Wind Energy

WIND ENERGY
Wind Energ. 2017; **20**:33–44
Published online 2 May 2016 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/we.1988

RESEARCH ARTICLE

Leveraging stochastic differential equations for probabilistic forecasting of wind power using a dynamic power curve

Emil B. Iversen, Juan M. Morales, Jan K. Møller, Pierre-Julien Trombe and Henrik Madsen
Technical University of Denmark, Asmussens Alle, Building 303B, DK-2800 Lyngby, Denmark

ABSTRACT

Short-term (hours to days) probabilistic forecasts of wind power generation provide useful information about the associated uncertainty of these forecasts. Standard probabilistic forecasts are usually issued on a per-horizon-basis, meaning that they lack information about the development of the uncertainty over time or the inter-temporal correlation of forecast errors for different horizons. This information is very important for forecast end-users optimizing time-dependent variables or dealing with multi-period decision-making problems, such as the management and operation of power systems with a high penetration of renewable generation. This paper provides input to these problems by proposing a model based on stochastic differential equations that allows generating predictive densities as well as scenarios for wind power. We build upon a probabilistic model for wind speed and introduce a dynamic power curve. The model thus decomposes the dynamics of wind power prediction errors into wind speed forecast errors and errors related to the conversion from wind speed to wind power. We test the proposed model on an out-of-sample period of 1 year for a wind farm with a rated capacity of 21 MW. The model outperforms simple as well as advanced benchmarks on horizons ranging from 1 to 24 h. Copyright © 2016 John Wiley & Sons, Ltd.

KEYWORDS
wind power; dynamic power curve; stochastic differential equations; probabilistic forecast

Correspondence
Emil B. Iversen, Technical University of Denmark, Asmussens Alle, Building 303B, DK-2800 Lyngby, Denmark.
E-mail: jebi@dtu.dk

Received 6 February 2015; Revised 21 February 2016; Accepted 24 February 2016

1. INTRODUCTION

Renewable energies have gained prominence in recent years as a sustainable solution to the world's growing energy demand. The largest share of renewable energy is produced by wind turbines in many countries.[1] The integration of wind power into the electricity grid is not without challenges. These are related to the variable and partly unpredictable behavior of the power generated by wind farms. To account for the stochastic nature of wind and to mitigate its potential adverse effects, accurate forecasts of future wind power generation are required. For the most efficient utilization of wind power, both in terms of grid stability and economic costs, forecasts that provide the full predictive distribution of the wind power generation are required.[2,3] Such forecasts are referred to as *probabilistic* in contrast to their deterministic counterpart, which consists of a single-value forecast, most typically the expected value or the most likely outcome.

1.1. Motivation

Figure 1 serves to motivate the present research work. In this figure, the normalized wind power is plotted against the observed wind speed. These observations come from the Klim wind farm in Denmark. It is apparent that the relationship between observed wind speeds and generated power is not deterministic. Two sequences, each consisting of 15 data points, are highlighted in red and blue, together with a local polynomial regression model of the relationship between wind speeds and generated power, shown in black, typically known as the power curve. The red sequence of data points indicates that the regressed curve consistently underestimates the power output. Conversely, the curve systematically overestimates the power output for the blue sequence of points. This reveals that the relationship between observed wind speed and observed power output exhibits some memory and may change over time. This may be attributed to a variety of factors such as the

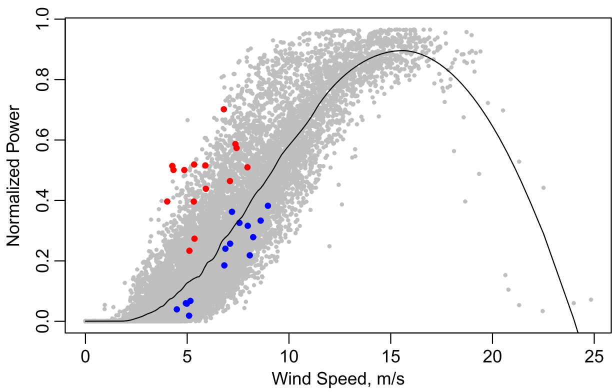

Figure 1. Normalized power plotted against wind speeds in gray both in hourly averages. The black line indicates a local polynomial regression model of the relationship between wind speeds and normalized power. The red points represent a sequence of 15 hourly observations. The blue dots are a sequence of 15 hourly observations 9 months later.

turbine blades being dirty, different local turbulence characteristics (tree foliage and wind shadowing) or the wind direction. These features have also been pointed out by other authors for other data sets.[4,5] The key takeaway, however, remains the same: If the power curve is to be used for forecasting, it should be dynamic and adaptive to accommodate the changes in the relationship between wind speed and the generated power.

There exists a wealth of approaches for generating wind power forecasts, albeit mostly centered on point predictions. Thorough reviews of the most relevant works in this field are given in Costa *et al.*,[6] Monteiro *et al.*[7] and Foley *et al.*[8]

Several studies document and emphasize the importance of probabilistic forecasts for wind power integration. Among such examples are Pinson *et al.*,[9] which illustrates the value of such forecasts for energy trading. In Bludszuweit and Domínguez-Navarro,[10] the usefulness of probabilistic forecasts for operating energy storage is highlighted. Another example is Matos and Bessa,[11] where determining the grid operating reserve is in focus. In response to this need for describing the uncertainty in the wind power forecast, probabilistic forecasts have received increased attention in recent years with the application of methods such as quantile regression[12,13] and the use of ensemble forecasts[14] and Box–Jenkins models.[15] Furthermore, forecasters are also focusing on generating scenarios and constructing models that capture the interdependence in time of forecast errors.[16−18]

Power curves for individual wind turbines are usually supplied by the wind turbine manufacturer. This power curve is typically deterministic. For wind farms, the aggregate power curve depends on many other factors, such as the topography, wind shading and wind direction. A stochastic power curve has been modeled in Jeon and Taylor,[4] where the time variation of the curve due to some of these factors is captured by an adaptive estimation procedure. A Markovian power curve is used in Anahua *et al.*[19] to simulate wind power fluctuations. In Gottschall and Peinke,[20] a stochastic power curve is also shown to yield a more accurate model of the wind power generation.

