecast.2015.03.001

forecasting, it is customary to distinguish between *parametric* and *nonparametric* methods. The former presuppose a certain standard distribution for the forecast error, and as a result, the modeling endeavor boils down to estimating the parameters that characterize such a distribution (Messner, Zeileis, Broecker, & Mayr, 2013; Pinson, 2012; Thorarinsdottir & Gneiting, 2010; Thorarinsdottir & Johnson, 2012). In contrast,nonparametric methods do not assume any pre-specified forecast error distribution and work directly with the empirical distribution instead (Bessa, Miranda, Botterud, Zhou, & Wang, 2012; Bremnes, 2004; Messner et al., 2013; Pinson, Nielsen, Møller, Madsen, & Kariniotakis, 2007). Recent work has also focused on generating realistic sample trajectories of the stochastic process and producing multi-horizon or multivariate probabilistic forecasts. To this end, the most popular approach is to fit a marginal predictive density for each univariate output variable (e.g., the wind speed in each time period of the prediction horizon) – see Lerch and Thorarinsdottir (2013) for a comparison of different regression-type models for wind speed – and then combine these marginals into a multivariate cumulative density using copula theory (Schefzik, Thorarinsdottir, & Gneiting, 2013).

Forecasts of wind power are improved markedly by the use of numerical weather predictions (NWPs) of wind speed. Foley et al. (2012) provide an overview of the different uses of NWPs for wind power forecasting. Specifically, they underline the fact that ensemble forecasts of wind speed can be used to obtain valuable information on the reliability of the wind power forecast. The usefulness of NWPs of wind speed for wind power forecasting has also been stressed, for instance, by De Giorgi, Ficarella, and Tarantino (2011) and Ramirez-Rosado, Fernandez-Jimenez, Monteiro, Sousa, and Bessa (2009). Taylor et al. (2009) and Pinson and Madsen (2009) use predictive densities of wind speeds to produce probabilistic forecasts of wind power. These predictive densities are estimated from the ensemble forecasts provided by a NWP system.

This paper describes a novel approach to wind speed probabilistic forecasting based on stochastic differential equations (SDEs). The proposed SDE model upgrades numerical weather predictions using wind speed data and provides plausible time-path trajectories of the wind speed process that are perfectly comparable to the ensemble forecasts obtained from a NWP system. Furthermore, our SDE model can generate these trajectories for a specific wind site swiftly. For these reasons, our SDE model can contribute significantly to wind power forecasting. More generally, SDEs offer a powerful and versatile modeling framework that allows us to issue point forecasts and all forms of probabilistic forecasts (namely quantiles, densities, and time-path trajectories) consistently using the same model. Moreover, the parameters characterizing the SDE can be interpreted intuitively, which makes it much easier to formulate model extensions on the basis of the specific physics of the underlying stochastic process or of observable statistical deficiencies. The proposed modeling framework naturally captures the time dependence of forecast errors and events with zero probability, such as negative wind speeds, and does not need to assume Gaussian innovations. Seen in a broader perspective, SDEs cover the large class of stochastic processes with continuous trajectories, and, in fact, many of the discrete time models used in classical time series theory can be seen as discrete-time versions of SDEs.

The application of stochastic differential equations to forecasting, and to wind power forecasting in particular, is a very recent topic, and consequently, the technical literature in this regard is scant. However, two studies should be mentioned here, namely those of Møller, Pinson, and Madsen (2013) and Zárate-Miñano, Anghel, and Milano (2013). SDEs are used fruitfully by Møller et al. (2013) for wind power forecasting, considering state-dependent diffusions and with a numerical weather prediction as external input. The SDE model proposed by the authors is fitted simultaneously to data from 1 to 48 h ahead, which makes it computationally intensive to estimate and limits the amount of data that can be used for fitting. In contrast, the SDE model proposed in this paper is estimated on one-step-ahead data, and thus, the approach that we use here resembles that of Box–Jenkins type models that are fitted using one-step-ahead information. Furthermore, our model focuses on wind speed, not wind power, and therefore the challenges are different. Zárate-Miñano et al. (2013) present a continuous time model for wind speed, which aims to simulate wind speed trajectories over very short time horizons. However, it cannot be used for forecasting, as the parameters are not estimated, no external inputs are allowed, and the SDE model is limited to a very simple structure. In addition, the SDE model that Zárate-Miñano et al. (2013) propose is designed to fit the long-term stationary distribution of wind speed. The stationary distribution, however, does not necessarily hold for the short term, as one can see for the case of the climatological forecast.

The approach to modeling wind speed taken in this paper is similar to the approach to modeling solar irradiance adopted by Iversen, Morales, Møller, and Madsen (2014), in that they both rely on SDEs. However, the weather phenomena considered in these two papers, namely wind speed and solar irradiance, are remarkably different, and each has its own challenges. Indeed, the modeling of wind speed differs substantially from that of solar irradiance in terms of the physical domain of the underlying process and its periodic nature. This results in distinct structures for the drift and diffusion terms. Moreover, in the case of wind speed, the mere introduction of the NWP as an exogenous input to the SDE model is not advantageous, as it results in the simulated process systematically lagging behind the NWP. This will be shown later on. We solve this issue by introducing the derivative of the NWP as well. Furthermore, in the present paper, we provide a straightforward methodology for generating predictive densities and time-path trajectories of wind speeds that is analogous to the ensemble forecasts obtained from a NWP system. These ensemble forecasts are of particular relevance for wind power forecasting, whereas time-path trajectories and predictive densities for solar irradiance are still of limited use, at least comparatively speaking.

The remainder of the paper is organized as follows: Section 2 provides a short introduction to stochastic differential equations (SDEs) and the parameter estimation procedure. Section 3 provides a model for wind speeds based on stochastic differential equations. This model is subsequently validated on training and test sets in Section 4. Section 5 concludes the paper and provides directions for future research.

