Journal of Forecasting, *J. Forecast.* (2016)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/for.2367

Probabilistic Forecasts of Wind Power Generation by Stochastic Differential Equation Models

JAN KLOPPENBORG MØLLER,* MARCO ZUGNO AND HENRIK MADSEN

DTU Compute, Matematiktorvet, Technical University of Denmark, DK-2800 Lyngby, Denmark

ABSTRACT

The increasing penetration of wind power has resulted in larger shares of volatile sources of supply in power systems worldwide. In order to operate such systems efficiently, methods for reliable probabilistic forecasts of future wind power production are essential. It is well known that the conditional density of wind power production is highly dependent on the level of predicted wind power and prediction horizon. This paper describes a new approach for wind power forecasting based on logistic-type stochastic differential equations (SDEs). The SDE formulation allows us to calculate both state-dependent conditional uncertainties as well as correlation structures. Model estimation is performed by maximizing the likelihood of a multidimensional random vector while accounting for the correlation structure defined by the SDE formulation. We use non-parametric modelling to explore conditional correlation structures, and skewness of the predictive distributions as a function of explanatory variables. Copyright © 2016 John Wiley & Sons, Ltd.

KEY WORDS non-linear forecasting; state space models; stochastic differential equations; wind power; probabilistic forecasting

INTRODUCTION

In 2013 wind power accounted for more than 33% of the total demand for electricity in Denmark, and in December (2013) about 53% of the electricity load was covered by wind power. Integrating this world-leading amount of wind power into the power systems calls for state-of-the-art methods for wind power forecasting. Methods for generating point forecasts of wind power for operational decision-making have been used since 1995 (Madsen *et al.*, 1995). However, the solution of the decision-making problems in electricity markets typically requires the full predictive density of future wind power production, and in some cases also its correlation structure (Morales *et al.*, 2014). A literature review with a focus on methods for point predictions and with some focus also on the meteorological aspects can be found in Giebel *et al.* (2011) and Costa *et al.* (2008). In practice, forecasts are needed for real-time monitoring, operation scheduling, production and maintenance planning, and energy trading (cf. Morales *et al.*, 2014). The corresponding lead times range from the very short term (say, 5 minutes) and up to the medium range (say, up to 5–7 days).

Still, even after about 20 years of research, wind power forecasting remains a challenge from a statistical point of view. This is firstly due to the nonlinear and double-bounded nature of wind power production (Pinson, 2012). Nonlinearity is a result of the sigmoidal dependence between wind power production and wind speed as well as, possibly, direction. In addition, nonlinearities may be present when describing both the conditional mean and the conditional variance (Trombe *et al.*, 2012). Finally, the process should generally be seen as non-stationary, with successive periods exhibiting different dynamics. In a probabilistic framework, conditional predictive densities may be generated in a parametric framework, for instance based on (censored) Gaussian, Beta, or mixtures of distributions (Pinson, 2012; Trombe *et al.*, 2012), or alternatively in a non-parametric framework, e.g. using quantile regression for modelling and predicting a set of quantiles with varying nominal proportions. Because of the complexity of the predictive distribution for long look-ahead times, the non-parametric framework is often preferred for lead times from hours to a few days ahead, e.g. the quantile regression approach of Nielsen *et al.* (2006) and time-adaptive quantile regression of Møller *et al.* (2008).

To further describe the interdependence between marginal predictive densities for a set of future lead times in a multivariate framework, Pinson *et al.* (2009) proposed to issue trajectories of wind power generation where the interdependence structure is modelled with a Gaussian copula. Such modelling then simplifies to the tracking of a time-varying covariance matrix. Alternatively, trajectories of wind power generation can be obtained as in Nielsen *et al.* (2006) by nonlinear transformation of ensemble forecasts of relevant meteorological variables (see Leutbecher and Palmer, 2008, for an overview of ensemble forecasting), hence accounting for the dynamic interdependence structure

*Correspondence to: Jan Kloppenborg Møller, DTU Compute, Matematiktorvet, Technical University of Denmark Building 303B, DK-2800 Lyngby, Denmark. E-mail: jkmo@dtu.dk

in the inherent uncertainty of meteorological forecasts used as input to wind power forecasting. These approaches are based on either static models or approaches with a simple tracking of the parameters, and hence they give a fairly static view of the interdependence structure.

In this paper, which is a further elaboration of the work presented in Møller *et al.* (2013) (proofs and comparison with benchmark models are included in the present work), a new method based on stochastic differential equations (SDEs) is suggested. The advantage of this method is that by selecting a proper specification and parametrization of the SDE, basically all the required properties, like nonlinearities, time variation, non-stationarity, double-bounded variations and varying quantiles, can be described. Furthermore, we demonstrate that this approach also offers a flexible description of the variation of interdependence structure of the predictions errors. Compared to the competing approaches described above, our method requires a significantly smaller number of parameters.

In our view, SDEs are a more suitable framework to describe the physics of wind power production than classical time series approaches. Indeed, the mechanics of dynamical systems are usually formalized as systems of differential equations. In the specific setting of wind power forecast, this implies that wind speed and wind direction can be included based on a mechanistic understanding of the problem. The results presented in this paper indicate that the combination of SDE models and multi-horizon forecasting has great potential.

Following this introductory section, the general framework is described in the next section, while the dataset used for the empirical investigation is introduced in the third section. Multivariate predictive densities issued in a Gaussian framework (possibly after some transformation), and with various modelling of the covariance structure, are presented in the fourth section. These Gaussian predictive densities are used as benchmark models. The fifth section gives a very short description of the general SDE framework used in the present work. The new method, which is the key contribution in the present work (in particular, Theorems 1 and 2), based on SDEs is presented in the sixth section, together with a number of candidates for the parametrization of the SDEs. Finally, numerical results are presented and compared in the seventh section, and concluding remarks are given in the eighth section.

GENERAL FRAMEWORK

We consider the problem of wind power forecasting with a look-ahead time of up to 2 days. As further described in the next section, a 48-hour prediction of wind power production is issued every 6th hour. Within this setup the observed wind power production (average of 1 hour) is considered as a realization of a 48-dimensional random variable, and hence the density of the observation is calculated in a multivariate setting.

After normalizing with the installed capacity, the 'true' underlying distribution will be contained in the hypercube $[0, 1]^n$, where n is the dimension of the variable. Here we consider forecast horizons $h = \{1, \ldots, 48\}$, i.e. $n = 48$.

Rather than testing different parametric distributions, the scope of this paper is to demonstrate the improvements that can be achieved by considering SDEs. This implies that we will not apply transformation to the original data, even though this might be beneficial. Hence we assume that the general distribution can be approximated by a multivariate normal distribution:

$$\boldsymbol{Y}_i \sim N(\hat{\boldsymbol{Y}}_i, \boldsymbol{\Sigma}_i) \tag{1}$$

where the index i refers to a realization of the 48-dimensional variable under some conditions (determined by the specific conditions, e.g. wind speed, and hence the index i), $\hat{\boldsymbol{Y}}_i$ and $\boldsymbol{\Sigma}_i$ are functions of meteorological forecasts and look-ahead time. In this setting we use point predictions provided by the widely used wind power prediction tool (WPPT) (see Madsen *et al.*, 1998; WPPT, 2012), which uses meteorological forecasts and local data as input for predicting wind power production, but provide only a point forecast in terms conditional expectation. The focus of this work is on describing the structure of the residuals. The 48-dimensional WPPT forecast issued for series i is denoted $\tilde{\boldsymbol{p}}_i = [\tilde{p}_{i,1}, \ldots, \tilde{p}_{i,48}]$, and a natural first approach would be to select

$$\hat{\boldsymbol{Y}}_i = \tilde{\boldsymbol{p}}_i; \quad \boldsymbol{\Sigma}_i = \boldsymbol{F}(\tilde{\boldsymbol{p}}_i, \boldsymbol{h}) \tag{2}$$

where $\boldsymbol{h} = [1, \ldots, 48]^T$ denotes the forecast horizons and $\boldsymbol{F}$ is a matrix function (with the natural restrictions of positive definiteness). We use these types of models as benchmark models, since the structure is similar to the SDE models presented in the fifth section. In the fourth section the covariance structure is obtained by a direct parametrization of $\boldsymbol{F}$, while $\boldsymbol{F}$ and $\hat{\boldsymbol{Y}}$ are obtained as the solution to SDEs in the fifth and sixth sections.

DATA

The data for this study consist of hourly averages of wind power production from the Klim wind power plant located in the northern part of Denmark. The dataset covers the period March 2001 through April 2003, and the total number

of 48-dimensional observations is about 3200. Production data are normalized with the installed capacity (21 MW). Every 6th hour a new forecast based on a 48-hour weather forecast is issued, i.e. a single observation in our setting consists of a 48-hour forecast and the corresponding (48) measurements. This implies that for every measurement there are up to seven forecasts issued at different times, but the combined set of 48-dimensional input vector and observation is unique.

Initially, all observations with missing data points are removed from the dataset, leaving a total of 2593 (48 dimensional) observations. The SDE estimation is computationally demanding and it was therefore decided to base the estimation on 150 observations only. The number 150 is quite arbitrary, but represents a compromise between the need for sufficient data to identify and estimate parameters and computation time. With the relatively low number of observations, we will need to ensure that the training set spans the different situations in the complete dataset. The selection algorithm below should be seen as a way to ensure this. Situations with high expected power are quite sparse and hence the algorithm aims at including these in the training dataset. The algorithm could be seen as a sort of experimental design, except that we cannot perform new experiments, but we simply have to choose from the total pool of data. There is also some resemblance with the way data are discarded in Møller *et al.* (2008).