The main contribution of this paper is the development of a dynamic power curve model that allows us to translate a probabilistic forecast of wind speed into a probabilistic forecast of wind power. For this purpose, the proposed power curve model is both probabilistic and dynamic, so as to reflect the changing characteristics of the wind farm. This is achieved by setting up our model in the form of stochastic differential equations.

The remainder of this paper is organized as follows: Section 2 gives a very brief introduction to stochastic differential equations and how to estimate their characteristic parameters. In Section 3, we present two stochastic differential equation models for wind power, which differ in whether wind speed observations are used or not. The following section, Section 4, deals with the performance of the generated forecasts. Section 5 concludes the paper and outlines future research.

2. STOCHASTIC DIFFERENTIAL EQUATIONS

Stochastic differential equations (SDEs) extend ordinary differential equations (ODEs) by including one or more stochastic terms. The solution to an SDE is, therefore, a stochastic process. SDE models have been employed in a variety of fields to describe systems with a large stochastic component. This type of models has been widely used in mathematical finance[21,22] and physics.[23,24] More recently, such models have also been used for modeling and forecasting solar irradiance[25] and wind speed.[26] In this section, we provide a very succinct description of SDEs and refer the reader to Øksendal[27] or Jazwinski[28] for general discussions on the topic and to Kristensen *et al.*[29] for more details on the specific approach followed in this paper.

Wind Energ. 2017; **20**:33–44 © 2016 John Wiley & Sons, Ltd.
DOI: 10.1002/we

The standard notation for SDEs is similar to that of ODEs and takes the following form.

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

where W_t is a Wiener process, and $f(\cdot)$ and $g(\cdot)$ are known as the drift and diffusion terms, respectively. Owing to the fact that the derivative of W_t, dW_t, does not exist, this equation should be interpreted as a convenient way of writing the following integral equation:

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

where the Itō interpretation of the second integral should be used.[27]

Solving the SDE yields a stochastic process, which characterizes the uncertainty in the process dynamics for every future time. The probability density function of the stochastic process at a certain time in the future can be obtained by solving a partial differential equation referred to as the Fokker–Plank or forward Kolmogorov equation; see Björk[21] for details. In general, this solution cannot be found analytically; however, a significant number of numerical solution approaches are currently available (e.g. Jazwinski[28]).

A convenient result about the generality of SDEs is the Lévy-Itō decomposition,[21] which essentially states that all stochastic processes with continuous trajectories can be written as special cases of SDEs. Indeed, many ordinary time series models can be seen as discrete time interpretations of SDEs.

Though continuous in nature, in practice, SDEs can only be observed or measured at discrete instances in time. For this reason, an *observation* or *measurement* equation, $h(\cdot)$, is introduced into the SDE system, which then becomes

$$dX_t = f(X_t, t)dt + g(X_t, t)dW_t \tag{3}$$

$$Y_k = h\left(X_{t_k}, t_k, e_k\right) \tag{4}$$

Notice that we have now defined the observation Y_k of the process at time t_k. This measurement equation also allows us to specify an observation error term, e_k, in our model.

The procedure that we use to estimate the SDE parameters hinges on a specific form of the observation equation, $h(\cdot)$, and the diffusion term, $g(\cdot)$. In particular, we require $g(X_t, t) = g(t)$ and $h\left(X_{t_k}, t_k, e_k\right) = h\left(X_{t_k}, t_k\right) + e_k$, where $e_k \sim \mathcal{N}\left(0, \sigma^2\right)$. In principle, these conditions limit our modeling framework significantly. However, they can be relaxed to a large extent. Indeed, the requirement that $g(\cdot)$ does not depend on X_t can be overcome by making use of Itō calculus to transform the original process with state-dependent diffusion to an equivalent one with *non*-state-dependent diffusione (e.g. Møller *et al.*[13] and Iversen *et al.*[25]). The second requirement, which implies that the observation error is additive and normally distributed, can be relieved by transforming the data (e.g. Box and Cox[30]).

The method we use for parameter estimation has been outlined in several previous papers and therefore, we will not describe it here. Instead, we refer the reader to Jazwinski[28] for a general discussion on parameter estimation for SDEs and to Møller and Madsen,[31] Iversen *et al.*[26] and Juhl *et al.*[32] for practical guidance on its application, which also includes an open source implementation. Note that the estimation procedure, as implemented in the R package, entails maximizing the likelihood function over the residuals computed one step ahead. We follow suit in this paper. Nonetheless, estimating SDE models using a multi-horizon approach is also possible, but at a high computational cost; see Møller *et al.*,[18] for instance.

3. MODELS FOR WIND POWER

In this section, we propose two models for wind power: one that assumes that both wind power and wind speed are measured and one where we only observe wind power. There are several reasons that motivate these two approaches: first of all, not all wind farms have access to wind speed data. Second, the difference in performance of the two models may highlight the benefit of having wind speed measurements available. Third, not observing wind speed allows us to investigate the feasibility of constructing a dynamic power curve model in such a case.

3.1. Model with wind speed observations

We start out with the following probabilistic model for wind speed

$$dX_t = \left(\left(1 - e^{-X_t}\right)\left(\rho_x \dot{p}_t + R_t\right) + \theta_x(p_t\mu_x - X_t)\right)dt + \sigma_x X_t^{0.5} dW_t \tag{5}$$

$$dR_t = -\theta_r R_t dt + \sigma_r dW_t \tag{6}$$

$$Y_{1,k} = X_{t_k} + \epsilon_k \tag{7}$$

which is a simplified version of the model developed in Iversen *et al.*[26] $X_t, R_t, Y_{1,k}$ and ϵ_k are stochastic variables, with X_t being the actual wind speed, $Y_{1,k}$ being the observed wind speed, R_t can be interpreted as a dynamic correction of the numerical weather prediction (NWP) p_t and $\epsilon_k \sim \mathcal{N}(0, \sigma_\epsilon)$ is an observation error. The first-order differentiation in the NWP of wind speed is introduced as $\dot{p}_t$. The term $\left(1 - e^{-X_t}\right)$ is included to keep the process in the proper domain, $\mathbb{R}^+$. ρ_x governs the emphasis put on $\dot{p}_t$. θ_x and θ_r govern the speed of reversion to $\mu_x p_t$ and zero for X_t and R_t, respectively. μ_x corrects the NWP from systematically over or under shooting the observed values of wind speed. σ_x, σ_r and σ_ϵ govern the size of the diffusions of X and R and the size of the observation error, respectively. The model described here succeeds in describing the dynamics of the wind speeds at a specific location. The diffusion term $\sigma_x X_t^{0.5}$ drops to zero when X_t is close to zero. In such a case, the drift term dominates and pushes the process away from zero with the result that X_t always takes on values within the domain $\mathbb{R}^+$.