2. Stochastic differential equations

A stochastic differential equation (SDE) is a differential equation with one or more stochastic terms that results in a solution that is in itself a stochastic process. SDEs are used to describe various phenomena that are driven by a large random component, and are especially prominent in mathematical finance (Björk, 2009; Mikosch, 1998) and physics (Adomian, 1988; Van Kampen, 1992). We give a very short introduction to SDEs here, and refer the interested reader to Øksendal (2010) for a thorough and mathematically rigorous discussion of the topic.

A SDE is commonly stated as

$$dX_t = f(X_t, t)dt + g(X_t, t)dW_t, \tag{1}$$

where W_t is a Wiener process and the functions $f(\cdot)$ and $g(\cdot)$ are known as the drift and diffusion terms, respectively. However, Eq. (1) is not well defined, as the derivative of W_t, dW_t, does not exist. Thus, Eq. (1) should actually be seen as a shorthand for the integral equation

$$X_t = X_0 + \int_0^t f(X_s, s)ds + \int_0^t g(X_s, s,)dW_s. \tag{2}$$

Here, we use the Itō interpretation of the second integral.

In contrast to the solution of an ordinary differential equation, which consists of a single time trajectory that represents the value of the modeled process at each point in time into the future, the solution to a SDE is a stochastic process, which characterizes the uncertainty in the system dynamics for each future time. For an Itō process given by the SDE in Eq. (1) with drift $f(X_t, t)$ and diffusion coefficient $g(X_t, t) = \sqrt{2D(X_t, t)}$, the density $j(x, t)$ of the random variable X_t in state x at time t is the solution to the partial differential equation (Björk, 2009):

$$\frac{\partial}{\partial t} j(x, t) = -\frac{\partial}{\partial x}\left[f(x, t)j(x, t)\right] + \frac{\partial^2}{\partial x^2}\left[D(x, t)j(x, t)\right], \tag{3}$$

which is known as the Fokker–Planck equation or the Kolmogorov forward equation. Hence, given a specific SDE, the probability density function of the process can be found by solving a partial differential equation (PDE). In general, this PDE cannot be solved analytically, but, fortunately, a wide range of numerical solution approaches do exist.

It can be shown (see the Lévy-Itō decomposition by Björk, 2009) that, under sufficient regularity conditions, all stochastic processes with continuous trajectories can be written as special cases of SDEs. Therefore, SDEs are a general class of stochastic processes. Indeed, many ordinary time series models can be interpreted as discrete versions of SDEs.

In practice, it is only possible to observe continuous-time systems in discrete time. For this reason, one defines the observation Y_k of the process at time t_k, which is found through some measurement equation $h(\cdot)$. We thus have the following system of equations:

$$dX_t = f(X_t, t)dt + g(X_t, t)dW_t \tag{4}$$

$$Y_k = h(X_{t_k}, t_k, e_k), \tag{5}$$

where e_k is some measurement error.

The procedure that we use in this paper to estimate Eqs. (4) and (5) relies on specific forms of $g(\cdot)$ and $h(\cdot)$. In particular, we require that $g(X_t, t) = g(t)$ and $h(X_{t_k}, t_k, e_k) = h(X_{t_k}, t_k) + e_k$, where $e_k \sim \mathcal{N}(0, \sigma^2)$. In principle, even though these two constraints may limit the modeling capability of our SDE framework considerably, they can actually be relaxed to a large extent. Indeed, the first condition, which establishes that the diffusion term must not depend on the current state X_t, can be overcome by transforming the original SDE into an equivalent one with non-state-dependent diffusion using Itō calculus and the so-called *Lamperti transform* (Iversen et al., 2014; Møller et al., 2008). The second condition, which requires the observation error to be additive and Gaussian, can be mitigated by transformations of the data (Box & Cox, 1964).

The procedure for estimating the vector θ of the parameters defining the SDE model in Eqs. (4)–(5) is based on the extended Kalman filter (see for example Welch & Bishop, 1995). The filter is applied to the equivalent SDE system that results from the aforementioned transformations. The first step in the estimation procedure is to find the one-step predictions of the mean and variance of the observations, which are defined as

$$\widehat{Y}_{k|k-1} = \mathbb{E}\left[Y_k|\mathcal{Y}_{k-1}, \theta\right] \tag{6}$$

$$R_{k|k-1} = \mathbb{V}\left[Y_k|\mathcal{Y}_{k-1}, \theta\right], \tag{7}$$

where $\mathbb{E}[\cdot]$ and $\mathbb{V}[\cdot]$ denote the expectation and variance, respectively, and $\mathcal{Y}_{k-1} = \{Y_0, \ldots, Y_{k-1}\}$. We use the extended Kalman filter (Jazwinski, 2007) to determine these predictions. We can now define the innovation

$$\epsilon_k = Y_k - \widehat{Y}_{k|k-1}, \tag{8}$$

in order to compute the likelihood. For a system satisfying the conditions on $g(\cdot)$ and $h(\cdot)$ that we stated above, the approximated likelihood is given by

$$L(\theta; \mathcal{Y}_N) = \left(\prod_{k=1}^{N} \frac{\exp\left(-\frac{1}{2}\epsilon_k^\top R_{k|k-1}^{-1}\epsilon_k\right)}{\sqrt{\det\left(R_{k|k-1}\right)}\left(\sqrt{2\pi}\right)^l}\right) p(Y_0|\theta), \tag{9}$$

where l is the dimension of the sample space, that is, the dimension of Y_k, N is the number of observations, $(\cdot)^\top$ denotes the vector transpose, and $p(Y_0|\theta)$ is the likelihood of seeing observation Y_0. We cannot compute this likelihood, because there are no observations previous to Y_0. To get around this issue, we instead optimize the logarithm of the likelihood function conditional on Y_0, which results in

$$\log\left(L(\theta; \mathcal{Y}_N|Y_0)\right) = -\frac{1}{2}\sum_{k=1}^{N}\left(\log(\det(R_{k|k-1})) + \epsilon_k^\top R_{k|k-1}^{-1}\epsilon_k\right) - \log(2\pi)\frac{Nl}{2}. \tag{10}$$

The parameter vector θ enters the log-likelihood function in Eq. (10) through ϵ_k and $R_{k|k-1}$. This parameter vector can now be estimated by maximizing Eq. (10), i.e.,

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \left(\log\left(L\left(\theta; \mathcal{Y}_N|Y_0\right)\right)\right), \tag{11}$$

where Θ is the feasible parameter space. A thorough introduction to parameter estimation and filtering is provided by Jazwinski (2007), and a more detailed description of the exact implementation of the extended Kalman filter and the parameter estimation procedure employed here is given by Kristensen, Madsen, and Jørgensen (2004). We note that the likelihood function is optimized for the one-step-ahead residuals. To estimate SDE models using a multi-horizon approach, we refer the interested reader to Møller et al. (2013).