The training set is chosen to span the different situations seen in data, according to the following algorithm:

1. Find the observation from the pool (initially the 2593 valid observations) of allowed observations with the largest average forecasted power over the 48-hour horizon ($\mu_{\max}$).
2. Calculate the numbers $\mu_j = \mu_{\max} \cdot \frac{j}{10}$, $j = \{10, \ldots, 0\}$.
3. For each μ_i choose the observation with average forecasted power closest to μ_i, while excluding all overlapping (in time) observations in each step.
4. Repeat steps 1–3 until the desired number of observations (150) has been found.

The observations not included in the training set are used as a test set. Because of the overlaps introduced by the use of a 48-hour horizon with a 6-hour sampling period, the test set set is further divided into two sets, and there are in total three sets which are used in this work, namely:

Train: observations chosen according to the algorithm described above (150 observations).
Test 1: observations with an overlap (in time) with the training set (1511 observations).
Test 2: observations with no overlap with the training set (932 observations).

The training set will be used for estimation of parameters, while the resulting model will be evaluated on the two test sets. Here overlap in time implies that observations are reused in different sets, but individual measurements will always refer to different prediction horizons and hence predicted power. The situation is illustrated in Figure 1, for each time there are several observations, but with different lead times.

The distribution of predicted power is rather different for the three sets considered here (Figure 2). In the complete dataset, most of the observations of normalized predicted wind power production are close to zero and very few points close to one. The training set has, by construction, a more even distribution of data, while Tests 1 and 2 include mostly observations close to zero. This is an effect of the sampling algorithm, and the scope of the sampling is that general inference is possible even with a relative small number of observations.

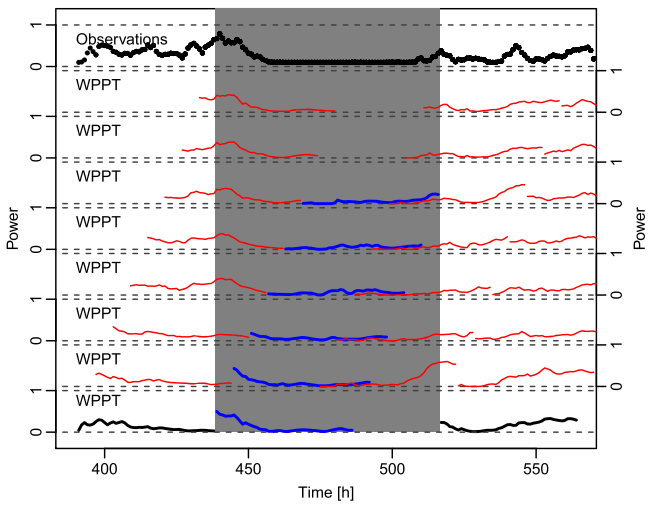

Figure 1. Illustration of training set (black lines), test1 (blue lines), test2 (red lines) and observations (black dots). Forecast sequences entirely in the grey area have no overlap with the training set

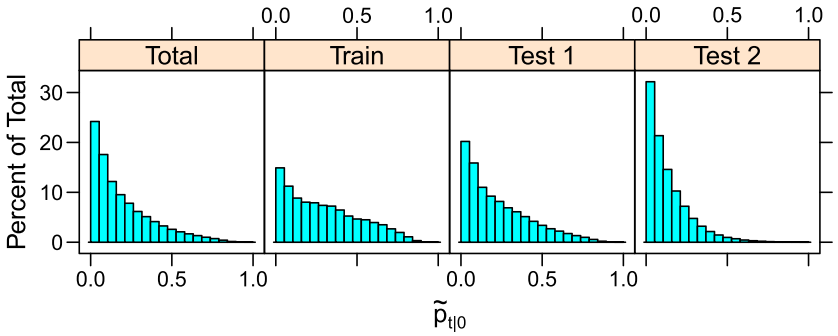

Figure 2. Histogram of predicted wind power production by WPPT for the complete dataset, training set and the two test sets

BENCHMARK MODELS

To study the effect of SDE models, later on we introduce benchmark models of the form

$$\boldsymbol{Y}_i | \tilde{\boldsymbol{p}}_i \sim N(\tilde{\boldsymbol{p}}_i, \boldsymbol{\Sigma}_i) \tag{3}$$

where $\boldsymbol{Y}_i \in \mathbb{R}^{48}$ is the observed power, $\tilde{\boldsymbol{p}}_i \in \mathbb{R}^{48}$ is the predicted power and $\boldsymbol{\Sigma}_i \in \mathbb{R}^{48\times 48}$ is the covariance matrix of $\boldsymbol{Y}_i$. We consider the following parametrization of $\boldsymbol{\Sigma}_i$:

$$\boldsymbol{\Sigma}_i = \boldsymbol{\sigma}_i \boldsymbol{R} \boldsymbol{\sigma}_i^T \tag{4}$$

where $\boldsymbol{\sigma}_i$ are the marginal standard deviations of series i and $\boldsymbol{R}$ is the correlation matrix. In the following, models with different structures (with increasing complexity) of $\boldsymbol{\sigma}_i$ and $\boldsymbol{R}$ are introduced. Since the used point forecast is a state-of-the-art point predictor (only providing the expected value of wind power production) the benchmark model can be viewed as a parametrization of the uncertainty for the point estimate. Although inspired by Gaussian approaches for describing the uncertainty, the choice of benchmark models is primarily attributed to the similarity with SDE formulation, and it allows us to study the potential benefit of using SDEs to construct the covariance matrix.

Estimation, testing and validation

The log-likelihood contribution from one observation (ignoring constant terms) is

$$l(\boldsymbol{\theta}; \boldsymbol{y}_i) = -\frac{1}{2}|\boldsymbol{\Sigma}_i(\boldsymbol{\theta})| - \frac{1}{2}(\boldsymbol{y}_i - \tilde{\boldsymbol{p}}_i)^T \boldsymbol{\Sigma}_i(\boldsymbol{\theta})^{-1}(\boldsymbol{y}_i - \tilde{\boldsymbol{p}}_i) \tag{5}$$

where $|\cdot|$ is the determinant. Now let the observations be indexed using the index set I, and let the training set and the test sets be indexed by I_{train}, I_{test1}, I_{test2}. The maximum likelihood estimate of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left(- \sum_{i \in I_{\text{train}}} l(\boldsymbol{\theta}; \boldsymbol{y}_i) \right) \tag{6}$$

and the corresponding log-likelihood of the training set is

$$l(\text{train}) = \sum_{i \in I_{\text{train}}} l(\hat{\boldsymbol{\theta}}; \boldsymbol{y}_i) \tag{7}$$

All models are fitted by optimizing the log-likelihood function, and the result of the estimation is shown in Table I. All the presented models are nested and the likelihood ratio test for model comparison is therefore relevant and presented in the table. In addition, cross-validation is performed by measuring the log-likelihood (or log density) of each of the test sets. This means that we calculate the sum in equation (7), but over the index sets I_{test1} and I_{test2}. These numbers are denoted $l(\text{test1})$, and $l(\text{test2})$.

All model extensions are significant, and in addition all models give improvements on the test sets. The parametrization of the models is explained next.

Table I. Likelihood table for the four different benchmark models

	d.f.	l(train) $n = 150$	$p-$value	l(test1) $n = 1511$	l(test2) $n = 932$	l(test1) + l(test2) $n = 2443$
BM 0	1	2709		35,050	27,933	62,983
BM 1	2	7361	< 0.0001	81,312	56,303	137,615
BM 2	7	8393	< 0.0001	88,690	67,777	156,467
BM 3	12	8419	< 0.0001	88,692	67,853	156,546

Note: The table contain the log-likelihood (l) for each of the models and each of the sets considered. d.f. refers to the number of parameters in each model; n refers to the number of observations in each set. The models are all nested and the p-values refer to likelihood ratio tests.

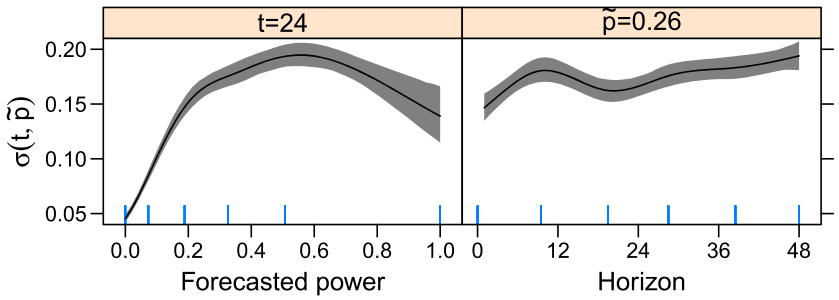

Figure 3. Functional relationship between uncertainty ($\sigma(\cdot,\cdot)$) and forecasted power (left panel) and look-ahead time (right panel) for fixed horizon ($t = 24$ h) and fixed forecasted power ($\tilde{p} = 0.26$, the median of forecasted power in the training set), respectively. Rugs indicate the knot placement (for the spline basis functions) and the grey areas indicate 95% confidence intervals of the relationship based on 10,000 bootstrap estimates of the curves

Table II. Parameter values for the benchmark models

	BM 0	BM 1	BM 2	BM 3
σ_0	0.166	0.164	0.049	0.04
	(0.163, 0.169)	(0.157, 0.17)	(0.046, 0.052)	(0.036, 0.044)
λ		0.161	0.161	0.169
		(0.148, 0.175)	(0.148, 0.176)	(0.154, 0.184)
f	0	0	Spline	Spline
g	0	0	0	Spline

Note: Numbers in parentheses give 95% confidence intervals for the parameters.

