Stochastic Wind Power Forecasting

Waled Alhaddad[1], Ahmed Kebaier[2], and Raúl Tempone[3]

[1,3]CEMSE Division, King Abdullah University of Science and Technology (KAUST), Saudi Arabia
[3]Alexander von Humboldt Professor, RWTH Aachen University, Germany
[2]Université Paris 13, Sorbonne Paris Cité, LAGA, CNRS (UMR 7539), Villetaneuse, France

February 1, 2020

Abstract

Reliable wind power generation forecasting is crucial for applications such as the allocation of energy reserves, optimization of electricity price and operation scheduling of conventional power plant. We propose a data driven model based on parametric Stochastic Differential Equations (SDEs) to captures real-world asymmetric dynamics of wind power forecast errors. Our SDE framework incorporates time derivative tracking of the forecast, time-dependent mean reversion parameter and an improved diffusion term. We are able to simulate future wind power production paths and to get sharp confidence bands. The method is forecast technology agnostic and enables the comparison between different forecasting technologies on the basis of information criteria. We apply the model to historical French and Uruguayan wind power production data and forecasts on the period (2017-2018).

Keywords: Indirect inference, wind power, probabilistic forecasting, stochastic differential equations, Lamperti transform, model selection, sensitivity.

1 Introduction

Reliable wind power generation forecasting is crucial for the following applications (see, for example, Giebel et al., 2011, p. 5, Chang, 2014, p. 162, Zhou et al., 2013):

- Allocation of energy reserves such as water levels in dams or oil, and gas reserves.
- Operation scheduling of non-controlable power plants.
- Optimization of the price of electricity for different parties such as electric utilities, Transmission system operator (TSOs), Electricity service providers (ESPs), Independent power producers (IPPs), and energy traders.
- Maintenance planning such as that of power plants components and transmission lines.

Different methods have been applied to wind power forecasting. They can be generally categorized as follows: physical models, statistical methods, artificial intelligence methods and hybrid approaches. The output of such methods is usually a deterministic forecast. Occasionally probabilistic forecasts are produced through uncertainty propagation in the data, parameters or through forecast ensembles. However, there is a lacking in simulating and producing data driven stochastic forecasts based on real-world performance of forecasting models. It is crucial to capture actual performance of a forecast as it has been known that different forecasting technologies exhibits different behavior for different wind farms and seasons [ref]. This is due to many factors which forecast are challenged to capture such as the surrounding terrains of the wind farm and the condition of the blades such as icing, wear and tear or dirt. It is known that complex terrains in both off shore and on shore locations decrease the accuracy of wind power forecasts significantly [ref]. It also has been shown that the performance of forecasts varies from month to month. Thus the performance of wind power forecasts is location and time dependent.

Many approaches have been taken to evaluate the uncertainty of a given forecast. There are two types of errors: level errors and phase errors. The use of mean or median errors in this context may be misleading as wind power forecast errors are asymmetric. This is a natural consequence of wind power being non-negative and bounded by the maximuim capacity of production. This is important as the associated cost to power forecast errors are also asymmetric due to different costs for up and down power regulations which are determined by the electricity market [ref].

We propose to model wind power forecasts errors using parametric stochastic differential equations (SDEs) whose solution defines a stochastic process. This resultant stochastic process describes the time evolution dynamics of wind power forecast errors while capturing properties such as a correlation structure and the inherent asymmetry. Additionally, the model we propose is agnostic of the forecasting technology and serves to complement forecasting procedures by providing a data driven stochastic forecast. Hence, we are able to evaluate wind power forecasts according to their real-world performance and we are able to compare different forecasting technologies. Most notably, we are able to simulate future wind power production given a deterministic wind power forecast. Future wind power production using Monte Carlo methods, as well as the analytic form of the proposed SDE, can be used in optimal control problems involving wind power production.