3. A SDE model for wind speed

In this section, we present a probabilistic model for wind speed. This section follows the general model-building approach presented by Iversen et al. (2014), which essentially applies general guidelines for system identification (see, e.g., Ljung, 1999; Madsen, 2008) to the context of stochastic differential equations. We start by introducing a basic SDE system that aims to track a given numerical weather prediction. We then highlight the shortcomings of this simple SDE model, and use them as a basis for justifying the different structural components of the more complex SDE model that we propose.

The data used in this study come from a meteorological station located in the western part of Denmark, and include hourly wind-speed measurements, together with predicted wind speeds based on a numerical weather prediction model from the Danish Meteorological Institute (Källén, 1996). The numerical weather prediction (NWP) provides 48-hour forecasts of the wind speed, and is updated every six hours. The data cover a period of three years, from 01/01/2009 to 31/12/2011, and are divided into two periods: a training set of two years for parameter estimation, and a test set spanning the remaining one year for evaluating the performance of the proposed SDE model.

We use the simple SDE model in Eqs. (12)–(13) below as a starting point.

$$dX_t = \theta_x(p_t\mu_x - X_t)dt + \sigma_x dW_t \tag{12}$$

$$Y_k = X_{t_k} + \epsilon_k. \tag{13}$$

Here, we aim to track the NWP supplied by the Danish Meteorological Institute. In Eqs. (12) and (13), wind speed observations are given by Y_k. We let p_t denote the numerical weather prediction at time t. The parameter μ_x is a local scaling of the numerical weather prediction, which corrects for the NWP either over- or undershooting on average. $\theta_x \geq 0$ is a time constant that governs how rapidly the model returns to the predicted wind speed. Essentially, this parameter controls the relative contributions of past observations and the NWP to the future wind speed. Indeed, a large θ_x results in a stochastic process that is driven mainly by the NWP (the process moves rapidly towards the NWP), while a small θ_x produces a stochastic process that is governed predominantly by the past observations of wind speed (the process moves slowly away from the previous wind speed value). The parameter σ_x characterizes the system noise. The data that we use are average hourly wind speeds, measured by an anemometer. We include the stochastic variable $\epsilon_k \sim \mathcal{N}(0, \sigma_\epsilon)$ in the observation equation to characterize the measurement error of the physical system.

The SDE model in Eqs. (12)–(13), however simple, suffers from grave deficiencies. First, the model assigns a positive probability to negative wind speeds, as can be seen in Fig. 1, which shows the one-step-ahead (i.e., one-hour-ahead) predictive distribution of the wind speed. This is clearly unrealistic. We will correct this obvious shortcoming by means of a state-dependent diffusion term.

Furthermore, the point forecast provided by the model in Eqs. (12)–(13) is shifted in time systematically with respect to the numerical weather prediction for horizons longer than one hour (see Fig. 2). We will resolve this issue by introducing the time derivative of the NWP as an input to our SDE model.

Finally, the autocorrelation function for the studentized residuals resulting from the model in Eqs. (12)–(13), shown in Fig. 3, reveals structures in the wind speed process (e.g., a daily pattern) that can be explained but are not being captured by this simple SDE model. Note that the y-axis in Fig. 3 has been scaled down so as to show the significance bounds better. We will address this issue by allowing for both a diurnal variation and a stochastic trend in the SDE system.

We propose the following SDE model for wind speeds:

$$\begin{aligned} dX_t = {} & \left(\left(\alpha_x \sin\left(\frac{2\pi}{24}t + \phi_x\right) + U_t + \rho_x \dot{p}_t\right)\left(1 - e^{-X_t}\right)\right. \\ & \left. + \theta_x(p_t\mu_x - X_t)\right) dt + \sigma_x X_t^{\beta_x} dW_{x,t} \end{aligned} \tag{14}$$

$$dU_t = \theta_u\left(\mu_u - U_t\right)dt + \sigma_u dW_{u,t} \tag{15}$$

$$Y_k = X_{t_k} + \epsilon_k. \tag{16}$$

The model in Eqs. (14)–(16) is characterized by the state-dependent diffusion term $\sigma_x X_t^{\beta_x} dW_{x,t}$, with the parameter $\beta_x \geq 0$ determining its shape. Furthermore, as the process X_t tends to zero, so does the diffusion term. Consequently, within the vicinity of null wind speeds, the stochastic process X_t is dominated by the drift term, which will eventually push the process away from zero. Indeed, under certain regularity conditions, related to the relative sizes of the diffusion term and the drift term, and given that p_t is always larger than zero, we find that X_t has a zero probability of taking on negative values. Note that the term $\left(1 - e^{-X_t}\right)$ guarantees that $\left(\alpha_x \sin\left(\frac{2\pi}{24}t + \phi_x\right) + U_t + \rho_x \dot{p}_t\right)$ does not force the wind speed process out of the domain $\mathbb{R}^+$, since $\left(1 - e^{-X_t}\right)$ tends to zero as X_t approaches zero.