Parametrization of the benchmark models

The benchmark models are of the form

$$\boldsymbol{\sigma}_{ij}(\tilde{p}_{i,j}, j) = \sigma_0 e^{f(\tilde{p}_{i,j})+g(j)} \tag{8}$$

$$\boldsymbol{R}_{lm} = e^{-\lambda\cdot|l-m|} \tag{9}$$

where i refers to the observation, $j \in \{1,\ldots,48\}$ refers to the horizon within this observation, $\boldsymbol{R}_{lm}$ is the correlation between forecast horizons l and m, and the nonlinear functions f and g, which describe the dependency on explanatory variables, the predicted wind power and horizon, are modelled as linear combinations of natural cubic spline basis functions:

$$f(\tilde{p}_{i,j}) = B_1^p(\tilde{p}_{i,j})\alpha_1 + \ldots + B_k^p(\tilde{p}_{i,j})\alpha_k \tag{10}$$

$$g(j) = B_1^t(j)\beta_1 + \ldots + B_q^t(j)\beta_q \tag{11}$$

where k and q are the number of degrees of freedom for the chosen basis. The knot placements defining the natural B-spline basis functions are chosen as (0.2,0.4,0.6,0.8)-quantiles of forecasted power in the training set (see Figure 3).

The complexity of the benchmark models is indicated in Table II. Model BM 0 is the baseline model (constant variance and no correlation). Introducing a first-order Markov correlation (BM 1) results in a very large improvement in the likelihood function, stressing that it is very important to include correlation structure in the model. The difference between BM 0 and BM 1 is illustrated in Figure 4, for the hour-to-hour prediction error, where it is seen that the prediction region of BM 1 captures the observed data much better than BM 0. The one-step-ahead correlation of the observed residuals is also clear from the figure.

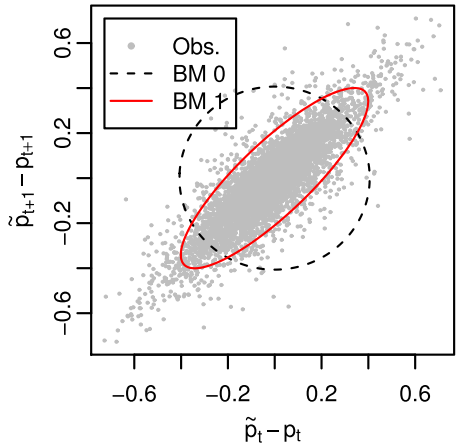

Figure 4. Observed prediction errors for the benchmark model and 95% prediction regions for the prediction errors for models BM 0 and BM 1

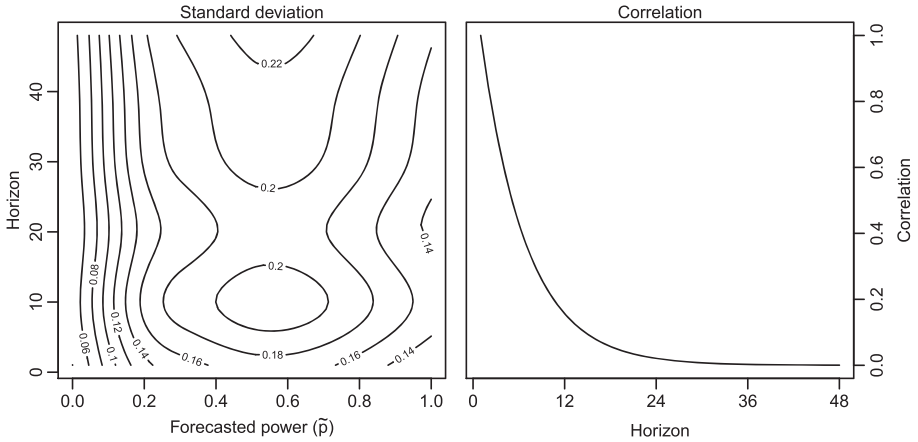

Figure 5. Standard deviation (left panel) as a function of forecasted power and horizon for Model BM 3

It is reasonable to assume that the variance scales with the predicted power (BM 2). Again a very large improvement in the likelihood is seen and hence the hypothesized relation between forecasted power and variance is supported by the likelihood ratio test. It is also reasonable to assume that the variance is linked to the prediction horizon (BM 3). Again, the knot placement defining the spline basis functions is shown in Figure 3.

All model extensions are significant on levels less than 0.01% (Table I). More complicated models also perform better on the test set; however, the gain between BM 2 and BM 3 is small. This conclusion is supported by Figure 3, where the functional relationship between standard deviation and forecasted power and horizon is drawn; the standard deviation is seen to vary strongly with forecasted power, while the dependence with horizon is small. It might be possible to describe the functional relationships in Figure 3 by fewer parameters than those used by the non-parametric description (the spline functions), but we will not consider this simplification exercise in this work. We can, however, conclude from Table I that some functional relationship is present.

The confidence bands in Figure 3 are constructed by considering the vector of parameters $\boldsymbol{\theta} = [\sigma_0, \lambda, \alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_l]^T$ as a Gaussian random vector with mean equal to $\hat{\boldsymbol{\theta}}$ and covariance matrix equal to the negative inverse Hessian of the log-likelihood evaluated at the optimum ($\hat{\boldsymbol{\theta}}$). The confidence intervals are calculated from point-wise quantiles of 10,000 curves drawn from this distribution. The confidence intervals are quite narrow and support the likelihood ratio test-based conclusion, namely that both horizon and forecasted power should be included in the model. The variability of the functional relationship between forecasted power and marginal variance is small for forecasted power close to zero, while the uncertainty is large for forecasted power close to 1, reflecting that there are more data for low forecasted power than for large forecasted power.

A contour plot of standard deviation as a function of both forecasted power and horizon is given in Figure 5, i.e. $\sigma(\tilde{p}, h)$ (see equation 8). The overall relationship is that the standard deviation is small for forecasted power close to 0 or 1, which, owing to the sigmoid nature of the power curve, is as expected. The surface also shows a local minimum at a horizon of approximately 20 hours and forecasted power around 0.55. The local minimum at 20 hours might be explained by overparametrization, but it might also indicate that the meteorological forecast is more accurate at 24 hours than at 12–16 hours. The maximum around forecasted power at 0.55 is as expected, since the steepest relation between wind speed and power output is around forecasted power close to 50% of installed

capacity, i.e. small changes in wind speed lead to large changes in produced power and therefore larger uncertainty (see also Lange, 2005).

Figure 5 also shows the correlation as a function of time lag. The correlation between neighbouring points is strong; e.g. it is close to about 0.5 for points 6 hours apart, and there is effectively no correlation between points more than 24 hours apart.

For the SDE-based models presented in the following two sections, the point predictions and correlation structure are calculated based on stochastic differential equations. The distributional assumption is, however, still the Gaussian density and hence the models presented in this section (in particular BM 3) serve as relevant benchmarks for the SDE models to be presented in the following two sections.

CONTINUOUS–DISCRETE TIME STATE-SPACE FORMULATION

Consider the stochastic differential equation (SDE)

$$\mathrm{d}x_{i,t} = f(x_{i,t}, u_{i,t}, t)\mathrm{d}t + \sigma(x_{i,t}, u_{i,t}, t)\mathrm{d}w_{i,t} \tag{12}$$

where $u_{i,t}$ is a known input (here $u_{i,t} = \tilde{p}_{i,t}$). The first term on the right-hand side of equation (12) is often referred to as the drift term, while the second term is referred to as the diffusion term. $\mathrm{d}w_{i,t}$ in the diffusion term is the differential of the standard Wiener process (Brownian motion); see, for example, Øksendal (2003) for details on properties of SDEs. The formulation (12) is short for its integral representation, which is only uniquely defined up to the integral interpretation; here we use the Itô interpretation, i.e.

$$x_{i,t} = x_{i,0} + \int_0^t f(x_{i,t}, u_{i,t}, t)\mathrm{d}t + \int_0^t \sigma(x_{i,t}, u_{i,t}, t)\mathrm{d}w_{i,t} \tag{13}$$

The Itô interpretation implies that the second term in equation (13) is a martingale and hence it would be natural to require that the drift term is zero for $x_{i,t} = \tilde{p}_{i,t}$, implying $f(\tilde{p}_{i,t}, \tilde{p}_{i,t}, t) = 0$, $\forall t$. Furthermore $\sigma(x_{i,t}, u_{i,t}, t)$ should depend on $x_{i,t}$ in such a way that the natural restriction ($x_{t,i} \in (0, 1)$) is fulfilled (implying $\sigma(0, \cdot) = \sigma(1, \cdot) = 0$). The precise structures are described later on in the following section. As discussed above, we aim at a second-order moment representation of $\boldsymbol{Y}_i$ and therefore of $\boldsymbol{x}_i = [x_{i,1}, \ldots, x_{i,48}]$. In general, when faced with an SDE with state-dependent diffusion one should try to transform the SDE into another SDE with state-independent diffusion (Iacus, 2008), such that the solution is better approximated by a Gaussian distribution. This applies to simulation (Iacus, 2008), but also for obtaining second-order moment approximations (Baadsgaard *et al.*, 1997).