Previous attempt by (Møller et al., 2016) considered stochastic wind power forecast models based on stochastic differential equations (SDEs). Here, we propose an improved model featuring time derivative tracking of the forecast, time-dependent mean reversion, modified diffusion and non-Gaussian approximations. We apply the model to French and Uruguayan wind power forecasts together with historical wind power production data pertaining to the year 2017-2018.

We use a year long data set from Uruguay based on 1000 observation paths, each of which is 72-hours long with observations recorded every 10 min. In total, it is a data set of approximatly half a million data points recorded in 2018. See Figure (1). The data is normalized with respect to the maximuim power capacity of wind power production in Uruguay. The asymmetry or skew nature of the data is clear when inspected for different power production levels. We split the aggregated data into a low, meduim and high power range. See Figure (2).

In this paper we present the phenomenological underlying model in Section 2 and describe the physical constraints in Section 3 and how these constrains can be met. Then, in Section 4, we will introduce an alternative formulation of the model in Lamperti space. In Section 5, we show our parameter estimation procedure and its results in Section 6. We compare alternative models in Section 7 and different forecast providers in Section 8.

2 Phenomenological Model

We introduce the following phenomenological model. Let $X = \{X_t, t > 0\}$ be the wind power generation forecasts stochastic process defined by the following parameterized stochastic differential equation (SDE)

$$\begin{cases} dX_t &= a(X_t; p_t, \dot{p}_t, \boldsymbol{\theta})\, dt + b(X_t; p_t, \boldsymbol{\theta})\, dW_t \\ X_0 &= x_0, \end{cases} \tag{1}$$

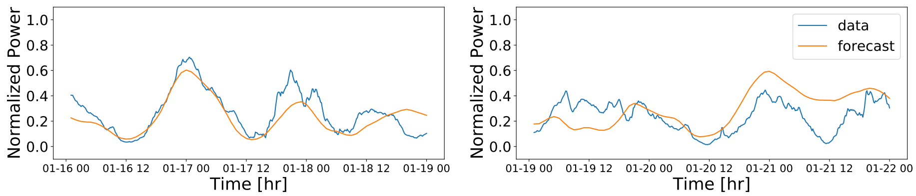

Figure 1: Two samples from the Uruguayan of 2018. Each sample compromises of two 72-hour paths. In yellow is an hourly wind power production forecast. In blue is the actual wind power production recorded in 10 minute intervals.

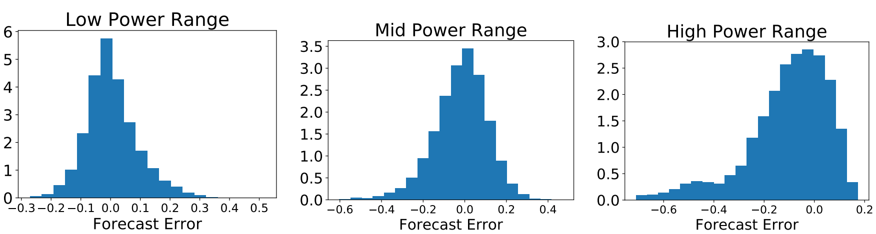

Figure 2: We see that forecast errors exhibit skewness near the boundries (i.e. low and high power production regimes.). Low power is when produced power is less than 0.3, mid power is when it is between 0.3 and 0.6 and high power when is larger than 0.6.

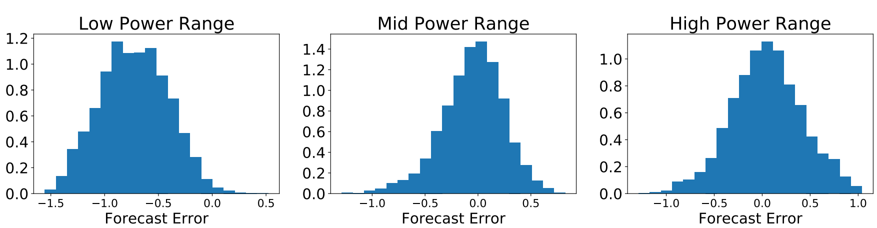

Figure 3: We observe that skewness has been greatly reduced after the Lamperti transformation. This motivates us to use a Gaussian transition density as a proxy density.