In order to remove the time shift between the numerical weather prediction and the point forecast of the SDE model, the time derivative of p_t, i.e., $\dot{p}_t$, is introduced. Similarly to μ_x, ρ_x is a factor that scales $\dot{p}_t$. It is important to stress that the model in Eqs. (14)–(16) is still causal, because $\dot{p}_t$, just like p_t, constitutes information that is available at time t as a by-product of the numerical weather

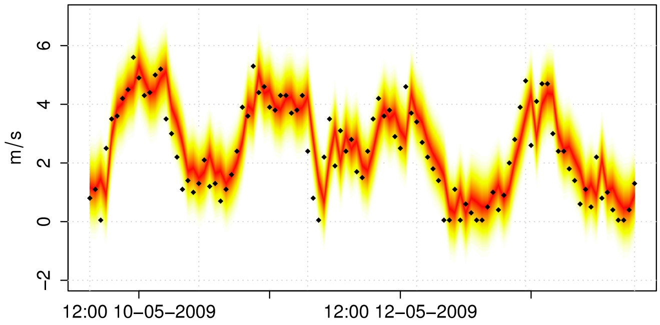

Fig. 1. One-step-ahead predictive density from the model specified by Eqs. (12) and (13), approximated by Monte Carlo simulation. Warmer colors represent a higher probability of seeing this realization. The black dots are the actual wind speed observations. (For the interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

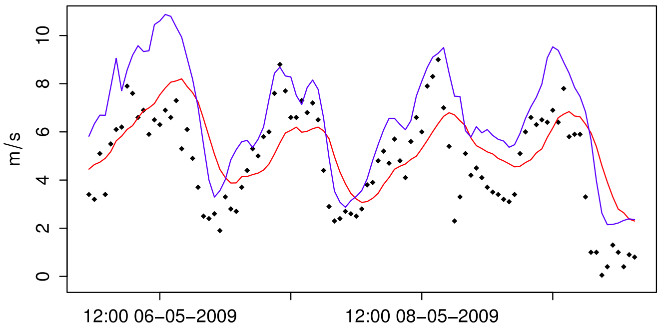

Fig. 2. The 24-hour-ahead point predictions obtained from the model defined by Eqs. (12)–(13) are shown in red. The NWP is shown in blue, with the observations in black. (For the interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

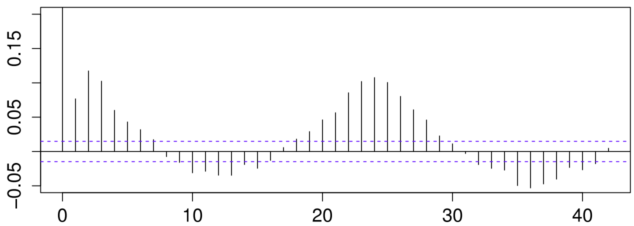

Fig. 3. The autocorrelation function for the studentized residuals of the SDE model defined by Eqs. (12)–(13).

prediction model. Therefore, $\dot{p}_t$ can be used as an input for predicting the future evolution of the stochastic process Y_t. In practice, $\dot{p}_t$ should be seen as the *increment* of the wind speed process over the next time period that is forecast by the numerical weather prediction model, insofar as the SDE model is specified in terms of increments.

Finally, U_t is a stochastic trend that captures the fact that periods of increasing (decreasing) wind speed tend to be followed by more periods where the wind speed is increasing (decreasing). In this line, μ_u describes the long-term trend of U_t, and θ_u determines how rapidly U_t reverts to this level. The parameter σ_u governs the diffusion of U_t. Eq. (14) includes a diurnal variation component, with α_x governing the amplitude and ϕ_x governing the phase. Both U_t and the diurnal variation are intended to improve on the NWP point forecast, which explains why they are introduced in the drift term.

4. Model validation

We first show that the shortcomings in the basic SDE model in Eqs. (12)–(13) are no longer present in the proposed SDE model (Eqs. (14)–(16)). In this vein, Fig. 4 displays the predictive densities obtained from the latter. Note that, except for some minor observation noise, negative wind speeds are now assigned a zero probability.

Fig. 5 shows the point predictions given by our SDE model, with the associated 95% prediction intervals. A comparison of this figure with Fig. 2 makes it clear that the time-shift problem highlighted in Section 3 is no longer an issue.

The autocorrelation function of the studentized residuals yielded by the proposed SDE model is depicted in Fig. 6. The range on the y-axis has been scaled to $(-0.05, 0.20)$ so as to illustrate the significance bands better, since they

Please cite this article in press as: Iversen, E. B., et al., Short-term probabilistic forecasting of wind speed using stochastic differential equations. International Journal of Forecasting (2015), http://dx.doi.org/10.1016/j.ijforecast.2015.03.001

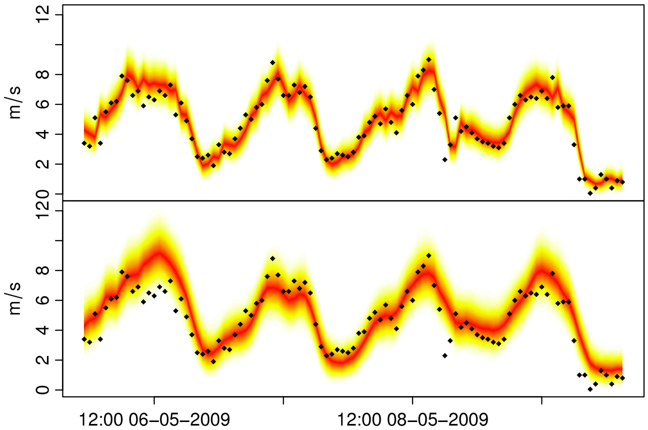

Fig. 4. 1-hour-ahead (top) and 24-hour-ahead (bottom) predictive densities of the proposed SDE model, with warmer colors indicating a higher probability of seeing this realization. The densities are approximated by Monte Carlo simulations. (For the interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