In order to find the second-order moment representation, the process (12) can be transformed into another process given by

$$\mathrm{d}z_{i,t} = \tilde{f}(z_{i,t}, u_{i,t}, t)\mathrm{d}t + \sigma_0(u_{i,t}, t)\mathrm{d}w_{i,t} \tag{14}$$

where the diffusion term of equation (14) does not depend on the state ($z_{i,t}$). The solution to equation (14) is a stochastic process and we can approximate the second-order moment representation by, for example, the prediction step of extended Kalman filter-inspired methods (see, for example, Jazwinski, 1970). The transformation $z_{i,t} = \psi(x_{i,t})$ is often referred to as the Lamperti transform (Iacus, 2008).

The SDE (12) is constructed such that the state space is equal to the natural state space (i.e. $x_{i,t} \in (0, 1)$), implying that the variance on the boundaries is equal to zero. Since there will be some uncertainty or variance regardless of the value of the forecast, we add a discrete time measurement equation:

$$Y_{i,j} = h(x_{i,j}) + e_{i,j} \tag{15}$$

where $e_{i,j}$ is an autocorrelated Gaussian process with covariance matrix $\boldsymbol{\Sigma}_i^{ee}$. If we choose $h(\cdot)$ as the identity, the complete multivariate model is now given by

$$\boldsymbol{Y}_i = \boldsymbol{\psi}^{-1}(\boldsymbol{z}_i) + \boldsymbol{e}_i \tag{16}$$

and the solution is approximated by

$$\boldsymbol{Y}_i \sim N(\boldsymbol{\psi}^{-1}(\boldsymbol{z}_i), \boldsymbol{C}_i^T \boldsymbol{P}_i \boldsymbol{C}_i + \boldsymbol{\Sigma}_i^{ee}) \tag{17}$$

where $\boldsymbol{C}_i = \frac{\partial \boldsymbol{\psi}^{-1}}{\partial \boldsymbol{z}_i}$, and $\boldsymbol{P}_i$ is the conditional covariance matrix of $(z_{i,1}, \ldots, z_{i,n})^T$. In the following section the mean and covariance functions are obtained from the same type of input as in the benchmark models, but the mean value and covariance matrix of the SDE solutions are based on integration of the SDE (14). The continuous–discrete time stochastic state-space model consists of the continuous time SDE in equation (12) or (14) and the discrete time measurement equation (15) or (16). A simple example of a continuous–discrete time stochastic differential equation

model might be found in Møller *et al.* (2011). The main advantage of a continuous time description of the dynamics is that natural restrictions on the state space can be build into equation (12), and results on correlation structure and drift can be derived directly from these formulations.

THE STOCHASTIC DIFFERENTIAL EQUATION MODEL

The continuous–discrete time stochastic state-space model introduced in the previous section consists of an SDE describing the dynamics, and an observation equation. The continuous time SDE has the form of equation (12). Initially we want the deterministic part of the model to have the stationary solution $x_{i,t} = \tilde{p}_{i,t}$ (this will be relaxed in the final model). This can be achieved by choosing

$$f(x_{i,t}, \tilde{p}_{i,t}, t) = -\theta_t(x_{i,t} - \tilde{p}_{i,t}) \tag{18}$$

where θ_t is allowed to depend on time and forecasted power. As indicated in equation (12) the diffusion term is allowed to depend on the state and the input, and since we want to kill the diffusion at the boundaries a minimal requirement is

$$\sigma(1, \tilde{p}_{i,t}, t) = \sigma(0, \tilde{p}_{i,t}, t) = 0 \tag{19}$$

The restrictions (19) are fulfilled for logistic type SDEs (e.g. Schurz, 2007). Here we select the special case

$$\sigma(x_{i,t}, \tilde{p}_{i,t}, t) = \sqrt{2\tilde{\sigma}(\tilde{p}_{i,t}, t)x_{i,t}(1 - x_{i,t})} \tag{20}$$

which evidently fulfils equation (19) (at least when $\tilde{\sigma}(\tilde{p}_{i,t}, t)$ is bounded). In addition to the restriction (19) we need a parametrization of equation (12) that allows for a stationary solution (this will be relaxed later on). equation (12) is a special case of the Pearson diffusion (see, for example, Forman and Sørensen, 2008; Iacus, 2008). A time-homogeneous version of equation (12), with drift term (18) and diffusion term (20) is

$$\mathrm{d}x_t = -\theta(x_t - \mu)\mathrm{d}t + \sqrt{2\theta a x_t(1 - x_t)}\mathrm{d}w_t \tag{21}$$

A stationary solution exists when (see Iacus, 2008)

$$\min(\mu, 1 - \mu) \geq a \tag{22}$$

Since in our case μ is a function of t, we will have to give a general function which fulfils equation (22). One way to obtain this is by choosing

$$a = \alpha\mu(1 - \mu) \tag{23}$$

with $\alpha \in [0, 1]$, with the advantage that a is differential w.r.t. μ.

For the time-inhomogeneous SDE we have $\mu_t = \tilde{p}_t$ and we choose

$$\tilde{\sigma}(\tilde{p}_{i,t}, t) = \theta_t a(\tilde{p}_{i,t}, t)\tilde{p}_{i,t}(1 - \tilde{p}_{i,t}) \tag{24}$$

with $a(\tilde{p}_{i,t}, t) \in (0, 1)$ for all t and $\tilde{p}_{i,t}$.

The SDE that forms the basis of the analysis in this paper is therefore

$$\mathrm{d}x_{i,t} = -\theta_{i,t}(x_{i,t} - \tilde{p}_{i,t})\mathrm{d}t + \sqrt{2\theta_{i,t}a_{i,t}\tilde{p}_{i,t}(1 - \tilde{p}_{i,t})x_{i,t}(1 - x_{i,t})}\mathrm{d}w_t \tag{25}$$

However, as previously mentioned, we want to transform equation (25) in such a way that the diffusion term of the transformed system is independent of the state. In order to do this we use the Lamperti transformation (e.g. Iacus, 2008):

$$z_{i,t} = \int \frac{1}{\sqrt{\xi(1 - \xi)}}\bigg|_{\xi = x_{i,t}} = \arcsin(2x_{i,t} - 1) \;\Rightarrow\; x_{i,t} = \frac{1}{2}(1 + \sin(z_{i,t})) \tag{26}$$

with $z_{i,t} \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.

The findings in the first part of this section are summarized in the following theorem.

Theorem 1. The Lamperti-transformed process corresponding to equation (25) is given by (with $\tilde{p}_{i,t} \in (0, 1)$)

$$\mathrm{d}z_{i,t} = \tilde{f}(z_{i,t}, \tilde{p}_{i,t}, t)\mathrm{d}t + \sqrt{2\theta_{i,t}a_{i,t}\tilde{p}_{i,t}(1 - \tilde{p}_{i,t})}\mathrm{d}w_t \tag{27}$$

with

$$\tilde{f}(z_{i,t}, \tilde{p}_{i,t}, t) = \frac{2\left(-\theta_{i,t}(\frac{1}{2}(1+\sin(z_{i,t})) - \tilde{p}_{i,t}) + \frac{1}{2}\sin(z_{i,t})a_{i,t}\theta_{i,t}\tilde{p}_{i,t}(1-\tilde{p}_{i,t})\right)}{\cos(z_{i,t})} \tag{28}$$

Moreover, the second-order moment representation of the linearized process can be solved for $a_{i,t} \in [0, 2)$.

Proof. Applying Itô's lemma (see, for example, Øksendal, 2003) gives the transformed state equation

$$\begin{aligned} \mathrm{d}z_{i,t} =& \frac{\left(-\theta_{i,t}(x_{i,t} - \tilde{p}_{i,t}) - \frac{1}{2}(1-2x_{i,t})a_{i,t}\theta_{i,t}\tilde{p}_{i,t}(1-\tilde{p}_{i,t})\right)}{\sqrt{x_{i,t}(1-x_{i,t})}}\mathrm{d}t \\ &+ \sqrt{2\theta_{i,t}a_{i,t}\tilde{p}_{i,t}(1-\tilde{p}_{i,t})}\mathrm{d}w_t \end{aligned} \tag{29}$$

$$\begin{aligned} =& \frac{2\left(-\theta_{i,t}(\frac{1}{2}(1+\sin(z_{i,t})) - \tilde{p}_{i,t}) + \frac{1}{2}\sin(z_{i,t})a_{i,t}\theta_{i,t}\tilde{p}_{i,t}(1-\tilde{p}_{i,t})\right)}{\sqrt{1-\sin^2(z_{i,t})}}\mathrm{d}t \\ &+ \sqrt{2\theta_{i,t}a_{i,t}\tilde{p}_{i,t}(1-\tilde{p}_{i,t})}\mathrm{d}w_t \end{aligned} \tag{30}$$

$$\begin{aligned} =& \frac{2\left(-\theta_{i,t}(\frac{1}{2}(1+\sin(z_{i,t})) - \tilde{p}_{i,t}) + \frac{1}{2}\sin(z_{i,t})a_{i,t}\theta_{i,t}\tilde{p}_{i,t}(1-\tilde{p}_{i,t})\right)}{\cos(z_{i,t})}\mathrm{d}t \\ &+ \sqrt{2\theta_{i,t}a_{i,t}\tilde{p}_{i,t}(1-\tilde{p}_{i,t})}\mathrm{d}w_t \end{aligned} \tag{31}$$

$$= \tilde{f}(z_{i,t}, \tilde{p}_{i,t}, t)\mathrm{d}t + \sqrt{2\theta_{i,t}a_{i,t}\tilde{p}_{i,t}(1-\tilde{p}_{i,t})}\mathrm{d}w_t \tag{32}$$

which proves equation (27).