Because we presume that a large part of the variation in the wind power output can be explained by the underlying wind speed process, we aim at capturing the speed-to-power relationship by introducing a dynamic power curve in the SDE system (5)–(7). This results in the following probabilistic model for wind power.

$$dX_t = \left(\left(1 - e^{-X_t}\right)\left(\rho_x \dot{p}_t + R_t\right) + \theta_x\left(p_t \mu_x - X_t\right)\right) dt + \sigma_x X_t^{0.5} dW_{x,t} \tag{8}$$

$$dR_t = -\theta_r R_t dt + \sigma_r dW_{r,t} \tag{9}$$

$$dQ_t = (S_t - \theta_q Q_t) dt + \sigma_q dW_{q,t} \tag{10}$$

$$dS_t = -\theta_s S_t dt + \sigma_s dW_{s,t} \tag{11}$$

$$Y_{1,k} = X_{t_k} + \epsilon_{1,k} \tag{12}$$

$$\begin{aligned} Y_{2,k} = &\left(0.5 + 0.5 \tanh(5(X_{t_k} - \gamma_1))\right)(0.5 - 0.5 \tanh(\gamma_2(X_{t_k} - \gamma_3))) \\ &\frac{\zeta_3}{1 + e^{-\zeta_1 (X_{t_k} - \zeta_2 + Q_{t_k})}} + \epsilon_{2,k} \end{aligned} \tag{13}$$

where equations (8), (9) and (12) correspond to our previous wind speed model. Notice that we have included a new observation, $Y_{2,k}$, which is the normalized wind power, while $Y_{1,k}$ in equation (12) still stands for the measured wind speed, with $\epsilon_{1,k}$ being the associated measurement error, which is assumed to follow a normal distribution with variance $\sigma^2_{\epsilon_1}$. The functional relationship between the wind speed and the normalized power is partly determined by the parameters γ_1, γ_2 and γ_3, which model the cut-in and cut-out wind speeds and how fast the wind turbines cut-out, and partly by ζ_1, ζ_2 and ζ_3, which govern the slope of the power curve in the domain where the wind turbines produce power. We also introduce the stochastic variable Q_t to capture the fact that the speed-to-power efficiency of the wind farm changes over time in a stochastic manner. This feature of the model is based on observations from the data, as shown in Figure 1, which may be caused by factors such as dirty blades, how laminar the wind flow is, turbine failures or turbine maintenance, all of which strongly affect the rate at which wind is converted into power. Thus, we add dynamics to the power curve to account for this shift over time through Q_t. S_t is a stochastic variable that helps model the dynamics of Q_t. Along with S_t and Q_t, we introduce the parameters $\theta_q, \theta_s, \sigma_q$ and σ_s that describe the evolution of these stochastic variables. The power curve in equation (13) is the result of an iterative model-building process supported by the physical understanding of the specific system we are modeling. There is a cut-in wind speed, γ_1, and a cut-out wind speed, γ_3. The structure of the power curve is further captured by $\zeta_3 \left(1 + e^{-\zeta_1 (X_{t_k} - \zeta_2 + Q_{t_k})}\right)^{-1}$, with ζ_3 determining the saturation level, and ζ_1 and ζ_2 determining the slope and location of the curve, respectively. Notice that the data required for this model is wind speed, Y_1, and wind power, Y_2, as well as a numerical weather prediction, p_t.

The parameters in the aforementioned model are collated in Table I along with their respective interpretation. Likewise, the stochastic variables present in the SDE model are listed and defined in Table II.

Finally, the subscripts '1', '2', 'x', 'r', 'q' and 's' are conventions used to name the different parameters of the model in relation to the specific stochastic variable they refer to. Subscripts t, t_k and k are related to time indices and are provided and explained in Table III.

Wind Energ. 2017; **20**:33–44 © 2016 John Wiley & Sons, Ltd.
DOI: 10.1002/we

Table I. List of parameters involved in the SDE models.

γ_1	Cut-in wind speed
γ_2	Transition to cut-out parameter
γ_3	Cut-out wind speed
ζ_1	Power curve shape parameter
ζ_2	Power curve shape parameter
ζ_3	Power curve shape parameter
σ_q	Rate of diffusion for Q
θ_q	Rate of reversion to mean for Q
σ_r	Rate of diffusion for R
θ_r	Rate of reversion to mean for R
σ_s	Rate of diffusion for S
θ_s	Rate of reversion to mean for S
$\hat{\rho}_x$	Factor normalizing the change in wind speed prediction, $\dot{p}_t$
$\hat{\mu}_x$	Factor normalizing the wind speed prediction, p_t
$\hat{\theta}_x$	Rate of reversion to mean for X
$\hat{\sigma}_x$	Rate of diffusion for X
$\hat{\sigma}^2_{\epsilon_1}$	Wind speed observation variance
$\hat{\sigma}^2_{\epsilon_2}$	Wind power observation variance

Table II. List and definition of stochastic variables.

X_t	Actual wind speed at time t
R_t	Auxiliary stochastic variable determining X_t
Q_t	Auxiliary stochastic variable determining Y_{2,t_k}
S_t	Auxiliary stochastic variable determining Q_t
Y_{1,t_k}	Observed wind speed at time t_k
Y_{2,t_k}	Observed wind power at time t_k
$\epsilon_{1,k}$	Wind speed observation error
$\epsilon_{2,k}$	Wind power observation error

Table III. Time and sample indices.

t	Time index for the stochastic processes (continuous).
k	Sampling index for the observations (discrete).
t_k	Sampling time index, the time t of sample k.