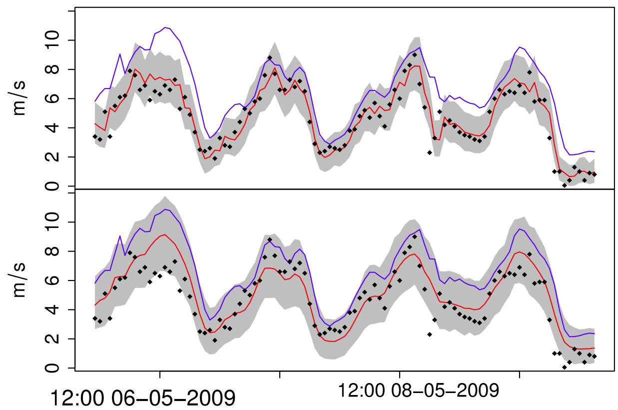

Fig. 5. 1-hour-ahead (top) and 24-hour-ahead (bottom) point predictions from the SDE model (in red), along with 95% prediction intervals (gray shaded area). These are obtained through Monte Carlo simulations. The blue line is the point prediction of the NWP. (For the interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

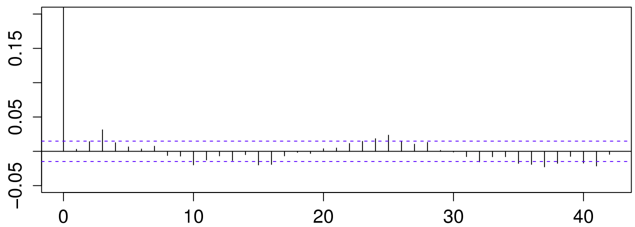

Fig. 6. The ACF of the proposed SDE model.

are quite small, because the data set is large. By comparing this figure with Fig. 3, one can conclude that the structure around lag 24 has almost disappeared, even though there are still some autocorrelations at a few lags that could be considered as being statistically significant, indicating that, naturally, the fit is not completely perfect.

On the basis of these results, the proposed SDE model is deemed satisfactory on the training set. It now remains only to validate the model on an out-of-sample test set.

Since our aim is to provide a conditional distribution of wind speeds at a future time, conventional evaluation methods, such as the mean absolute error (MAE) or root mean squared error (RMSE), are not appropriate in this context, as they consider only the deviation from the point forecast. Instead, we evaluate the models in terms of likelihoods on the training and test sets, and in terms of the continuous rank probability score (CRPS) on the test set. We compute the likelihood as described in Section 2, whereas the CRPS is calculated according to the definition provided by Gneiting and Raftery (2007). Both of these scores are proper scoring rules (Gneiting & Raftery, 2007), meaning that a better fit of the data will produce a better score.

Please cite this article in press as: Iversen, E. B., et al., Short-term probabilistic forecasting of wind speed using stochastic differential equations. International Journal of Forecasting (2015), http://dx.doi.org/10.1016/j.ijforecast.2015.03.001

In assessing the performance of the proposed SDE model, we consider a number of benchmarks, namely the climatological model, where we use the empirical density of the training set to predict the next observation; the persistence prediction, where the wind speed forecast is given by the present value; and a family of autoregressive (AR) models. To be more precise, we consider a standard AR model and an ARX model, with the numerical weather prediction as an exogenous input. This is then extended to have a truncated Gaussian innovation. Lastly, we consider an ARX-GARCH model, also with an extension to a truncated Gaussian innovation. These benchmarks are classical time series models (Box, Jenkins, & Reinsel, 2013; Hamilton, 1994; Madsen, 2008). In particular, AR-GARCH methods have been used by Tol (1997) for forecasting daily wind speeds. For shorter horizons, Box–Jenkins type models have been used for modeling and forecasting wind speeds by Huang and Chalabi (1995) and Katz and Skaggs (1981), among others.

The persistance model is defined as:

$$Y_k = Y_{k-1} + \epsilon_k, \qquad \epsilon_{k+1} \sim \mathcal{N}(0, \sigma^2). \tag{17}$$

Here, we have used a Gaussian innovation. While the persistence forecast does not require any associated distribution of the innovation, this is needed for computing a likelihood. We choose the innovation to be Gaussian, in order to be in line with the classical AR model setup.

The AR model is specified as:

$$Y_k = \psi_0 + \sum_{i=1}^{p} \psi_i Y_{k-i} + \epsilon_k, \quad \text{where } \epsilon_k \sim \mathcal{N}(0, \sigma^2). \tag{18}$$

The ARX model takes the form:

$$Y_k = \psi_0 + \sum_{i=1}^{p} \psi_i Y_{k-i} + \phi p_k + \epsilon_k, \quad \epsilon_k \sim \mathcal{N}(0, \sigma^2). \tag{19}$$

For the ARX with the truncated Gaussian innovation, ϵ_k follows a Gaussian distribution truncated at $-\big(\psi_0 + \sum_{i=1}^{p} \psi_i Y_{k-i} + \phi p_k\big)$. This means that the process is confined to $\mathbb{R}^+$.

The ARX-GARCH model is of the form:

$$Y_k = \psi_0 + \sum_{i=1}^{p} \psi_i Y_{k-i} + \phi p_k + \epsilon_k, \quad \epsilon_k \sim \mathcal{N}(0, \sigma_k^2), \tag{20}$$

$$\sigma_k^2 = \alpha_0 + \sum_{i=1}^{\tilde{p}} \alpha_i \sigma_{k-i}^2 + \sum_{i=1}^{\tilde{q}} \beta_i \tilde{\epsilon}_{k-i} \quad \tilde{\epsilon}_k \sim \mathcal{N}(0, \tilde{\sigma}^2), \tag{21}$$

where the Gaussian innovation, $\epsilon_k \sim \mathcal{N}(0, \sigma_k^2)$, is truncated at $-\big(\psi_0 + \sum_{i=1}^{p} \psi_i Y_{k-i} + \phi p_k\big)$, for the variant of this benchmark with a truncated innovation.