Figure 4: insert figure showing the difference with and without derivative tracking

where

- $a(\cdot; p_t, \dot{p}_t, \boldsymbol{\theta}) : [0,1] \to \mathbb{R}$ denotes a drift function,
- $b(\cdot; p_t, \boldsymbol{\theta}) : [0,1] \to \mathbb{R}$ a diffusion function,
- $\boldsymbol{\theta}$ is a vector of parameters,
- p_t is a time-dependent scalar value and $\dot{p}_t$ is its time derivative at time t,
- $\{W_t, t > 0\}$ is a standard Wiener process in $\mathbb{R}$.

In this work, p_t is to be considered as a deterministic forecast of the wind power generation at time t which is available from an official source.

Our goal is to provide a specification of the model (1) to follow closely the available wind power forecasts while ensuring its unbiasedness with respect to the forecast. It is straightforward to show that $\mathbb{E}[X_t] = p_t$, where p_t is the available deterministic wind power forecast for the time t.

2.1 Physical Constrains

Let p_t be a numerical wind power forecast, which is an input to this approach. Assume that the wind power generation forecasts are modeled as solutions to the following Itô stochastic differential equation

$$\begin{aligned} dX_t &= \dot{p}_t \, dt - \theta_t (X_t - p_t) \, dt + b(X_t; \boldsymbol{\theta}) \, dW_t \,, \quad t > 0 \\ X_0 &= x_0 \,. \end{aligned} \tag{2}$$

According to the SDE specification in (2), the process X_t satisfies the two following properties:

- it is mean reverting to the wind power forecast p_t, and
- it tracks the time derivative wind power forecast $\dot{p}_t$.

Observe that a mean reverting model without derivative tracking exhibits consistent lags, as it is shown in Figure (4). See Section 7 for comparisons.

We normalize the forecast and production data to Uruguay installed power capacity at the time of observation. Thus our process must be limited to the range $[0,1]$. To enforce this constraint, our drift and diffusion terms must satisfy certain rules.

Let $\boldsymbol{\theta} = (\theta_0, \alpha)$. We want a process that follow the wind forecast, thus we choose mean reverting drift term which also tracks the derivative of p_t, which is an input to our model.

$$a(x; p_t, \boldsymbol{\theta}) = \dot{p}_t - \theta_t (x - p_t) \tag{3}$$

where $\theta_t > 0$ is a time-dependent parameter that controls the speed of reversion.

We would like a diffusion term that vanishes at the boundaries to prevent the process from escaping the region $[0,1]$.

$$b(x; p_t, \boldsymbol{\theta}) = \sqrt{2\theta_t \alpha x (1-x)} \tag{4}$$

where $\alpha > 0$ is a constant parameter that controls the path variability. This diffusion term belongs to the Pearson diffusion family and , in particular, it is a Jacobi type diffusion.

To further ensure that the process does not escape the region $[0,1]$, the mean reversion parameter has to be selected according to the following rule. Observe that the time derivative term $\dot{p}_t$ is not controlled to maintain that X_t stays a.s. inside the range $[0,1]$. In other words, the zero drift line defined by $a(x; p_t, \boldsymbol{\theta}) = 0$, which an attractor, must be contained inside the range $[0,1]$. Thus, we must have that

$$\frac{-|\dot{p}_t|}{p_t} \leq \theta \leq \frac{|\dot{p}_t|}{1-p_t} \tag{5}$$

which is satisfied by choosing a time-dependent θ_t as follows,

$$\theta_t = \max\left(\theta_0 \ , \ \frac{|\dot{p}_t|}{\min(p_t, 1-p_t)}\right), \quad \theta_0 > 0 \tag{6}$$

Change of Variables:

To avoid differentiation of the forecast p_t and simplify, we apply the change of variables

$$V_t = X_t - p_t \, .$$

The model becomes,

$$\begin{aligned} dV_t &= -\theta_t V_t \, dt + \sqrt{2\theta_t \alpha (V_t + p_t)(1 - V_t - p_t)} \, dW_t \\ V_0 &= v_0 \, . \end{aligned} \tag{7}$$

3 State independent diffusion: Lamperti transform

Our model (7) for the forecast error has a diffusion term that depends on the state variable V_t. To estimate the unknown model parameters, a recommended technique is to modify the SDE (7) by applying the so-called Lamperti transform (see Iacus, 2008, pp. 40–41, Møller and Madsen, 2010, Särkkä and Solin, 2019, pp. 98–100) to the process V to obtain a SDE for the transformed process whose diffusion term does not depend anymore on the state variable.

We consider the following Lamperti transformation

$$Z_t = h(V_t) = \int_{\xi}^{x} \frac{1}{\sqrt{(u+p_t)(1-u-p_t)}}\, du \Bigg|_{x=V_t}, \tag{8}$$

where ξ is an arbitrary point of the state space of the process V. The choice of $\xi = \frac{1}{2} - p_t$, where p_t is a known input, leads to the process

$$Z_t = h(V_t) = \arcsin(2(V_t + p_t) - 1)\,, \tag{9}$$

that, after applying Itô's formula on $h(V_t)$, gives the state-independent diffusion SDE

$$dZ_t = \left[\frac{-\theta_t V_t}{\sqrt{(V_t+p_t)(1-V_t-p_t)}} - \frac{1}{2}\frac{\theta_t \alpha\,(1-2(V_t+p_t))}{\sqrt{(V_t+p_t)(1-V_t-p_t)}} \right] dt + \sqrt{2\theta_t \alpha}\, dW_t\,. \tag{10}$$

After replacing $V_t = \frac{\sin(Z_t)+1}{2} - p_t$ in (10), we obtain that the process Z satisfies the SDE

$$dZ_t = \frac{-\theta_t(\sin(Z_t) + 1 - 2p_t) + \theta_t \alpha \sin(Z_t)}{\cos(Z_t)}\, dt + \sqrt{2\theta_t \alpha}\, dW_t\,. \tag{11}$$

We can see in Figure (3) the effect of the Lamperti transformation upon the forecast error data. The Lamperti transformation has greatly reduced the forecast error skewness, ensuring that the process stays in the range $[0, 1]$. Therefore, in this case, the transition densities of the process Z can be adequately approximated through Gaussian densities.

4 Likelihood in V space

4.1 Likelihood

Suppose that any of M non-overlapping paths of the continuous-time Itô process $V = \{V_t, t > 0\}$ is sampled at N equispaced discrete points with length interval Δ_N, and let $V^{M,N} = \{V_{t_1^{M,N}}, V_{t_2^{M,N}}, \dots, V_{t_N^{M,N}}\}$ denote this random sample.

Let $\rho_i(v|v_{j,t_{i-1}};\boldsymbol{\theta})$ denote the conditional probability density of V_{j,t_i} given $V_{j,t_{i-1}} = v_{j,t_{i-1}}$ evaluated at v, where $\boldsymbol{\theta} = (\theta_0, \alpha)$ are the unknown model parameters.