3.2. Model without wind speed observations

An obvious variant of the former wind power model results from considering that wind speed measurements are not directly available (e.g. through an anemometer), and thus, only the power output of the wind farm is observed. This model would be useful if we were to apply the model on a wind farm where we do not have access to wind speed observations. This could, in principle, be performed by simply removing equation (12) from the aforementioned model. However, if we proceeded in this way, we would end up with an unidentifiable model, the parameters of which could not be estimated and for which the underlying states could not be filtered. The idea here is that we can decompose model (8)–(13) into a system with fast dynamics (the wind speed process) and a system with slow dynamics (the conversion from speed to power). However, if we do not observe the wind speed, this decomposition turns out to be infeasible. Ljung[33] addresses the topic of identifiability for stochastic models. Consequently, in the absence of wind speed measurements, the SDE model (8)–(13) has to be simplified substantially, resulting in the following model:

$$dX_t = \left(\left(1 - e^{-X_t}\right)(\rho_x \dot{p}_t + R) + \theta_x(p_t \mu_x - X_t)\right) dt + \sigma_x X_t^{0.5} dW_t \tag{14}$$

$$dR_t = -\theta_r R_t dt + \sigma_r dW_t \tag{15}$$

$$Y_{2,k} = (0.5 - 0.5\tanh(\gamma_2(X_{t_k} - \gamma_3)))\frac{\zeta_3}{1 + e^{-\zeta_1(X_{t_k} - \zeta_2)}} + \epsilon_{2,k} \tag{16}$$

In the model specified by equations (14)–(16), the power curve (16) is not dynamic as it remains unchanged over time. In this case, X_t can no longer be interpreted as the wind speed per se but instead as a state variable that we may call the *effective wind speed*, that is, the wind speed that would yield the measured power output given the deterministic power curve defined in equation (16). Notice that the data required for this model is wind power, Y_2, as well as a numerical weather prediction, p_t.

4. MODEL EVALUATION

In this section, we focus on validating the proposed model (8)–(13) while also considering the difference between this model and model (14)–(16).

The data used in this study originate from the Klim Fjordholme wind farm and consists of power measurements and predicted wind speeds based on an NWP model from the Danish Meteorological Institute.[34] The data are sampled hourly. The power data from the Klim wind farm have been used in previous studies; see, for instance, Pinson.[3] The NWP model provides a wind speed forecast up to 48 h ahead and is updated every 6 h. The data covers 3 years, from 1 January 1999 to 31 December 2001 and is divided into two periods: a training set spanning 2 years, which is used for estimation, and a test set covering the remaining 1 year to evaluate the performance of the proposed model.

The first step in the validation task is to inspect the parameter estimates of the model, which are easily interpretable and, as such, provide a quick check of how (un)reasonable our model is at a first glance. The parameter values are shown in Table IV. The parameter estimation procedure takes approximately 2 h on a server with 24 1.9-GHz-cores and 64 GB of memory. Besides, the optimization routine that is employed for the estimation is parallelized to some extent.

The shape of the stationary power curve, which is depicted in Figure 2 along with the observations, is determined by the parameters $\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3, \hat{\xi}_1, \hat{\xi}_2$ and $\hat{\xi}_3$. Notice that the cut-out wind speed, which is given by γ_3 in the model, is estimated to be 19.63 $\mathrm{m\,s^{-1}}$, that is, close to the rated cut-out wind speed for the Vestas V44 wind turbines installed at Klim Fjordholme, which is 20 $\mathrm{m\,s^{-1}}$. It is important to stress that the power curve that is modeled here corresponds to that of the wind farm and not to the power curve for the individual wind turbine. We have normalized the wind farm power output with its rated capacity of 21 MW. However, the actual nominal maximal production seems closer to $\hat{\xi}_3 \cdot 21 = 20.00$ MW, which may be because of turbine aging. Notice also that $\hat{\mu}_x$ is somewhat close to one indicating that the predicted wind speed for longer horizons should tend to the numerical weather prediction, albeit with some bias.

To give an indication of the SDE model performance, we compare it against some simple and some more advanced benchmarks. We focus on standard time series benchmarks ranging from a persistence model to an ARX-GARCH. All the benchmark models are applied directly to observed power generation. The persistence model of order j is given by

$$Y_{2,k} = Y_{2,k-j} + \epsilon_{2,k}, \qquad \epsilon_{2,k} \sim \mathcal{N}\left(0, \sigma^2\right) \tag{17}$$

Table IV. Parameter estimates for model (8)–(13).

$\hat{\gamma}_1$	2.648	$\hat{\sigma}_q$	1.093	$\hat{\rho}_x$	0.3087
$\hat{\gamma}_2$	0.657	$\hat{\theta}_r$	1.358	$\hat{\mu}_x$	0.7120
$\hat{\gamma}_3$	19.63	$\hat{\sigma}_r$	1.476	$\hat{\theta}_x$	0.2317
$\hat{\xi}_1$	0.612	$\hat{\theta}_s$	0.0629	$\hat{\sigma}_x$	0.0884
$\hat{\xi}_2$	9.312	$\hat{\sigma}_s$	0.7195	$\hat{\sigma}^2_{\epsilon_1}$	0.1324
$\hat{\xi}_3$	0.952	$\hat{\theta}_q$	1.788	$\hat{\sigma}^2_{\epsilon_2}$	0.0001259

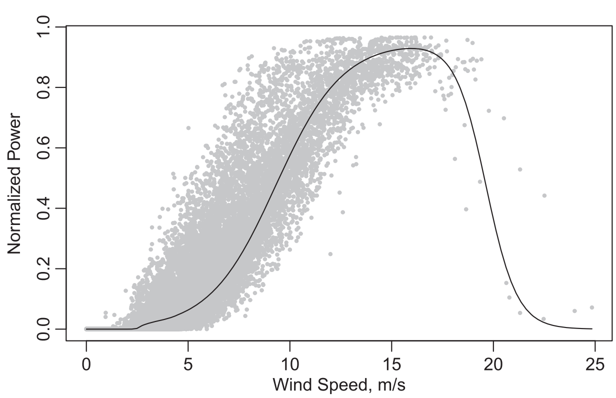

Figure 2. The power curve obtained when the states Q_t and S_t are at their long-term stationary levels.