The scores for the SDE model and the benchmarks are provided in Table 1. Clearly, the SDE model outperforms the benchmarks on the training and test sets in terms of both likelihoods and CRPS values. For the climatology benchmark, the empirical log-likelihood is used, because it is a non-parametric method, see for instance Bera and Bilias (2002). The SDE model thus provides significant improvements over both simple and complex benchmarks, with only a slight increase in the number of parameters with respect to the benchmark with the best performance.

The parameter estimates of the fitted SDE model are collated in Table 2.

The values of these estimates are reasonable. For example, $\hat{\mu}_x$ and $\hat{\rho}_x$ in the SDE model are close to 1, indicating that the wind speed tends toward the numerical weather prediction, albeit with some bias.

The SDE model can also be used to generate multi-horizon predictive densities. These predictive densities can be approximated using Monte Carlo simulations, and are shown in Fig. 7, where we see that the density spreads out as the horizon increases. Fig. 7 also includes a single simulated trajectory of the wind speed process (scenario), along with its actual realization. Scenarios can be simulated easily from our SDE model using one of the many numerical approaches provided by Jazwinski (2007) and Øksendal (2010). Nonetheless, we limit ourselves here to evaluating the predictive densities of the wind speed, and thus, we do not go into the realm of verification methods for time-path trajectories. For a more detailed discussion on the treatment and evaluation of time-path trajectories, we refer the interested reader to Pinson and Girard (2012) as a relevant practical case among recent developments in multivariate probabilistic forecast verification (see Table 3).

As was indicated in Section 2, the parameters in the SDE model are fitted to 1-hour-ahead data. Nevertheless,

Table 1
The log-likelihoods and CRPS values of the proposed SDE model, along with those of the selected benchmarks. The empirical log-likelihood is used for the climatology benchmark. The acronym TN stands for "truncated normal" (distribution).

Models	Parameters	Training set	Test set	
		LL	LL	CRPS
Climatology	–	−36 346	−19 609	1.2577
Persistence	1	−23 496	−12 023	0.5223
AR	6	−22 903	−11 764	0.5098
ARX	7	−22 087	−11 332	0.4847
ARX-TN	8	−21 445	−11 012	0.4882
ARX-GARCH	9	−21 715	−11 081	0.4717
ARX-GARCH-TN	10	−21 117	−10 784	0.4734
SDE Model	11	**−20 599**	**−10 433**	**0.4468**

Table 2
Parameter estimates for the SDE model.

$\hat{\theta}_x$	0.425	$\hat{\alpha}_x$	0.269
$\hat{\mu}_x$	0.817	$\hat{\phi}_x$	−0.191
$\hat{\sigma}_x$	0.421	$\hat{\theta}_u$	0.0185
$\hat{\sigma}_\epsilon$	0.0623	$\hat{\mu}_u$	−0.0502
$\hat{\beta}_x$	0.423	$\hat{\sigma}_u$	0.0439
$\hat{\rho}_x$	0.904		

Table 3
The CRPS for the SDE model and the two ARX-GARCH benchmarks. The non-iterative ARX-GARCH is fitted to each forecast horizon specifically. The iterative ARX-GARCH model is fitted to 1-hour-ahead data and then run iteratively until the desired prediction horizon is reached.

Models	Horizon			
	1 h	4 h	12 h	24 h
ARX-GARCH	0.4717	0.6168	0.6339	0.6357
ARX-GARCH-iterative	0.4717	0.6863	0.7464	0.7465
SDE model	0.4468	0.5713	0.6172	0.6236

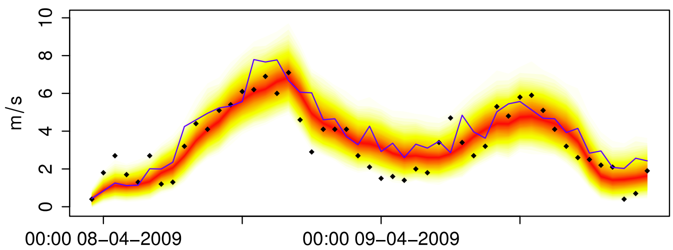

Fig. 7. The multi-step predictive densities of the SDE model for 1–48 hours ahead, obtained via Monte Carlo simulation, with warmer colors indicating higher probabilities. The blue line is a simulated plausible time-path trajectory (scenario) of the wind speed process, and the black dots are the actual realized values. (For the interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

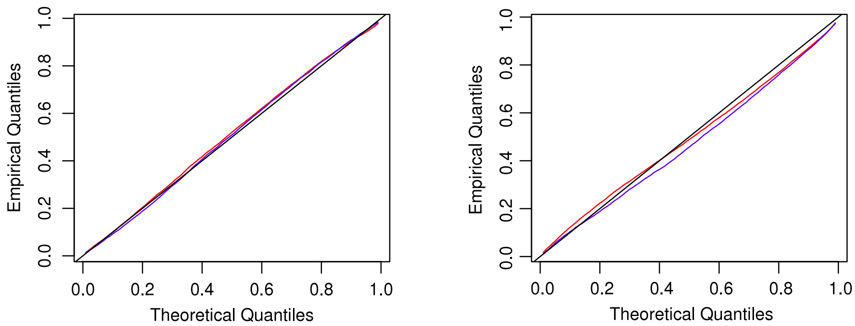

Fig. 8. QQ plot for the proposed SDE model for 1 h ahead (on the left) and 24 h ahead (on the right). For each one, we show the quantiles for the training set in red and the quantiles for the test set in blue. (For the interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

we also investigate the performance of our SDE model for prediction horizons of longer than one hour, in case our model were to be used for multi-horizon forecasting in future. For this purpose, the predictive densities obtained from the SDE model for four different lead times are compared with those given by an ARX-GARCH model with parameters fitted specifically for each lead time and those provided by an ARX-GARCH model with parameters fitted to 1-hour-ahead data, which is then run iteratively until the targeted prediction horizon is reached. The comparison is carried out in terms of the CRPS. These ARX-GARCH models are employed as benchmarks here because they perform the best among all of the benchmark models listed in Table 1. Note that the proposed SDE model outperforms the two ARX-GARCH models. This is particularly remarkable because one of the ARX-GARCH models is fitted expressly to the specific lead time under consideration. Furthermore, this ARX-GARCH model cannot be used to generate proper wind speed trajectories, while the proposed SDE model (and the "iterative" ARX-GARCH) can.