The Itô process V defined by the SDE (7) is Markovian, then the likelihood function of the data can be written as the following product of transition densities:

$$\mathcal{L}(\boldsymbol{\theta}; V^{M,N}) = \prod_{j=1}^{M}\prod_{i=1}^{N} \rho_i(V_{j,t_i}|V_{j,t_{i-1}};\boldsymbol{\theta})\, \rho_0(V_{j,t_0})\,. \tag{12}$$

The exact computation of the likelihood (12) relies on the availability of a closed-form expression for the transition densities of V that, on the basis of the Markovian property of V, are characterized, for $t_{j,i-1} < t < t_{j,i}$, as solutions of the Fokker-Planck-Kolmogorov equation (Iacus, 2008, p. 36, Särkkä and Solin, 2019, pp. 61–68):

$$\begin{aligned} \frac{\partial f}{\partial t}\rho_i(v_{j,i}, t|v_{j,i-1}, t_{j,i-1};\boldsymbol{\theta}) &= -\frac{\partial}{\partial v}(-\theta_t v\, \rho_i(v, t_{j,i}|v_{j,i-1}, t_{j,i-1};\boldsymbol{\theta})) \\ &+ \frac{1}{2}\frac{\partial^2}{\partial v^2}(2\theta_t\alpha(v+p_t)(1-v-p_t)\, \rho_i(v, t_{j,i}|v_{j,i-1}, t_{j,i-1};\boldsymbol{\theta}))\,, \end{aligned} \tag{13}$$

subject to the initial conditions $\rho_{i-1}(v, t_{j,i-1}, \boldsymbol{\theta}) = \delta(v - V_{j,t_{i-1}})$, where $\delta(v - V_{j,t_{i-1}})$ is the Dirac-delta generalized function centered at $V_{j,t_{i-1}}$, and suitable boundary conditions.

Nevertheless, closed-form solutions to this initial-boundary value problem can be obtained only in a few cases. Besides, solving numerically (13) at every transition step is computationally expensive. Therefore, under the likelihood-based inferential paradigma, many techniques have been devised to obtain approximate maximum likelihood estimates for the unknown parameters of continuous-time SDE models with discrete observations.

4.2 Approximate Likelihood

Solving for transition densities of the process V_t requires solving the Fokker-Planck equation at every step which is computationally prohibitive. A common choice is performing a Gaussian approximation of the transition densities, but this is inappropriate here due to physical constraints which give rise to asymmetric forecasting errors as seen in figure (2).

We propose a proxy transition density. We match the moments of our SDE model with that of the proxy density. Using Itô formula, we arrive at the following iterative ODEs for the state dependent diffusion formulation (7)

$$\frac{d\mathbb{E}[V_t^k]}{dt} = -k\theta_t\mathbb{E}[V_t^k] + \frac{k(k-1)}{2}\mathbb{E}[V_t^{k-2}b(V_t^k;\theta_t,\alpha)] \tag{14}$$

For $t \in [t_{n-1}, t]$, the first two moments are given by

$$\begin{aligned}
\frac{dm_1(t)}{dt} &= -m_1(t)\theta_t \\
\frac{dm_2(t)}{dt} &= -2m_2(t)\theta_t(1+\alpha) + 2\alpha\theta_t m_1(t)(1-2p_t) \\
&\quad + 2\alpha\theta_t p_t(1-p_t)\,,
\end{aligned} \tag{15}$$

with initial conditions $m_1(t_{n-1}) = v_{n-1}$ and $m_2(t_{n-1}) = v_{n-1}^2$.

And for the state independent diffusion formulation, similarly, we obtain a system of ODEs to determine the centered moments of the Lamperti transformed process V_t. Due to the non-linearity in the drift, we can only approximate the centered moments by the following ODEs,

$$\begin{aligned}
\frac{dm_1(t)}{dt} &= -m_1(t)\theta_t(1-\alpha) - \theta(1-2p_t) \\
\frac{dvar(t)}{dt} &= 2var(t)\theta_t(2p_t-1)\tan(m_1(t))\sec(m_1(t)) \\
&\quad + \theta_t(\alpha-1)\sec^2(m_1(t)) + 2\theta_t\alpha\,.
\end{aligned} \tag{16}$$

A suitable candidate for a proxy transition density is a Beta probability distribution as it is compactly supported and can morph into symmetric and asymmetric shapes.