Wind Energ. 2017; **20**:33–44 © 2016 John Wiley & Sons, Ltd.
DOI: 10.1002/we

While a persistence model does not require any specific associated distribution of the innovation term, we assume that it is normally distributed to fit into the classical auto-regressive (AR) model setup and to compute the continuous ranked probability score for the persistence model.

We specify an AR model of order q as

$$Y_{2,k} = \psi_0 + \sum_{i=1}^{q} \psi_i Y_{2,k-i} + \epsilon_{2,k}, \qquad \epsilon_{2,k} \sim \mathcal{N}\left(0, \sigma^2\right) \tag{18}$$

An AR with external input (ARX) model takes the form

$$Y_{2,k} = \psi_0 + \sum_{i=1}^{q} \psi_i Y_{2,k-i} + \phi p_{t_k} + \epsilon_{2,k}, \qquad \epsilon_{2,k} \sim \mathcal{N}\left(0, \sigma^2\right) \tag{19}$$

where $\tilde{p}_{t_k}$ is the wind power prediction at time t_k and where $\tilde{p}_{t_k} = j\left(p_{t_k}\right)$. $j(\cdot)$ is approximated by a polynomial regression model to find the relationship between predicted wind speed and power (similar to Figure 1). We also consider an ARX model with truncated normal distributed innovation (ARX-TN), where $\epsilon_{2,k}$ follows a truncated normal, such that the process is confined to $[0, 1]$.

Lastly, we consider an ARX model and a generalized AR conditional heteroskedastic variance term (ARX-GARCH) that takes the following form:

$$Y_{2,k} = \psi_0 + \sum_{i=1}^{q} \psi_i Y_{2,k-i} + \phi \tilde{p}_{t_k} + \epsilon_{2,k}, \qquad \epsilon_{2,k} \sim \mathcal{N}\left(0, \sigma_k^2\right) \tag{20}$$

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

where the normally distributed innovation, $\epsilon_{2,k} \sim \mathcal{N}\left(0, \sigma_k^2\right)$ is truncated such that the process is confined to $[0, 1]$ for the variant of this benchmark with truncated innovation (ARX - GARCH - TN).

Models with the same general structure as the benchmarks earlier have been used to forecast wind power production in several scientific publications such as Duran *et al.*,[35] Jeon and Taylor,[4] Taylor *et al.*,[36] Lau and McSharry[37] and Trombe *et al.*[15]

In Table V, the performance for the benchmarks and the proposed SDE model are shown for one-step (i.e.1 h) ahead forecasts. We evaluate in terms of various scores the mean absolute error (MAE), which is computed as

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{22}$$

and the root mean squared error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \tag{23}$$

where N is the number of observations, y_i is the i'th wind power observation and $\hat{y}_i$ is the corresponding forecast power.

Table V. The MAE, RMSE and CRPS scores for benchmarks as well as for the proposed model.

		Test set			
Models	Parameters	MAE	RMSE	CRPS	SS_{CRPS}
Climatology	—	0.2208	0.2693	0.1417	−2.4900
Persistence	1	0.0509	0.0835	0.0428	−0.0541
AR	4	0.0527	0.0820	0.0417	−0.0270
ARX	5	0.0510	0.0795	0.0406	0.0000
ARX - TN	7	0.0648	0.0848	0.0444	−0.0935
ARX - GARCH	9	0.0505	0.0797	0.0382	0.0591
ARX - GARCH - TN	11	0.0575	0.0823	0.0401	0.0123
Model (8)–(13)	19	**0.0471**	**0.0773**	**0.0327**	**0.1945**
Model (14)–(16)	12	0.0553	0.08988	0.0399	0.0172

Also, we use the continuous ranked probability score (CRPS) as defined in Gneiting and Raftery[38] to rank and compare the models in a probabilistic sense. In particular, the CRPS is defined as

$$CRPS = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}\mathbb{E}_{F_i}|Y - Y'| - \mathbb{E}_{F_i}|Y - y_i| \tag{24}$$

where F_i is the forecast probability cumulative density function for observation y_i, and Y and Y' are independent copies of the random variable with distribution F_i. Further, we introduce a skill score for the CRPS to facilitate easy comparison between models. The skill score on the CRPS (SS_{CRPS}) is defined as $SS_{CRPS} = 1 - CRPS_{model}/CRPS_{ref}$, where the ARX model is used as reference model.

The energy score (ES) is, in turn, given by

$$ES = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}\mathbb{E}_{F_i}||Y - Y'||^{\beta} - \mathbb{E}_{F_i}||Y - y_i||^{\beta} \tag{25}$$

where Y' and $Y \in \mathbb{R}^m$ and $||\cdot||$ defines the euclidean norm. If $\beta = 1$ and $m = 1$, the ES boils down to the CRPS. We choose $\beta = 1$ in subsequent uses of the ES. It follows from the aforementioned definitions that both the CRPS and the ES evaluate *predictive densities* and not point predictions as the RMSE and MAE do.

We see here that the SDE model (8)–(13) outperforms simple as well as stronger benchmarks. Comparing model (8)–(13) and model (14)–(16), we note that there is a clear benefit from having wind speed observations along with a dynamic power curve, this is both in terms of the point forecast, as measured by the MAE and the RMSE, as well as in terms of the predictive density, as measured by the CRPS.

The 1 h ahead and 24 h ahead predictive densities of wind power that the proposed SDE model provides are shown in Figure 3. Three observations are in order. First, note that the predictive densities are more spread out for the 24 h ahead forecasts, revealing a higher degree of uncertainty for longer prediction horizons, as expected. Second, observe that the density is also more spread out when the predicted normalized power is around 0.5. This follows from the fact that this is the power value around which the power curve has the steepest slope and therefore, a certain variation in wind speed will, all other things equal, yield a larger variation in the power produced. Third, notice that the predictive density seems to become quite sharp around zero predicted power. This can be explained by means of Figure 2, where we can see that the power curve is essentially flat and equal to zero for low wind speeds, thus yielding zero power and very little variation for low wind speed values. This is a clear advantage over Box–Jenkins-type models, as the (8)–(13) model can associate a NWP of small wind speeds with very little variability in the prediction of generated power.