The expected value $\hat{z}_{i,t}$ is given by the solution to

$$\frac{\mathrm{d}\hat{z}_{i,t}}{\mathrm{d}t} = \frac{2\left(-\theta_{i,t}(\frac{1}{2}(1+\sin(\hat{z}_{i,t})) - \tilde{p}_{i,t}) + \frac{1}{2}\sin(\hat{z}_{i,t})a_{i,t}\theta_t\tilde{p}_{i,t}(1-\tilde{p}_{i,t})\right)}{\cos(\hat{z}_t)} \tag{33}$$

Considering the limit $x_t \to 0$ $\left(z_t \to -\frac{\pi}{2}\right)$, we get

$$\lim_{\hat{z}_{i,t}\to-\frac{\pi}{2}} \frac{\mathrm{d}\hat{z}_{i,t}}{\mathrm{d}t} = \lim_{\hat{z}_{i,t}\to-\frac{\pi}{2}} \frac{2\left(-\theta_{i,t}(\frac{1}{2}(1+(-1)) - \tilde{p}_{i,t}) + \frac{1}{2}(-1)a_{i,t}\theta_{i,t}\tilde{p}_{i,t}(1-\tilde{p}_{i,t})\right)}{\cos(\hat{z}_{i,t})} \tag{34}$$

$$= \lim_{\hat{z}_{i,t}\to-\frac{\pi}{2}} \frac{2\theta_{i,t}\left(\tilde{p}_{i,t} - \frac{1}{2}a_{i,t}\tilde{p}_{i,t}(1-\tilde{p}_{i,t})\right)}{\cos(\hat{z}_{i,t})} \tag{35}$$

$$= \lim_{\hat{z}_{i,t}\to-\frac{\pi}{2}} \frac{2\theta_{i,t}\tilde{p}_{i,t}\left(1 - \frac{1}{2}a_{i,t}(1-\tilde{p}_{i,t})\right)}{\cos(\hat{z}_{i,t})} \tag{36}$$

$$> \lim_{\hat{z}_{i,t}\to-\frac{\pi}{2}} \frac{2\theta_{i,t}\tilde{p}_{i,t}\left(1 - \frac{1}{2}a_{i,t}\right)}{\cos(\hat{z}_{i,t})} = \infty \text{ for } a_{i,t} \in [0, 2) \tag{37}$$

and by similar calculations for $x_t \to 1$, we get

$$\lim_{\hat{z}_{i,t}\to\frac{\pi}{2}} \frac{\mathrm{d}\hat{z}_{i,t}}{\mathrm{d}t} < \lim_{\hat{z}_{i,t}\to\frac{\pi}{2}} \frac{-2\theta_{i,t}(1-\tilde{p}_{i,t})\left(1 - \frac{1}{2}a_{i,t}\right)}{\cos(\hat{z}_{i,t})} = -\infty \text{ for } a_{i,t} \in [0, 2) \tag{38}$$

which shows that equation (33) has a solution for $a_{i,t} \in [0, 2)$.

By linearization, inspired by the extended Kalman filter approach (see, for example, Jazwinski, 1970), we are able to find the variance ($P_{i,t}$) of $z_{i,t}$ as the solution to the ODE:

$$\frac{\mathrm{d}P_{i,t}}{\mathrm{d}t} = 2\left(\frac{\partial \tilde{f}}{\partial z}\right)\Bigg|_{z=\hat{z}_{i,t}} P_{i,t} + \tilde{\sigma}_{i,t}^2 \tag{39}$$

$$= -2\theta_{i,t}\frac{\sin(\hat{z}_{i,t}) + 1 - 2\tilde{p}_{i,t}\sin(\hat{z}_{i,t}) - a_{i,t}\tilde{p}_{i,t} + a_{i,t}\tilde{p}_{i,t}^2}{\cos^2(\hat{z}_{i,t})} P_{i,t} + 2\theta_{i,t}a_{i,t}\tilde{p}_{i,t}(1-\tilde{p}_{i,t}) \tag{40}$$

$$= -2\theta_{i,t}\frac{1 + \sin(\hat{z}_{i,t})(1-2\tilde{p}_{i,t}) - a_{i,t}\tilde{p}_{i,t}(1-\tilde{p}_{i,t})}{\cos^2(\hat{z}_{i,t})} P_{i,t} + 2\theta_{i,t}a_{i,t}\tilde{p}_{i,t}(1-\tilde{p}_{i,t}) \tag{41}$$

$$= 2A_{i,t}P_{i,t} + \tilde{\sigma}_{i,t}^2 \tag{42}$$

By similar reasoning, as in equations (37) and (38), it can be shown that $A_{i,t} < 0 \;\forall\; \tilde{p}_{i,t}$ when $a_{i,t} \in [0, 2)$.

Based on Theorem 1, the restriction of equation (23) is therefore relaxed and we consider the SDE formulation:

$$\mathrm{d}x_{t,i} = -\theta(x_{t,i} - \tilde{p}_{t|0,i})\mathrm{d}t + 2\sqrt{\theta\alpha\tilde{p}_{i,t}(1-\tilde{p}_{i,t})x_{t,i}(1-x_{t,i})}\mathrm{d}w_{t,i} \tag{43}$$

with $\alpha \in [0, 1)$; and, with $x_{t,i}$ given by equation (26), the Lamperti transform leads to the process

$$\mathrm{d}z_{t,i} = -\frac{\theta[x_{t,i} - \tilde{p}_{i,t} + \alpha\tilde{p}_{i,t}(1-\tilde{p}_{i,t})(1-2x_{t,i})]}{\sqrt{x_{t,i}(1-x_{t,i})}}\mathrm{d}t + 2\sqrt{\theta\alpha\tilde{p}_{i,t}(1-\tilde{p}_{i,t})}\mathrm{d}w_{t,i} \tag{44}$$

The transformed process includes a 'bias' ($\alpha\tilde{p}_{i,t}(1-\tilde{p}_{i,t})(1-2x_t)$) term introduced by the diffusion, and we consider a compensator for this bias. The resulting SDE is given by

$$\begin{aligned}\mathrm{d}x_{t,i} = &-\theta(x_{t,i} - \tilde{p}_{i,t} - c\,\tilde{p}_{i,t}(1-\tilde{p}_{i,t})(1-2x_{t,i}))\mathrm{d}t \\ &+ 2\sqrt{\theta\alpha\tilde{p}_{i,t}(1-\tilde{p}_{i,t})x_{t,i}(1-x_{t,i})}\mathrm{d}w_{t,i}\end{aligned} \tag{45}$$

where $c \geq 0$ is a constant. Again the results are given in a theorem.

Theorem 2. The second-order moment representation of the linearized Lamperti-transformed process corresponding to equation (45) can be solved for $\alpha \in [0, 1)$ and $c \geq 0$.

The complete solution (second-order moment representation of the Lamperti-transformed process) is given by (for given $z_{0|0}$)

$$\hat{z}_t = \arcsin\left\{\left(\int_0^t \theta(2\tilde{p}_{i,s} - 1)e^{\int_0^s g_u\mathrm{d}u}\mathrm{d}s + \sin(z_{0|0})\right)e^{-\int_0^t g_s\mathrm{d}s}\right\} \tag{46}$$

where

$$g_t = \theta_t[1 - 2(1-c_t)\alpha_t\tilde{p}_{i,t}(1-\tilde{p}_{i,t})] \tag{47}$$

and the covariance is given by (with $P_0 = 0$)

$$P_t = \int_0^t \tilde{\sigma}_s^2 e^{-2\int_0^s A_u\mathrm{d}u}\mathrm{d}s \cdot e^{2\int_0^t A_s\mathrm{d}s} \tag{48}$$

with

$$A_t = \frac{\partial \tilde{f}(z)}{\partial z}\Bigg|_{z=\hat{z}_t} \tag{49}$$

Proof. The proof is similar to that for Theorem 1 and is therefore omitted. It can be completed by computing the Lamperti-transformed process and considering the limits $\hat{z}_{i,t} \to \pm\frac{\pi}{2}$ (it suffices to consider $\alpha = 1$ and $\tilde{p}_{i,t} = \frac{1}{2}$).

Solving the linearized second-order moment equations gives the full solution (46)–(49).

In addition to the theorem above we use (correlation of the linearized process after stationarity is reached)

$$\rho_{t,t+s} = e^{\int_t^{t+s} A_u \mathrm{d}u} \tag{50}$$

to approximate the correlation structure.

These equations are solved by numerical integration (with $\Delta t = 0.001$) and the result is a mean value vector $\hat{\boldsymbol{z}}_i = [\hat{z}_{1,i}, \ldots, \hat{z}_{48,i}]$ and a covariance matrix $\boldsymbol{P}_i \in \mathbb{R}^{48\times 48}$ for the ith observation vector.