Moment Matching

To approximate the transition densities of the process V_t by a Beta distribution, we match its moments with the shape parameters ξ_1, ξ_2 of a Beta proxy density on $[-1, 1]$. The any moment of the process V_t is given by solving the corresponding ODE system for the k^{th} moment,

The shape parameters are given by

$$\xi_1 = -\frac{(1+\mu_t)(\mu_t^2+\sigma_t^2-1)}{2\sigma_t^2}\,, \quad \xi_2 = \frac{(\mu_t-1)(\mu_t^2+\sigma_t^2-1)}{2\sigma_t^2}\,, \tag{17}$$

where $\mu_t = m_1(t)$ and $\sigma_t^2 = m_2(t) - m_1(t)^2$.

4.3 Optimization

To initialize the optimization process, we solve the following least-squares problem which gives us first estimates of the mean reversion parameter θ_0:

$$\arg\min \sum_{i}^{M} \sum_{j}^{N} \left(v_{i+1,j} - v_{i,j} - \left(-\theta_t v_{i,j}\right)\left(t_{i+1,j} - t_{i,j}\right)\right)^2 , \tag{18}$$

where $v_{i,j} = x_{i,j} - p_{i,j}$, $x_{i,j}$ is historical wind power production and $p_{i,j}$ is the wind power forecast.

By assuming ergodicity, we can obtain a first estimate on the product of the parameters as follows,

$$\theta_0 \alpha = \frac{1}{M} \sum_{i}^{M} \frac{\sum_{j}^{N} (x_{i+1,j} - x_{i,j})^2}{2 \sum_{j}^{N} x_{i,j}(1 - x_{i,j})} . \tag{19}$$

Solving for α, we have both first estimates to quick start the inference processes given as follows:

Step 1. initialize;

Step 2. optimize the log-likelihood function using Nelder-Mead optimization algorithm on a mini-batch sampled with replacement;

Step 3. check if accuracy threshold is reached. Else, re-initialize the optimization with the most recent result on a larger mini-batch.

5 Model Comparison

We compare two candidate models to find the best-fit that maximizes the retained information,

- Model 0: This model does not feature derivative tracking.

$$\begin{aligned} dX_t &= -\theta_0 (p_t - X_t)\ dt + \sqrt{2\theta\alpha X_t(1 - X_t)}\ dW_t \\ X_0 &= x_0 , \end{aligned} \tag{20}$$

  with $\theta_t = \theta_0$.

Model	parameters (θ_0, α)	AIC	BIC
Model 0	$([insert], [insert])$	[insert]	[insert]
Model 2	$([insert], [insert])$	[insert]	[insert]

Table 1: We compare the different models based on information criterion. [insert]

Forecast Provider	parameters (θ_0, α)	AIC	BIC
Provider A	$([insert], [insert])$	[insert]	[insert]
Provider B	$([insert], [insert])$	[insert]	[insert]

Table 2: [insert]

- Model 1: This model features derivative tracking, i.e. it is equivalent to (2).

$$\begin{aligned} dV_t &= -\theta_t V_t \; dt + \sqrt{2\theta_t \alpha (V_t + p_t)(1 - V_t - p_t)} \; dW_t \\ V_0 &= v_0 \,, \end{aligned} \tag{21}$$

with θ_t given by (6).

6 Forecast Provider Comparison

We compare forecasts from two different companies for the same period. In Figure ([insert])-([insert]), we see that forecast provider A is of better quality than provider B. This is confirmed by both the AIC and BIC information criteria.

Model	parameters (θ_0, α)
low frequency data (hourly)	$([insert], [insert]) \pm ([insert], [insert])$
high frequency data (every 10 minutes)	$([insert], [insert]) \pm ([insert], [insert])$

Table 3: [insert]confidence interval obtained using bootstrap

7 Results

We were able to obtain the following for

Formulation	parameters (θ_0, α)
Without Lamperti transform	$(12, 0.3) \pm ([insert], [insert])$
With Lamperti transform	$(12, 0.29) \pm ([insert], [insert])$

Table 4: We compare the parameters obtained in both the original and Lamperti space. Parameters have been obtained based on [insert]data points from the Uruguayan pertaining to the year 2018 [insert]

.