As hinted in Figure 3, SDEs automatically provide a framework for providing multi-horizon forecast densities. In this line, Figure 4 displays the predictive densities for 1–24 h ahead. These can be obtained by solving the Fokker–Planck equation that specifies the predictive densities of our SDE model.[21] The solution is found by a numerical approximation using a Monte Carlo simulation method. Specifically, we have used an Euler method to simulate a large number of time-path trajectories (10, 000), which are subsequently used to approximately construct the predictive densities.

It is important to note, however, that even though the proposed SDE model is estimated based on the one-step ahead residuals, it also readily provides good predictive densities for horizons beyond the first hour. Notwithstanding this,

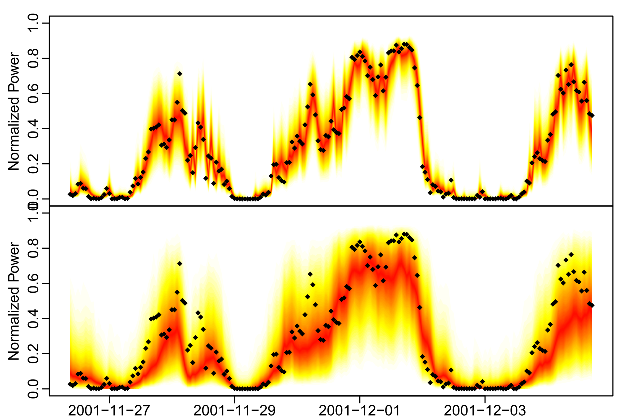

Figure 3. 1 h (top) and 24 h (bottom) ahead densities predicted by model (8)–(13), with warmer colors indicating a higher probability of seeing this realization. The densities are approximated by Monte Carlo simulations.

Wind Energ. 2017; **20**:33–44 © 2016 John Wiley & Sons, Ltd.
DOI: 10.1002/we

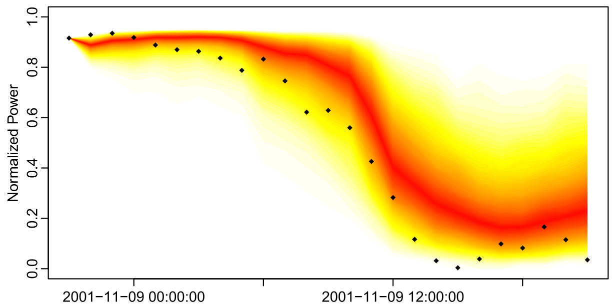

Figure 4. Multi-horizon predictive densities for normalized wind power generation from model (8)–(13). Warmer colors indicate a higher probability of seeing this realization. The density is approximated by Monte Carlo simulations.

Table VI. The CRPS and energy score for SDE models (8)–(13) and (14)–(16) and for the ARX-GARCH benchmarks. The ARX-GARCH benchmark is fitted specifically for each forecast horizon. The iterative ARX-GARCH-TN model is fitted to 1 h ahead data and then run iteratively until the desired horizon is reached.

	CRPS for different horizons				
Models	1 h	4 h	12 h	24 h	Energy score
ARX - GARCH	0.0382	0.0704	0.0787	**0.0789**	1.180
ARX - GARCH - TN - iterative	0.0401	0.0783	0.1043	0.1225	1.945
Model (8)–(13)	**0.0327**	**0.0641**	**0.0779**	0.0836	**0.739**
Model (14)–(16)	0.0399	0.0685	0.0784	0.0793	0.745

we evaluate next the performance of the proposed SDE model beyond 1 h horizons and compare it with the best performing benchmarks on these same horizons. The results of this comparison are collated in Table VI.

We compare in terms of the CRPS and the ES, both defined in Gneiting and Raftery,[38] and we consider two benchmarks. First, an ARX-GARCH model, which is specified previously, but now fitted specifically to each prediction horizon. On the one hand, this confers an advantage on this benchmark model with respect to the proposed one, because, as we have just mentioned, the SDE model is only fitted for 1 h ahead data and forecasts for longer horizons are obtained by extrapolation. On the other hand, the so-fitted ARX-GARCH model fails to capture the time structure of forecast errors. This brings us to the second benchmark, the ARX-GARCH-TN-iterative, where we run the ARX-GARCH-TN in an iterative fashion to obtain forecasts for the desired horizons. This benchmark is able to provide both trajectories and multi-horizon forecasts. We choose the truncated version of the ARX-GARCH model, as the standard ARX-GARCH model becomes unstable for a large number of iterations.

In Table VI, we see that model (8)–(13) outperforms the two benchmarks for all lead times shorter than 24 h in terms of the CRPS. However, the ARX-GARCH benchmark performs the best on the 24 h horizon. We insist, though, that the ARX-GARCH model is specifically designed to provide power forecasts on the desired horizon while failing to produce multi-horizon forecasts and time-path trajectories with the proper time structure. The proposed model (8)–(13) also outperforms the two benchmarks in terms of the energy score. Further insight can be gained in Table VI by comparing model (8)–(13) and model (14)–(16). The former exploits wind speed observations, while the latter does not. Notice that including wind speed observations improves the forecast significantly for short horizons, but this is not the case for longer horizons. This is so because information on recently past wind speed proves to be useful to identify and predict the range of the power curve in which the wind farm is operating. In practice this allows a more advantageous use of the numerical weather prediction for predicting wind speeds. Naturally, the predictive power of wind speed measurements is first diluted and finally vanishes with longer prediction horizons.

Table VII is analogous to Table VI, but in terms of MAE, which is a metric to evaluate the point-prediction performance of the different wind power models. Note that both tables convey similar messages.