To assess how the quantiles implied by the proposed SDE model match up with the empirical ones, Fig. 8 shows two quantile–quantile plots, corresponding to the 1- and 24-hour-ahead predictions. These quantiles have been computed for both the training and test sets, by using Monte Carlo simulations with 10 000 samples for each time step. It is apparent that the 1-hour-ahead quantiles agree accurately with the empirical ones, whereas the 24-hour-ahead predictions underestimate the low quantiles and overestimate the larger quantiles slightly.

To conclude this section, we would like to mention that there do exist other modeling approaches that are designed specifically for multi-horizon forecasting. One of the most popular is to fit a marginal predictive density for each specific lead time and then combine them into a multivariate cumulative density using copula theory (Bessa et al., 2012; Schefzik et al., 2013). Copula-based methods are able to capture a wide range of distributions and inter-dependencies. However, they are computationally demanding and require the estimation of a large number of parameters. Stochastic differential equations can also be used for multi-horizon forecasting (Møller et al., 2013). This approach reduces the issue of the large number of parameters to be estimated, but is still very computationally intensive and can only handle a limited amount of data. That being said, we find that our SDE approach, albeit based on a one-step-ahead estimation procedure, can also be considered to be a reasonable method for multi-horizon forecasting, given that it relies on relatively few parameters and demands a computational time that is comparable to that of standard time series models, while showing a better performance.

5. Concluding remarks

A modeling framework based on stochastic differential equations for describing wind speeds in continuous time is proposed. We build a model that captures the bounded nature of the wind speed and its changing variability over time, and that incorporates numerical weather predictions and stochastic trends. The model outperforms both simple and complex benchmarks on both the training and test sets. The proposed modeling framework, based on SDEs, allows for an easy formulation of model extensions based on the physics of the system or statistical analysis. The different outputs of the model include accurate point forecasts, predictive densities and forecasts for multiple horizons that capture the interdependence in prediction errors, prediction intervals, and scenarios. Because scenarios can be generated readily from the proposed SDE model using standard numerical techniques, and because they preserve the same distribution as that observed in the data, the scenarios generated would be able to replace ensemble forecasts, which are used widely as an input to wind power forecasting.

Further research within the field of modeling energy systems using stochastic differential equations could be directed at the co-modeling of wind speeds and solar irradiance, as these are the physical phenomena underlying two of the most rapidly developing renewable energy sources: wind and solar power. The aggregation of power production from larger areas would also be of great relevance. The framework of SDEs may also allow for the spatio-temporal modeling of wind power by using stochastic partial differential equations. Another possibly fruitful topic for future research would be to extend the proposed wind speed model to predict wind power as well, possibly by introducing a dynamic power curve.

Acknowledgments

DSF (Det Strategiske Forskningsråd) is to be acknowledged for partly funding the work of Emil B. Iversen, Juan M. Morales and Henrik Madsen through the Ensymora project (no. 10-093904/DSF). Furthermore, Juan M. Morales and Henrik Madsen are funded partly by the Research Centre CITIES (no. 1035-00027B), which is also supported by DSF (Det Strategiske Forskningsråd). Finally, we want to thank the three anonymous reviewers for their insightful and constructive comments, which have helped to improve the paper significantly.