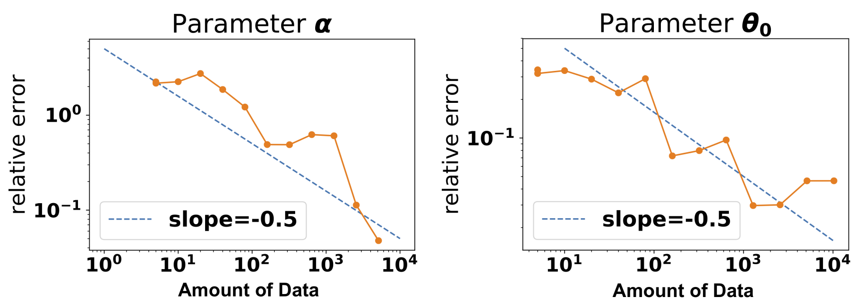

Figure 5: We show self-convergence of our algorithm applied to model 2. We conlcude that the rate matches that of of Monte Carlo. Data is from Uruguay pertaining to year 2018.

We are able to obtain the parameters based on the complete data sets. Using the different models variations, we are able to simulate wind power production given a forecast. We see in figures ()-() five possible wind power production paths for each model.

In Figures ()-(), we show point-wise empirical confidence bands for the different models.

8 Conclusions

Application: Using this tool, it is possible to iterate over hyper parameters for the wind power forecasting, reaching an optimal forecasting.

We have proposed a method to produce stochastic wind power forecasts based on parametric SDEs. This method is agnostic of the wind power forecasting technology. Using this method, we were able to simulate future wind power production paths and obtain confidence bands. We conclude that Model 2 is a best-fit model. It features time-derivative tracking of

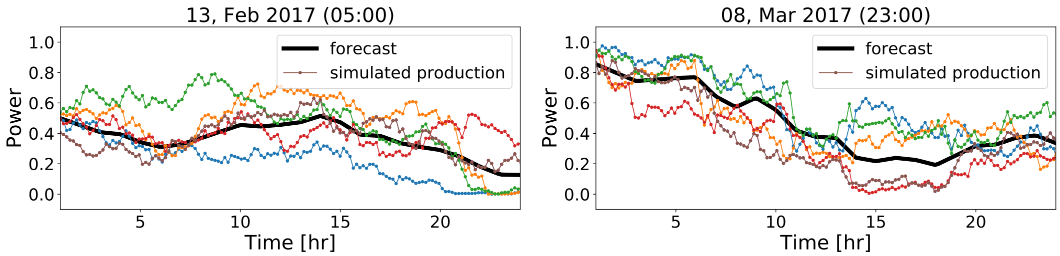

Figure 6: We simulate following model 2 five possible future wind power production paths using the obtained optimal parameters $(\theta_0, \alpha) = (12, 0.3)$. Forecast is from Uruguay pertaining to year 2018.

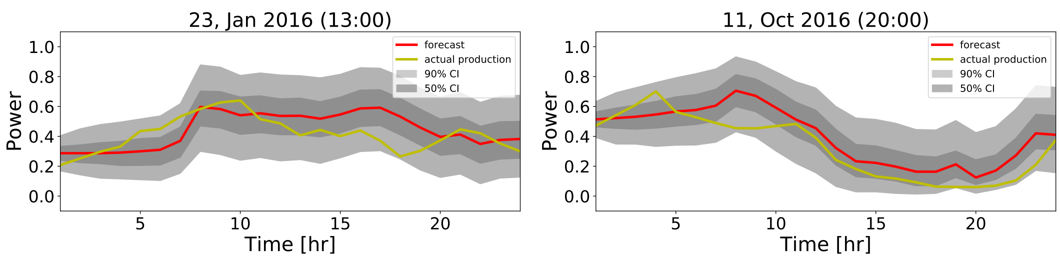

Figure 7: We obtain confidence intervals following model 2 for future wind power production using the obtained optimal parameters $(\theta_0, \alpha) = (12, 0.3)$. Actual production plotted in retrospect. Forecast and data is from Uruguay pertaining to year 2018.

the forecast, time-dependent mean reversion parameter, and a more natural diffusion term. Moreover, the model preserves the asymmetry of wind power forecast errors and their correlation structure.