The observation equation

In addition to the SDE model we introduce the observation equation given by

$$\boldsymbol{Y}_i = \hat{\boldsymbol{x}}_i + \boldsymbol{e}_i \tag{51}$$

$$= \boldsymbol{h}(\hat{\boldsymbol{z}}_i) + \boldsymbol{e}_i \tag{52}$$

where $\hat{\boldsymbol{x}}_i$ is found by the integration described in the previous section and $\boldsymbol{e}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}^e)$. In our case the observation equation does not account for the inability to measure the exact power output, but rather model deficiencies introduced by the natural restrictions on the state space, which leads to zero variance at the boundary. We assume that the correlation structure of $\boldsymbol{e}_i$ is equal to the correlation structure introduced by the process (other parametrizations, like the first-order Markov correlation in the bench mark models, could have been chosen). Using the error propagation law and assuming that $\boldsymbol{Y}_i$ is Gaussian we get

$$\boldsymbol{Y}_i \sim N\left(\frac{1}{2}(1 + \sin(\hat{\boldsymbol{z}}_i)), \boldsymbol{C}_i \boldsymbol{P}_i \boldsymbol{C}_i + \boldsymbol{\Sigma}_i^e\right) \tag{53}$$

where $\boldsymbol{C}_i = \frac{\partial \boldsymbol{h}(\hat{\boldsymbol{z}}_i)}{\partial \hat{\boldsymbol{z}}_i} = \frac{1}{2}\mathrm{diag}(\cos(\hat{\boldsymbol{z}}_i))$ and $\boldsymbol{P}_i$ is constructed by combining equations (48) and (50). The covariance function of $\boldsymbol{e}_i$ is modelled by $\mathrm{cov}(e_{t_1,i}, e_{t_2,i}) = s^2\rho(e_{t_1,i}, e_{t_2,i})$, where s^2 is a constant to be estimated and $\rho(e_{t_1,i}, e_{t_2,i})$ is determined from the correlation of the covariance matrix $\boldsymbol{C}_i \boldsymbol{P}_i \boldsymbol{C}_i$ by

$$\rho(e_{t_1,i}, e_{t_2,i}) = \frac{(\boldsymbol{C}_i \boldsymbol{P}_i \boldsymbol{C}_i)_{t_1,t_2}}{\sqrt{(\boldsymbol{C}_i \boldsymbol{P}_i \boldsymbol{C}_i)_{t_1,t_1}(\boldsymbol{C}_i \boldsymbol{P}_i \boldsymbol{C}_i)_{t_2,t_2}}} \tag{54}$$

Time-varying parameters

As for the benchmark models, we consider a sequence of nested models. For all of the models the observation variance (s) is a constant (to be estimated). The parameters α, θ and c are allowed to depend on horizon (t) and forecasted power ($\tilde{p}_{i,t}$), and the functional relationship is modelled by

$$\alpha(\tilde{p}_{i,t}, t) = \frac{1}{1 + \exp(-\alpha_0 - f_\alpha(\tilde{p}_{i,t}) - g_\alpha(t))} \tag{55}$$

$$\theta(\tilde{p}_{i,t}, t) = \frac{K_\theta}{1 + \exp(-\theta_0 - f_\theta(\tilde{p}_{i,t}) - g_\theta(t))} \tag{56}$$

$$c(\tilde{p}_{i,t}, t) = \frac{K_c}{1 + \exp(-c_0 - f_c(\tilde{p}_{i,t}) - g_c(t))} \tag{57}$$

where the functions in f and g are modelled by natural cubic splines. We use the same knot placement as in the benchmark models. The range of the spline functions (f_c, f_θ, f_c and g_c) is difficult to control inside the optimization algorithm, and K_θ and K_c are introduced to control the range of c and θ (we use $K_\theta = K_c = 20$, which is far from the maximum estimates of θ and c (see Figures 7 and 8 below).

Estimation

Parameter estimation is carried out considering the joint log-likelihood of the observations in the training set, assuming the distribution (53) and using the general-purpose optimiser `optim` in R.

Table III. Parameter estimates for each of the eight SDE models

	SDE 0	SDE 1	SDE 2	SDE 3	SDE 4	SDE 5	SDE 6	SDE 7
s	0.058	0.048	0.045	0.057	0.048	0.041	0.037	0.037
	(0.054, 0.061)	(0.044, 0.052)	(0.041, 0.049)	(0.05, 0.064)	(0.042, 0.055)	(0.036, 0.046)	(0.032, 0.042)	(0.033, 0.042)
α	0.307	0.472	0.568	0.433				
	(0.286, 0.327)	(0.425, 0.52)	(0.499, 0.633)	(0.374, 0.494)				
θ	0.206	0.129	0.109					
	(0.194, 0.219)	(0.116, 0.144)	(0.097, 0.123)					
c	0	2.879						
		(2.503, 3.312)						
f_c		0	0	0	0	0	0	Spline
g_c		0	Spline	Spline	Spline	Spline	Spline	Spline
f_θ	0	0	0	Spline	Spline	Spline	Spline	Spline
g_θ	0	0	0	0	0	Spline	Spline	Spline
f_α	0	0	0	0	Spline	Spline	Spline	Spline
g_α	0	0	0	0	0	0	Spline	Spline

Note: Appropriate confidence intervals are included in parentheses.

RESULTS

The models are developed by including more and more structure in the parameters, starting from the simple model (SDE 0) as indicated in Table III. We use the same test and training sets as in the 'Data' section above. In each step all possible inclusions are tested and the best model in terms of likelihood is chosen, i.e. inclusions are chosen by forward selection. The simplest model (SDE 0) contains only a constant observation variance (s), a time constant (θ) and a constant diffusion parameter (α). The log-likelihood value for SDE 0 is seen to be between BM 2 and BM 3 (compare Tables I and IV). Including the compensator (SDE 1), c, gives a log-likelihood which is far better than BM 3, with only 4 degrees of freedom (considering cross-validation SDE 0 actually perform better than BM 3).

Further inclusion of time-varying parameters (either directly as horizon or by including a function of $\tilde{p}_{i,t}$) gives significant improvements in all steps, and the best-performing model on the test sets is SDE 5 or 6 (depending on the test set). Overall, the best model on the test set is SDE 5, indicating some over-parametrization for the last models (SDE 6 and SDE 7). Table III summarizes the eight different models presented in the following. In SDE 2–7 all spline functions are included one by one and the best model is chosen in each step.

In SDE models 0 and 1 the parameters are time-homogeneous; the standard deviation increases very fast for the first horizons and finds a stationary level after a few hours (Figure 6); the effect of the compensator (c) is clear from Figure 6, where the bias is illustrated. Including time-dependent compensator (SDE 2–7) results in a slower increase in the standard deviation (Figure 6, SDE 5), high bias for short look-ahead time (Figure 6, SDE 5) and low bias for long look-ahead time (opposite of SDE 0). Including a WPPT forecast -dependent ($\tilde{p}$) time constant (SDE 3, not shown) maintains a symmetric standard deviation (around $\tilde{p}_{i,t} = 0.5$), but the increase is slower and the standard deviation is smaller for large look-ahead times, and the overall picture for the bias is similar (even though not symmetric).

Introducing a WPPT forecast-dependent diffusion parameter (SDE 4, not shown) gives an asymmetric standard deviation and asymmetric bias. Generally, including more structure gives more complex dependence of standard deviation and bias but, as indicated in Table IV, introducing too much structure results in over-parametrization (SDE 6 and 7).

Table IV. Value of the log-likelihood function for the training set and test sets

	d.f.	l(train)	p-value	l(test1)	l(test2)	l(test)
SDE 0	3	8408.4		92,376	69,222	161,598
SDE 1	4	8532.4	< 0.0001	94,292	70,719	165,011
SDE 2	9	8645.3	< 0.0001	94,396	70,930	165,326
SDE 3	14	8673.9	< 0.0001	94,391	70,982	165,373
SDE 4	19	8703.3	< 0.0001	94,771	71,227	165,998
SDE 5	24	8728.9	< 0.0001	**94,857**	71,439	**166,296**
SDE 6	29	8760.5	< 0.0001	94,781	**71,480**	166,262
SDE 7	34	**8770.6**	0.0012	94,715	71,409	166,124

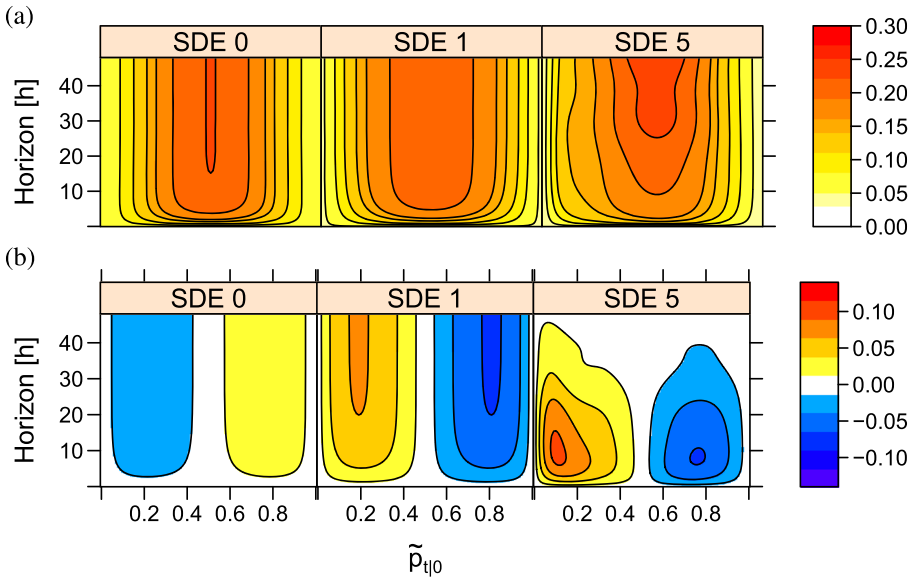

Figure 6. Standard deviation (upper panel) and bias (lower panel) as a function of $\tilde{p}_{i,t}$ and horizon (for $\tilde{p}_{i,t}$ in time) for models SDE 0, 1 and 5