Other approaches exist for producing multi-horizon forecasts of wind power production. One method that is particularly popular is to fit a marginal predictive density for each specific lead time and then to combine them into a cumulative distribution using copulas.[17,39] Copula methods are, however, computationally demanding and require estimation of a large number of parameters. Another approach is to simultaneously fit an SDE model to different horizons, thus addressing the issue of the large number of parameters to be estimated.[18] However, this method is still computationally burdensome and can only handle a limited amount of data. Hence, we claim that our SDE approach, although relying on a one-step-ahead

Table VII. The MAE for SDE models (8)–(13) and (14)–(16) and for the ARX-GARCH benchmarks. The ARX-GARCH benchmark is fitted specifically for each forecast horizon. The iterative ARX-GARCH-TN model is fitted to 1 h ahead data and then run iteratively until the desired horizon is reached.

	MAE for different horizons			
Models	1 h	4 h	12 h	24 h
ARX - GARCH	0.0505	0.0958	0.117	**0.111**
ARX - GARCH - TN - iterative	0.0575	0.109	0.135	0.143
Model (8)–(13)	**0.0471**	**0.0913**	**0.113**	0.122
Model (14)–(16)	0.0553	0.0976	0.113	0.115

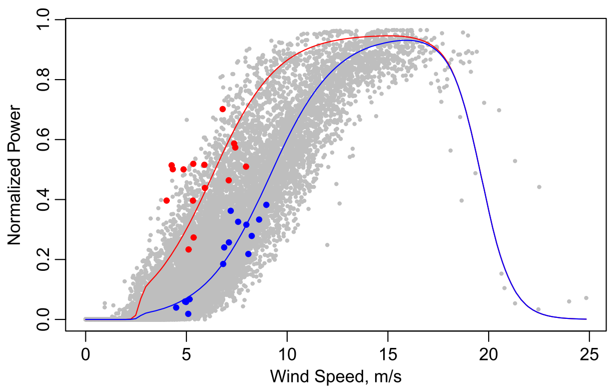

Figure 5. Normalized power plotted against wind speeds in gray both in hourly averages, similar to Figure 1. The blue and red lines are the average power curves from model (8)–(13) for the times when the blue and the red observations were recorded.

estimation procedure, can be considered a reasonable method for multi-horizon forecasting; it requires relatively few parameters and is comparable in computation time to standard time series models.

To conclude this section, we plot again (Figure 5) the two sets of 15 data points with 9 months between them that we showed in Figure 1 in the introduction. This time, however, instead of the power curve fitted with the local regression in Figure 1, we depict the adaptive power curve that is specific for each time period. The resulting power curves are shown in Figure 5. As it turns out, these power curves seem to much better capture the time-varying relationship between wind speed and wind power.

5. CONCLUDING REMARKS

Probabilistic forecasts of wind power generation provide useful input information to a variety of decision-making problems concerning management and trading of wind-generated electricity. Standard forecasts do not take into account the development of the uncertainty over time, making them of little use for a large class of operational tasks that involve time-dependent decisions. Such problems are related to the large-scale integration of wind power in electrical energy systems, where the dynamics of the uncertainty in the wind generation is of critical importance.

In this paper, we suggest a new framework for modeling, simulating and forecasting wind power generation for multiple horizons. The proposed framework provides wind power forecasts by combining a dynamic power curve with a stochastic model for wind speed based on stochastic differential equations. This results in an automatically bounded wind power process with time-varying uncertainty. The proposed model outperforms simple as well as complex benchmarks on an out-of-sample period of 1 year on horizons ranging from 1 to 24 h.

There are factors that could affect the speed-to-power conversion efficiency of the wind farm, but that we have not *explicitly* considered in the proposed SDE framework, for example, the age of the wind turbines, among others. Explicitly incorporating such factors into forecasting models is a highly difficult task in many aspects. In particular, we lack clear observational evidences on how these factors relate to wind turbine performance in the sense that many of these factors are not directly observable but rather are the consequence of other complex factors or phenomena, for instance, harsh weather conditions, weak effects, wind turbine components of poor manufacturing quality, faulty wind sensors, etc. Instead,

Wind Energ. 2017; **20**:33–44 © 2016 John Wiley & Sons, Ltd.
DOI: 10.1002/we

our SDE model (8)–(13) relies on an *implicit* modeling of some unobservable variables that are believed to influence the dynamical conversion from wind speed to power. In particular, the stochastic state variable Q_t is included for that specific purpose. Alternatively, for model (14)–(16), which does not include Q_t, one may think of estimating the parameters of this model in a recursive manner together, potentially with a forgetting factor. In recursive estimation frameworks, parameters are time-varying and can be updated as new observations become available.[40] The development of such estimation techniques for SDE models is the focus of present research.

One could also capture seasonal effects with our SDE framework, but this would require fitting models on time series spanning extensive periods of time, thereby increasing the computational cost of simulations, without necessarily increasing the overall forecasting skill of the model in the short term.

On a different front, using stochastic differential equations for forecasting lends itself well to several extensions. One such an extension is to consider spatio-temporal forecasting for capturing the interdependence between different sites in space and time. Another extension is to consider forecasting multiple outputs such as wind power generation, solar power generation and perhaps power load in a single complete model.

ACKNOWLEDGEMENTS

Det Strategiske Forskningsråd (DSF) 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 partly funded by the Research Center CITIES (no. 1035-00027B), which is also supported by DSF.