References

Adomian, G. (1988). *Nonlinear stochastic systems: theory and application to physics, Vol. 46*. Springer.
Bera, A. K., & Bilias, Y. (2002). The MM, ME, ML, EL, EF and GMM approaches to estimation: a synthesis. *Journal of Econometrics*, *107*(1), 51–86.
Bessa, R. J., Miranda, V., Botterud, A., Zhou, Z., & Wang, J. (2012). Time-adaptive quantile-copula for wind power probabilistic forecasting. *Renewable Energy*, *40*(1), 29–39.
Björk, T. (2009). *Arbitrage theory in continuous time. Oxford finance series*. Oxford: OUP, URL: http://books.google.dk/books?id=N0ImZs3HpwIC.
Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *26*, 211–252.
Box, G. E., Jenkins, G. M., & Reinsel, G. C. (2013). *Time series analysis: forecasting and control*. John Wiley & Sons.
Bremnes, J. B. (2004). Probabilistic wind power forecasts using local quantile regression. *Wind Energy*, *7*(1), 47–54.
Bremnes, J. B. (2006). A comparison of a few statistical models for making quantile wind power forecasts. *Wind Energy*, *9*(1–2), 3–11.
Costa, A., Crespo, A., Navarro, J., Lizcano, G., Madsen, H., & Feitosa, E. (2008). A review on the young history of the wind power short-term prediction. *Renewable and Sustainable Energy Reviews*, *12*(6), 1725–1744.
De Giorgi, M. G., Ficarella, A., & Tarantino, M. (2011). Assessment of the benefits of numerical weather predictions in wind power forecasting based on statistical methods. *Energy*, *36*(7), 3968–3978.
Foley, A. M., Leahy, P. G., Marvuglia, A., & McKeogh, E. J. (2012). Current methods and advances in forecasting of wind power generation. *Renewable Energy*, *37*(1), 1–8.
Gneiting, T., & Raftery, A. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378.
Hamilton, J. D. (1994). *Time series analysis, Vol. 2*. Princeton: Princeton University Press.
Huang, Z., & Chalabi, Z. (1995). Use of time-series analysis to model and forecast wind speed. *Journal of Wind Engineering and Industrial Aerodynamics*, *56*(2), 311–322.
Iversen, E. B., Morales, J. M., Møller, J. K., & Madsen, H. (2014). Probabilistic forecasts of solar irradiance using stochastic differential equations. *Environmetrics*, *25*, 152–164.
Jazwinski, A. H. (2007). *Stochastic processes and filtering theory*. Courier Dover Publications.
Källén, E. (1996). *Hirlam documentation manual; 1996. system 2.5*. Tech. rep. Sweden: Norrkøping, URL: http://www.hirlam.org/.
Katz, R. W., & Skaggs, R. H. (1981). On the use of autoregressive-moving average processes to model meteorological time series. *Monthly Weather Review*, *109*(3), 479–484.
Kristensen, N., Madsen, H., & Jørgensen, S. (2004). Parameter estimation in stochastic grey-box models. *Automatica*, *40*(2), 225–237.
Lerch, S., & Thorarinsdottir, T. L. (2013). *Comparison of nonhomogeneous regression models for probabilistic wind speed forecasting*. arXiv Preprint arXiv:1305.2026.
Ljung, L. (1999). *System identification*. Wiley Online Library.
Madsen, H. (2008). *Time series analysis, Vol. 72*. CRC Press.
Messner, J. W., Zeileis, A., Broecker, J., & Mayr, G. J. (2013). Probabilistic wind power forecasts with an inverse power curve transformation and censored regression. *Wind Energy*, *17*, 1753–1766.
Mikosch, T. (1998). *Elementary stochastic calculus: with finance in view, Vol. 6*. World Scientific.
Møller, J. K., Nielsen, H. A., & Madsen, H. (2008). Time-adaptive quantile regression. *Computational Statistics and Data Analysis*, *52*(3), 1292–1303.
Møller, J., Pinson, P., & Madsen, H. (2013). *Probabilistic forecasts of wind power generation by stochastic differential equation models*. Tech. rep. Technical University of Denmark, URL: http://www.statistics.gov.hk/wsc/STS019-P5-S.pdf.
Monteiro, C., Bessa, R., Miranda, V., Botterud, A., Wang, J., Conzelmann, G., et al. (2009). *Wind power forecasting: state-of-the-art 2009*. Tech. rep. Argonne National Laboratory (ANL).
Morales, J. M., Conejo, A. J., Madsen, H., Pinson, P., & Zugno, M. (2014). *Integrating renewables in electricity markets—operational problems. International series in operations research & management science, Vol. 205*. New York: Springer.
Nielsen, H. A., Nielsen, T. S., Madsen, H., Giebel, G., Badger, J., Landbergt, L., et al. (2006). From wind ensembles to probabilistic information about future wind power production—results from an actual application. In *International conference on probabilistic methods applied to power systems, 2006. PMAPS 2006* (pp. 1–8). IEEE.
Øksendal, B. (2010). *Stochastic differential equations: an introduction with applications. Universitext*. Springer, URL: http://books.google.dk/books?id=kXw9hB4EEpUC.
Pinson, P. (2012). Very-short-term probabilistic forecasting of wind power with generalized logit–normal distributions. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, *61*(4), 555–576.
Pinson, P., & Girard, R. (2012). Evaluating the quality of scenarios of short-term wind power generation. *Applied Energy*, *96*, 12–20.
Pinson, P., & Kariniotakis, G. (2010). Conditional prediction intervals of wind power generation. *IEEE Transactions on Power Systems*, *25*(4), 1845–1856.
Pinson, P., & Madsen, H. (2009). Ensemble-based probabilistic forecasting at Horns Rev. *Wind Energy*, *12*(2), 137–155.
Pinson, P., Nielsen, H., Møller, J., Madsen, H., & Kariniotakis, G. (2007). Non-parametric probabilistic forecasts of wind power: required properties and evaluation. *Wind Energy*, *10*(6), 497–516.
Ramirez-Rosado, I. J., Fernandez-Jimenez, L. A., Monteiro, C., Sousa, J., & Bessa, R. (2009). Comparison of two new short-term wind-power forecasting systems. *Renewable Energy*, *34*(7), 1848–1854.
Schefzik, R., Thorarinsdottir, T. L., & Gneiting, T. (2013). Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statistical Science*, *28*(4), 616–640.
Taylor, J. W., McSharry, P. E., & Buizza, R. (2009). Wind power density forecasting using ensemble predictions and time series models. *IEEE Transactions on Energy Conversion*, *24*(3), 775–782.
The European Wind Energy Association (2013). EWEA annual report 2012. Online. URL: http://www.ewea.org/publications/reports/ewea-annual-report-2012/.

Thorarinsdottir, T. L., & Gneiting, T. (2010). Probabilistic forecasts of wind speed: ensemble model output statistics by using heteroscedastic censored regression. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, *173*(2), 371–388.
Thorarinsdottir, T. L., & Johnson, M. S. (2012). Probabilistic wind gust forecasting using nonhomogeneous Gaussian regression. *Monthly Weather Review*, *140*(3), 889–897.
Tol, R. (1997). Autoregressive conditional heteroscedasticity in daily wind speed measurements. *Theoretical and Applied Climatology*, *56*(1–2), 113–122.
Van Kampen, N. G. (1992). *Stochastic processes in physics and chemistry, Vol. 1*. Elsevier.
Welch, G., & Bishop, G. (1995). An introduction to the Kalman filter. Online. URL: http://clubs.ens-cachan.fr/krobot/old/data/positionnement/kalman.pdf.
Zárate-Miñano, R., Anghel, M., & Milano, F. (2013). Continuous wind speed models based on stochastic differential equations. *Applied Energy*, *104*, 42–49.
Zhou, Z., Botterud, A., Wang, J., Bessa, R., Keko, H., Sumaili, J., & Miranda, V. (2013). Application of probabilistic wind power forecasting in electricity markets. *Wind Energy*, *16*(3), 321–338.

Please cite this article in press as: Iversen, E. B., et al., Short-term probabilistic forecasting of wind speed using stochastic differential equations. International Journal of Forecasting (2015), http://dx.doi.org/10.1016/j.ijforecast.2015.03.001