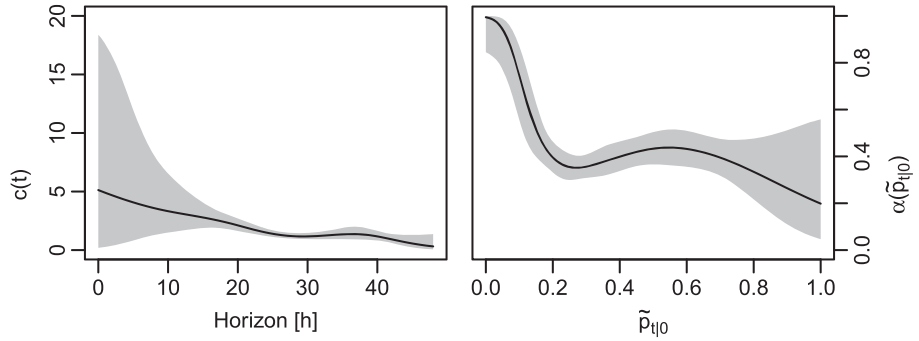

Figure 7. The compensator function (c) as a function of horizon (left panel) and α as a function of $\tilde{p}_{i,t}$ (right panel), for model SDE 5

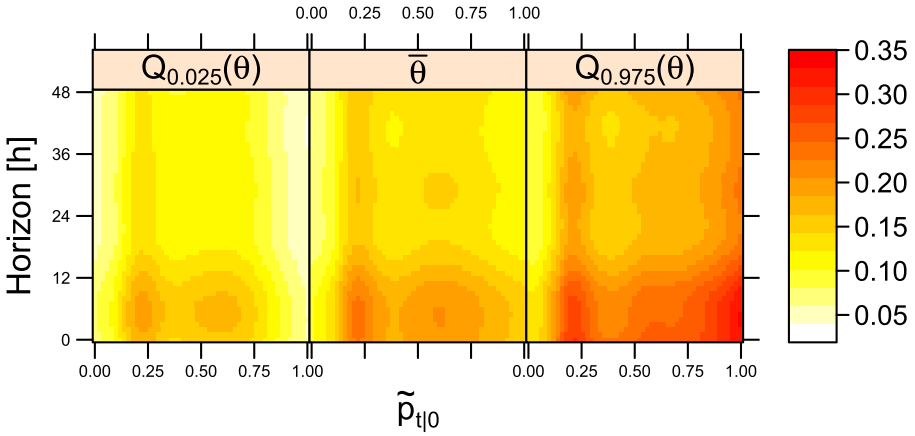

Figure 8. θ as a function of horizon and $\tilde{p}_{i,t}$ for model SDE 5; centre panel is $\hat{\theta}$, and left and right panels indicate 95% confidence limits for the estimate (based on boot strapping from $N(\hat{\theta}, \Sigma_\theta)$, where Σ_θ is obtained from the Hessian of the negative log-likelihood)

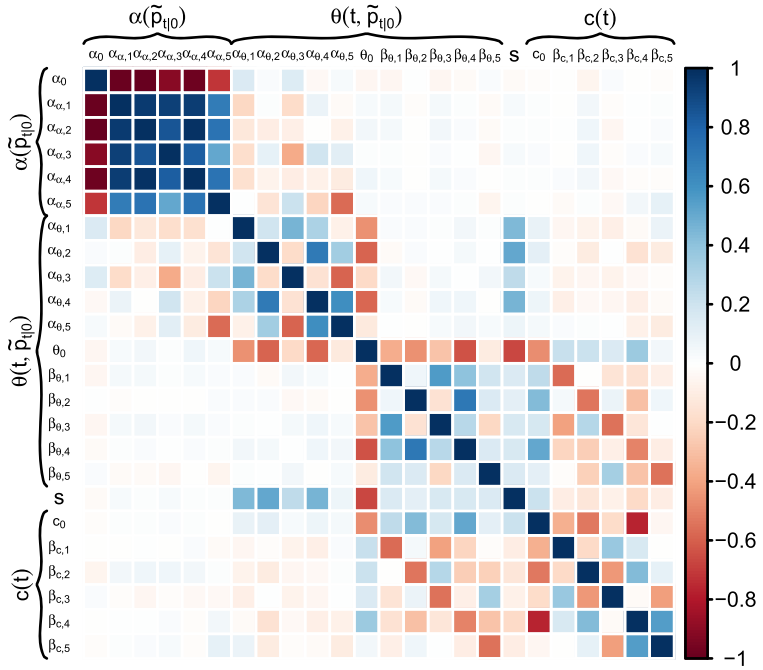

Figure 9. Parameter correlation matrix for the parameters of model SDE 5

The time-dependent parameters are shown for model SDE 5 in Figure 7. The compensator (Figure 7, left panel) is large for short look-ahead times and small for long look ahead times. This might be an effect of the WPPT forecast itself ($\tilde{p}_{i,t}$) being closer to 0.5 (away from the boundary) for large look-ahead times. The diffusion parameter (α) (Figure 7, right panel) is large for $\tilde{p}_{i,t}$ close to zero and small for large values of $\tilde{p}_{i,t}$, but to get the

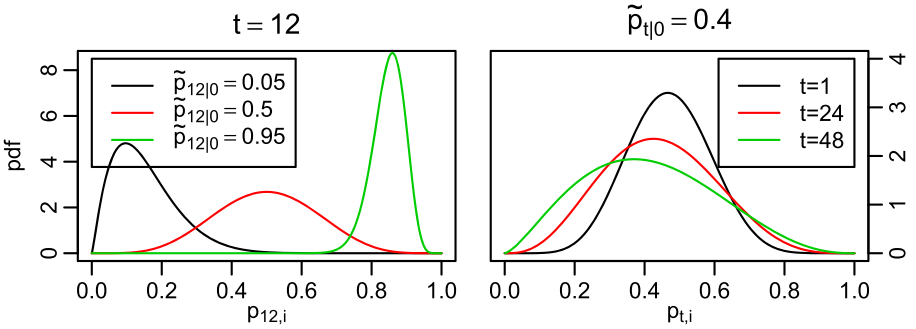

Figure 10. Stationary probability density function for look-ahead time (t) equal to 12 hours for different values of WPPT forecasts (left panel), and for WPPT forecast ($\tilde{p}_{i,t}$) equal to 0.4 and different horizons (right panel)

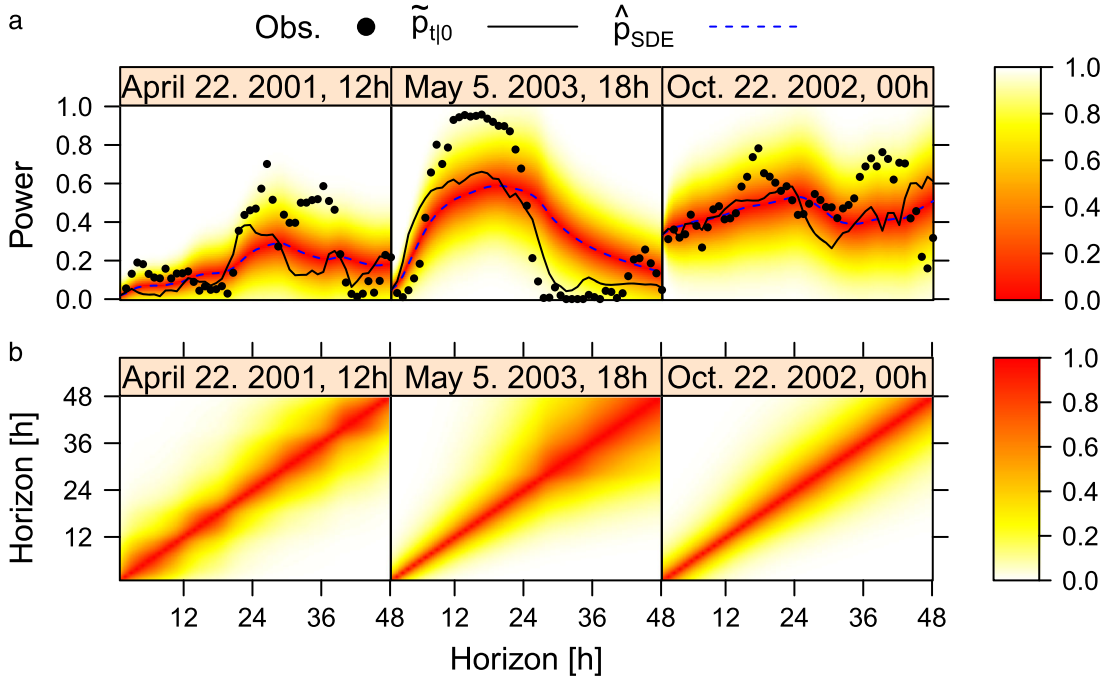

Figure 11. Predictions and prediction intervals (upper panels), and correlation structure (lower panels) for three different time points

full effect on the variance it should be multiplied by $\tilde{p}_{i,t}(1 - \tilde{p}_{i,t})$, and integrated w.r.t. time (as in Figure 6). θ (Figure 8) is large (implying a small time constant) for short look-ahead time, implying that the process will converge to $\tilde{p}_{i,t}$ fast in the beginning of each time series, indicating that the quality of the forecast is better for short look-ahead times.

In general, there are large uncertainties for some regions of the parameters; small look-ahead time for $c(t)$ (Figure 7, left panel), WPPT forecast close to 1 for α (Figure 7, right panel), and small look-ahead times and large WPPT forecast for θ (Figure 8). The correlation between some parameters (Figure 9) is also large. The very strong correlation is, however, within the coefficients for the same spline functions, e.g. the correlations between the parameters for α are very large, but the uncertainties of the resulting curve are not very large (Figure 7, right panel). These parameter uncertainties and correlations, of course, suggest that the model could be reduced. On the other hand, the likelihood ratio test and cross-validation indicate that all effects (c, α, and θ), should be present in the model. We will not pursue the issue of reducing the number of parameters further here, but rather keep the flexible setup of the spline basis function.