REFERENCES

1. The European Wind Energy Association, EWEA annual report 2012. Online, 2013. Available from: http://www.ewea.org/publications/reports/ewea-annual-report-2012/ (Accessed 15 April 2016).
2. Morales JM, Conejo AJ, Madsen H, Pinson P, Zugno M. *Integrating Renewables in Electricity Markets—Operational Problems*, International Series in Operations Research & Management Science, vol. 205. Springer: New York, 2014.
3. Pinson P. Wind energy: forecasting challenges for its operational management. *Statistical Science* 2013; **28**(4): 564–585.
4. Jeon J, Taylor JW. Using conditional kernel density estimation for wind power density forecasting. *Journal of the American Statistical Association* 2012; **107**(497): 66–79.
5. Wächter M, Milan P, Mücke T, Peinke J. Power performance of wind energy converters characterized as stochastic process: applications of the langevin power curve. *Wind Energy* 2011; **14**(6): 711–717.
6. Costa A, Crespo A, Navarro J, Lizcano G, Madsen H, Feitosa E. A review on the young history of the wind power short-term prediction. *Renewable and Sustainable Energy Reviews* 2008; **12**(6): 1725–1744.
7. Monteiro C, Bessa R, Miranda V, Botterud A, Wang J, Conzelmann G, *Wind power forecasting: state-of-the-art 2009*. Argonne National Laboratory (ANL), Argonne, Illinois, USA, 2009.
8. Foley AM, Leahy PG, Marvuglia A, McKeogh EJ. Current methods and advances in forecasting of wind power generation. *Renewable Energy* 2012; **37**(1): 1–8.
9. Pinson P, Chevallier C, Kariniotakis GN. Trading wind generation from short-term probabilistic forecasts of wind power. *IEEE Transactions on Power Systems* 2007; **22**(3): 1148–1156.
10. Bludszuweit H, Domínguez-Navarro JA. A probabilistic method for energy storage sizing based on wind power forecast uncertainty. *IEEE Transactions on Power Systems* 2011; **26**(3): 1651–1658.
11. Matos MA, Bessa RJ. Setting the operating reserve using probabilistic wind power forecasts. *IEEE Transactions on Power Systems* 2011; **26**(2): 594–603.
12. Bremnes JB. Probabilistic wind power forecasts using local quantile regression. *Wind Energy* 2004; **7**(1): 47–54.
13. Møller JK, Nielsen HA, Madsen H. Time-adaptive quantile regression. *Computational Statistics & Data Analysis* 2008; **52**(3): 1292–1303.
14. Pinson P, Madsen H. Ensemble-based probabilistic forecasting at horns rev. *Wind Energy* 2009; **12**(2): 137–155.
15. Trombe PJ, Pinson P, Madsen H. A general probabilistic forecasting framework for offshore wind power fluctuations. *Energies* 2012; **5**(3): 621–657.
16. Pinson P, Madsen H, Nielsen HA, Papaefthymiou G, Klöckl B. From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind energy* 2009; **12**(1): 51–62.
17. Bessa RJ, Miranda V, Botterud A, Zhou Z, Wang J. Time-adaptive quantile-copula for wind power probabilistic forecasting. *Renewable Energy* 2012; **40**(1): 29–39.

18. Møller JK, Pinson P, Madsen H, *Probabilistic forecasts of wind power generation by stochastic differential equation models*, Technical University of Denmark, Kgs. Lyngby, Denmark, 2013. Available from: http://www.statistics.gov.hk/wsc/STS019-P5-S.pdf (Accessed 15 April 2016).
19. Anahua E, Barth St, Peinke J. Markovian power curves for wind turbines. *Wind Energy* 2008; **11**(3): 219–232.
20. Gottschall J, Peinke J. Stochastic modelling of a wind turbine's power output with special respect to turbulent dynamics. *Journal of Physics: Conference Series* 2007; **75**: 1–8.
21. Björk T. *Arbitrage Theory in Continuous Time*, Oxford Finance Series. Oxford University Press: Oxford, United Kingdom, 2009.
22. Mikosch T. *Elementary Stochastic Calculus: With Finance in View*, Vol. 6. World Scientific: Singapore, 1998.
23. Van Kampen NG. *Stochastic Processes in Physics and Chemistry*, North-Holland Personal Library, vol. 1. Elsevier: Amsterdam, 1992.
24. Adomian G. *Nonlinear Stochastic Systems Theory and Application to Physics*, Vol. 46. Springer: Heidelberg, Germany, 1988.
25. Iversen EB, Morales JM, Møller JK, Madsen H. Probabilistic forecasts of solar irradiance by stochastic differential equations. *Environmetrics* 2014; **25**(1): 152–164.
26. Iversen EB, Morales JM, Møller JK, Madsen H. Short-term probabilistic forecasting of wind speed using stochastic differential equations. *International Journal of Forecasting* 2015. DOI: 10.1016/j.ijforecast.2015.03.001.
27. Øksendal B. *Stochastic Differential Equations: An Introduction with Applications*. Springer: Heidelberg, Germany, 2010.
28. Jazwinski AH. *Stochastic Processes and Filtering Theory*. Courier Dover Publications: New York, USA, 2007.
29. Kristensen NL, Madsen H, Jørgensen SB. Parameter estimation in stochastic grey-box models. *Automatica* 2004; **40**(2): 225–237.
30. Box GEP, Cox DR. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)* 1964; **26**(2): 211–252.
31. Møller JK, Madsen H, *From state dependent diffusion to constant diffusion in stochastic differential equations by the Lamperti transform*, Technical University of Denmark, Kgs. Lyngby, Denmark, 2010. Available from: http://orbit.dtu.dk/fedora/objects/orbit:82985/datastreams/file_5175845/content.
32. Juhl R, Kristensen NR, Bacher P, Mø ller JK, Madsen H. CTSM-R user guide. Technical University of Denmark, 2013.
33. Ljung L. *System Identification*, Wiley Encyclopedia of Electrical and Electronics Engineering. John Wiley & Sons, Inc.: New Jersey, USA, 2001.
34. Källén E, *Hirlam documentation manual; 1996. system 2.5*, Norrkøping, Sweden, 1996. Available from: http://www.hirlam.org/ (Accessed 15 April 2016).
35. Duran MJ, Cros D, Riquelme J. Short-term wind power forecast based on ARX models. *Journal of Energy Engineering* 2007; **133**(3): 172–180.
36. Taylor JW, McSharry PE, Buizza R. Wind power density forecasting using ensemble predictions and time series models. *IEEE Transactions on Energy Conversion* 2009; **24**(3): 775–782.
37. Lau A, McSharry P. Approaches for multi-step density forecasts with application to aggregated wind power. *The Annals of Applied Statistics* 2010; **4**(3): 1311–1341.
38. Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* 2007; **102**(477): 359–378.
39. Schefzik R, Thorarinsdottir TL, Gneiting T. Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statistical Science* 2013; **28**(4): 616–640.
40. Madsen H. *Time Series Analysis*, Vol. 72. CRC Press: Boca Raton, USA, 2008.