We were also able to compare two different forecast providers with respect to their real-world performance on the aggregated data set and on specific wind farm sites. Finally, the model paves the way for stochastic optimal control methods enabling optimal decision making under uncertainty.

References

Baadsgaard, M., Nielsen, J. N., Spliid, H., Madsen, H., & Preisel, M. (1997). Estimation in Stochastic Differential Equations with a State Dependent Diffusion Term. *IFAC Proceedings Volumes*, *30*(11), 1369–1374.

Chang, W.-Y. (2014). A Literature Review of Wind Forecasting Methods. *Journal of Power and Energy Engineering*, *2*(4), 161–168. https://doi.org/10.4236/jpee.2014.24023

Elkantassi, S., Kalligiannaki, E., & Tempone, R. (2017). Inference and Sensitivity in Stochastic Wind Power Forecast Models. In M. Papadrakakis, V. Papadopoulos, & G. Stefanou (Eds.), *2nd ECCOMAS Thematic Conference on Uncertainty Quantification in Computational Sciences and Engineering* (pp. 381–393). Rhodes Island, Greece, Eccomas Proceedia UNCECOMP 2017. https://doi.org/10.7712/120217.5377.16899

Giebel, G., Brownsword, R., Kariniotakis, G. N., Denhard, M., & Draxl, C. (2011). *The State of the Art in Short-Term Prediction of Wind Power. a Literature Overview* (2nd edition, tech. rep. ANEMOS.plus and SafeWind projects). https://doi.org/10.13140/RG.2.1.2581.4485

Hurn, A., Jeisman, J., & Lindsay, K. (2007). Seeing the wood for the trees: A critical evaluation of methods to estimate the parameters of stochastic differential equations. *Journal of Financial Econometrics*, *5*(3), 390–455. https://doi.org/10.1093/jjfinec/nbm009

Iacus, S. M. (2008). *Simulation and Inference for Stochastic Differential Equations: With R Examples*. New York, Springer.

Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand J of Stat*, *24*, 211–229.

Møller, J. K., & Madsen, H. (2010). *From State Dependent Diffusion to Constant Diffusion in Stochastic Differential Equations by the Lamperti Transform* (tech. rep. IMM-Technical Report-2010-16). Technical University of Denmark, DTU Informatics, Building 321. Kgs. Lyngby, Denmark.

Møller, J. K., Zugno, M., & Madsen, H. (2016). Probabilistic Forecasts of Wind Power Generation by Stochastic Differential Equation Models. *Journal of Forecasting*, *35*(3), 189–205.

Ravn, H. F. (2006). Short term wind power prognosis with different success criteria, In *9th International Conference on Probabilistic Methods Applied to Power Systems*, IEEE. KTH, Stockholm, Sweden. https://doi.org/10.1109/PMAPS.2006.360291

Särkkä, S., & Solin, A. (2019). *Applied Stochastic Differential Equations*. Cambridge University Press.

Uchida, M., & Yoshida, N. (2014). Adaptive Bayes type estimators of ergodic diffusion processes from discrete observations. *Stat Infer Stoch Process*, *17*, 181–219.

Zhou, Z., Botterud, A., Wang, J., Bessa, R., Keko, H., Sumaili, J., & Miranda, V. (2013). Application of probabilistic wind power forecasting in electricity markets. *Wind Energy*, *16*(3), 321–338. https://doi.org/10.1002/we.1496