Stationary distribution

Since the parameters are all functions of time, the presented SDEs are always in the transient phase. It is, however, illustrative to look at the stationary distribution. If the (time-varying) parameters are fixed, the stationary distribution of the underlying SDE is a Beta distribution with parameters $\frac{1+c_t\cdot(1-\tilde{p}_{i,t})}{2\alpha_t\cdot(1-\tilde{p}_{i,t})}$ and $\frac{1+c_t\cdot\tilde{p}_{i,t}}{2\alpha_t\cdot\tilde{p}_{i,t}}$. This is visualized in Figure 10 (for model SDE 5), for different WPPT point forecasts (Figure 10, left panel) and for different horizons (Figure 10, right panel). The figure clearly illustrates the skewness obtained by the state-dependent formulation of the diffusion, but also the drift in location is observed.

Scenarios

Figure 11 presents three different forecasts and for each period WPPT forecast ($\tilde{p}_{i,t}$), observations and SDE predictions (Model 5) are compared. Notably, SDE models are able to capture the dynamic properties of wind power production in different situations. The correlation structure is closely linked to the time constant, and the introduction of correlation implies that the mode of the SDE solution lags behind the WPPT point forecast. The ability to model very complex correlation structures is seen in Figure 11. The correlation structure depends on the history of point forecast and not just the horizon as for the benchmark models.

DISCUSSION AND CONCLUSION

The starting point of the modelling is a logistic-type SDE driven by a state-of-the-art forecasting algorithm providing point forecasts of wind power production (which itself is driven by a MET forecast). By considering natural restrictions on the diffusion process, we are able to formulate a set of ODEs that govern the second-order moment representation of wind power production. The considered SDE models perform far better in terms of likelihood than a variety of benchmarks with a larger number of parameters.

The SDE approach offers an intuitive split between parameters that govern bias (c), variance (α) and correlation (θ). In the present work we include the parameters as non-parametric functions (splines), while the investigation of parametric formulations (or indeed simplification), possibly considering data from multiple locations, is left for future work. The focus of this paper is to describe how SDEs can be used for scenario forecasting when correlation is taken into account. In general, it might be advisable to use some kind of transformation, e.g. using quantiles, before the SDE formulation is applied, but in order to keep the formulation and reasoning simple we have chosen not to consider such transformations. Instead we concentrate on simple SDE formulations and development of the covariance structure. In principle, it would be possible to apply some transformation to both the benchmark models and the SDE models.

Furthermore, the SDE framework is able to provide a simultaneous description of the interdependence structure as well as the sequence of bounded and asymmetrical state-dependent predictive densities for future wind power production.

The SDE-based state-space formulation offers a simple and flexible framework for constructing very complex correlation or interdependence structures that are difficult to formulate explicitly in a direct parametric setup. The introduced framework for forecasting is able to provide an integrated description of the interdependence structure as well as the sequence of bounded and asymmetrical state-dependent predictive densities for future wind power production.

The mapping from wind speed to wind power is very complex, and depends on factors not taken into account in the presented formulation, such as how dirty the blades of the turbines are. Hence state-of-the-art methods for wind power forecasting of today should include some methods for adaptive estimation of the parameters, like those described in Madsen (2008). It is expected that such techniques will improve the performance (as in Møller *et al.*, 2008, for example) if integrated with the SDE approach.

We believe that the framework presented here has great potential. However, as indicated above, more work related to model development and optimization of algorithms is needed before the method is ready for online settings.

REFERENCES

Baadsgaard M, Nielsen JN, Spliid H, Madsen H, Preisel M. 1997. Estimation in stochastic differential equations with state dependent diffusion term. In *11th IFAC Symposium on System Identification, SYSID '97*, IFAC, Kitakyushu, Fukuoka Japan; 1425–1430.

Costa A, Crespo A, Navarro J, Lizcano G, Madsen H, Feitosa E. 2008. A review of the young history of the wind power short-term prediction. *Renewable and Sustainable Energy Reviews* **12**(6): 1725–1744.

Forman J, Sørensen M. 2008. The Pearson diffusion: a class of statistically tractable diffusion processes. *Scandinavian Journal of Statistics* **35**: 438–465.

Giebel G, Brownsword R, Kariniotakis G, Denhard M, Draxl C. 2011. *The State-of-the-Art in Short-Term Prediction of Wind Power: A Literature Overview*. ANEMOS.plus Project: Paris.

Iacus SM. 2008. *Simulation and Inference for Stochastic Differential Equations: With R Examples*. Springer Series in Statistics. Springer: New York.

Jazwinski AH. 1970. *Stochastic Processes and Filtering Theory*. Academic Press: New York.

Lange M. 2005. On the uncertainty of wind power predictions: analysis of the forecast accuracy and statistical distribution of errors. *Journal of Solar Energy Engineering* **127**: 177–184.

Leutbecher M, Palmer TN. 2008. Ensemble forecasting. *Journal of Computational Physics* **227**: 3515–3539.

Madsen H. 2008. *Time Series Analysis*, Texts in Statistical Science. Chapman and Hall: Boca Raton, FL.

Madsen H, Sejling K, Nielsen TS, Nielsen HAa, Nielsen US. 1995. *Wind Power Prediction Tool in Control Dispatch Centres*. ELSAM: Skaerbaek, Denmark.

Madsen H, Nielsen TS, Tøfting J. 1998. WPPT: a tool for online wind power prediction. In *EPRI-DOE-NREL Wind Power Forecasting Workshop*, San Francisco, CA. Available: http://henrikmadsen.org/publications/publications-1990-1999/.

Møller JK, Nielsen HAa, Madsen H. 2008. Time-adaptive quantile regression. *Computational Statistics and Data Analysis* **52**(3): 1292–1449.

Møller JK, Madsen H, Carstensen J. 2011. Parameter estimation in a simple stochastic differential equation for phytoplankton modelling. *Ecological Modelling* **222**: 1793–1799.
Møller JK, Pinson P, Madsen H. 2013. Probabilistic forecasts of wind power generation by stochastic differential equation models. In *59th ISI World Statistical Congress,* Hong Kong; 1471–1476.
Morales JM, Conejo AJ, Madsen H, Pinson P, Zugno M. 2014. *Integrating Renewables in Electricity Markets*. Springer: New York.
Nielsen HAa, Nielsen TS, Madsen H, Giebel G, Badger J, Landberg L, Sattler K, Voulund L, Tøfting J. 2006. From wind ensembles to probabilistic information about future wind power production - results from an actual application. In *Proceedings of the 9th International Conference on Probabilistic Methods Applied to Power Systems,* Stockholm; 1–8.
Nielsen HAa, Madsen H, Nielsen TS. 2006. Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. *Wind Energy* **9**(2): 95–108.
Øksendal B. 2003. *Stochastic Differential Equations: An Introduction with Applications* (6th edn). Springer: Berlin.
Pinson P. 2012. Very short-term probabilistic forecasting of wind power with generalized logit-Normal distributions. *Journal of the Royal Statistical Society, Series C* **61**(4): 555–576.
Pinson P, Madsen H, Nielsen HAa, Papaefthymiou G, Klöckl B. 2009. From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy* **12**(1): 51–62.
Schurz H. 2007. Modelling, analysis and discretization of stochastic logistic equations. *International Journal of Numerical Analysis and Modelling* **4**(2): 178–197.
Trombe P-J, Pinson P, Madsen H. 2012. A general probabilistic forecasting framework for offshore wind power fluctuations. *Energies* **5**: 621–657.
WPPT. 2012. Wind power prediction tool. Available: (http://www.enfor.dk/products/wppt.aspx) [accessed on 26 August 2015].

Authors' biographies:

Jan Kloppenborg Møller recieved his PhD degree in 2011 and is no Associate Professor in Stochastic Dynamical Systems Department of Applied Mathematics and Computer Science at the Technical University of Denmark. The methodological focus of his research is stochastic dynamical systems in continuous time (Stochastic Differential Equations). He has been involved in applications of SDE's in diverse areas of applications like; ecosystems, urban draniage, water treatment, and wind and solar power forecast.

Marco Zugno received the M.Sc. degrees in Electrical Engineering from the Technical University of Denmark in 2008 and in Automation Engineering from the University of Padova, Italy, in 2009. He holds a Ph.D. degree awarded in 2013 by the Department of Applied Mathematics and Computer Science at the Technical University of Denmark. He is currently a postdoctoral fellow with the same department. His research interests include modeling and analytics for energy systems, stochastic programming, robust optimization and hierarchical optimization.

Henrik Madsen got his PhD in Statistics at the Technical University of Denmark in 1986. He was appointed Ass. Prof. in Statistics in 1986, Assoc. Prof. in 1989, and Professor in Mathematical Statistics with a special focus on Stochastic Dynamical Systems in 1999. His main research interest is related to analysis and modelling of stochastic dynamics systems. This includes signal processing, time series analysis, identification, estimation, grey-box modelling, prediction, optimization and control. The applications are mostly related to Energy Systems, Informatics, Environmental Systems, Bioinformatics, Biostatistics, Process Modelling and Finance. He has authored or co-authored approximately 500 papers and 12 books. The most recent books are Time Series Analysis (2008); General and Generalized Linear Models (2011); Integrating Renewables in Electricity Markets (2013), and Statistics for Finance (2015).

Authors' addresses:

Jan Kloppenborg Møller, **Marco Zugno** and **Henrik Madsen**, DTU Compute, Matematiktorvet, Technical University of Denmark Building 303B, DK-2800 Lyngby, Denmark.

