Stochastic Wind Power Forecasting

Waleed Alhaddad*   Ahmed Kebaier‡   Raúl Tempone*†

*CEMSE Division, King Abdullah University of Science and Technology (KAUST), Saudi Arabia
†Alexander von Humboldt Professor, RWTH Aachen University, Germany
‡Universit Paris 13, Sorbonne Paris Cit, LAGA, CNRS (UMR 7539) , Villetaneuse , France

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. The SDE proposed 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 confidence bands. The method is forecast technology agnostic and enables the comparison between different forecasting technologies on the basis of an information criteria. We apply the model to historical Uruguayan and French wind power production data and various forecasts on the period (2017-2018).

1. Introduction

Reliable wind power generation forecasting is crucial for the following applications [ref]:

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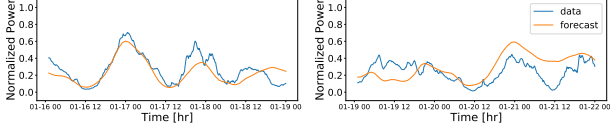
Figure 1. Two samples from the Uruguayan of 2018. Each sample compromises of two 72-hour paths. In yellow is an hourly wind power production forecast. In blue is the actual wind power production recorded in 10 minute intervals.
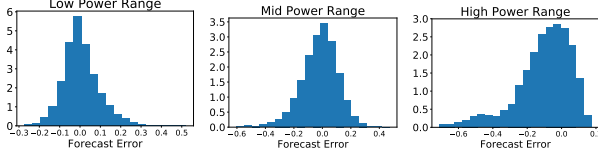

Figure 2. We see that forecast errors exhibit skewness near the boundries (i.e. low and high power production regimes.). Low power is when produced power is less than 0.3, mid power is when it is between 0.3 and 0.6 and high power when is larger than 0.6.

different forecasting technologies. Most notably, we are able to simulate future wind power production given a deterministic wind power forecast. Future wind power production as well as the analytical form of the proposed SDE can be used in optimal control problems involving wind power production.

Previous attempt by Møller et al. (Møller, Zugno, & Madsen, 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 Uruguayan and french 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 approximately 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 underlaying 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_t be the wind power generation forecasts stochastic process defined by the following parameterized stochastic differential equation (SDE),

$$dX_t = a(X_t; p_t, \dot{p}_t, \boldsymbol{\theta})\,dt + b(X_t; p_t, \boldsymbol{\theta})\,dW_t \quad t > 0$$
$$X_0 = x_0 \tag{1}$$

- $a(\cdot; p_t, \dot{p}_t, \boldsymbol{\theta}) : [0,1] \to \mathbb{R}$ a drift function.

- $b(\cdot; p_t, \boldsymbol{\theta}) : [0,1] \to \mathbb{R}$ a diffusion function.

- $\boldsymbol{\theta}$: a vector of parameters.

- p_t time-dependent scalar value and $\dot{p}_t$ is its time derivative at time t. (in our case p_t is a deterministic forecast).

- W_t: Standard Wiener random process in $\mathbb{R}$.

We would like to have this model closely follow wind power forecasts while ensuring that it is unbiased with respect to the forecast. It straightforward to show that $\mathbb{E}X_t = p_t$ where p_t is a deterministic given wind power forecast in the model above.

2.1. Physical Constrains

Let p_t be a numerical wind power forecast, which is an input to this approach. Then the model is given by the following Itô stochastic differential equation,

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

We have that the process X_t mean reverts to the wind power forecast p_t and tracks its the time derivative $\dot{p}_t$. An example such mean reverting model without derivative tracking exhibits consistent lags as shown in Figure (). See section (5) for comparisons.

We normalize the forecast and production data to installed power capacity. Thus our process must be limited to the range $[0,1]$. To enforce this constraint, we must have drift and diffusion control.

Diffusion Control:

The physical constraint to the range $[0,1]$ is respected by choosing diffusion coefficient $b(x; \boldsymbol{\theta}) = \sqrt{2\theta_t \alpha x(1-x)}$ which is zero on the boundaries of $[0,1]$. We also have that α is a path variability constant parameter to be determined.

Drift Control:

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 $a(X_t; \boldsymbol{\theta}) = 0$, which an attractor, must be contained inside the range $[0,1]$. Thus, we must have that

$$\frac{-|\dot{p}_t|}{p_t} \leq \theta \leq \frac{|\dot{p}_t|}{1 - p_t} \tag{3}$$

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

$$\theta_t = \max\left(\theta_0 \, , \, \frac{|\dot{p}_t|}{\min(p_t, 1 - p_t)} \right) \tag{4}$$

Change of Variables:

In order to avoid differentiation of the forecast $\dot{p}_t$ and simplify, we apply a change of variables

$$V_t = X_t - p_t$$

The model becomes,

$$dV_t = -\theta_t V_t \, dt + \sqrt{2\theta_t \alpha (V_t + p_t)(1 - V_t - p_t)} \, dW_t$$
$$V_0 = v_0 \tag{5}$$

with θ_t given by (4).

3. Data

We apply the model to Uruguayan wind power data and their corresponding wind power forecasts. The wind power actual production data set is recorded on intervals of ten minutes for each wind farm in the country. The associated wind power forecasts are provided by two different entities. Each forecast is provided on an hourly basis and is computed for a 72 hour horizon every few hours. Sample data is shown in Figure ().

4. Inference and Optimization

4.1. Likelihood

The SDE above defines the stochastic process V_t. Consider a set of M paths with N observations each, $V^{M,N} = \{V_{t_1}^{M,N}, V_{t_2}^{M,N}, \ldots, V_{t_N}^{M,N}\}$ observed in intervals of Δ_N. Since V_t defined by the SDE in (5) is Markovian, the likelihood function can be written as product of transition densities.

$$\mathscr{L}(\boldsymbol{\theta}; V) = \prod_{j=1}^{M} \prod_{i=1}^{N} \rho(V_{j,i+1} | V_{j,i}, \boldsymbol{\theta}) \rho(V_{j,0}) \tag{6}$$

The transition densities can be exactly obtained by solving the following parametric Fokker-Planck equation,

$$\frac{\partial f}{\partial t}(y, t | x, s, \theta_t, \alpha) = -\frac{\partial}{\partial y}(a(y; \dot{p}_t, p_t, \theta_t) f(y, t | x, s, \theta_t, \alpha))$$
$$+ \frac{1}{2} \frac{\partial^2}{\partial y^2}(b^2(y; \theta_t, \alpha) f(y, t | x, s, \theta_t, \alpha)) \quad t < s \tag{7}$$

The above parametric PDE is computationally prohibitive to be solved for each wind production data point and for all parameter space which is to be optimized. Therefore, we consider approximate likelihood methods.

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 and the asymmetric nature of forecast errors.

Therefore, we propose to approximate likelihood using a Beta distribution. As the process X_t and p_t take on values in $[0,1]$, then V_t takes on values in $[-1,1]$. Therefore, we translate and scale the Beta distribution to have compact support on the interval $[-1,1]$.

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 the Beta distribution on $[-1,1]$. Then, the shape parameters are given by,

$$\xi_1 = -\frac{(1 + \mu_t)(\mu_t^2 + \sigma_t^2 - 1)}{2\sigma_t^2} \quad \xi_2 = \frac{(\mu_t - 1)(\mu_t^2 + \sigma_t^2 - 1)}{2\sigma_t^2} \tag{8}$$

where $\mu_t = \mathbb{E}[V_t]$ and $\sigma_t^2 = \mathbb{E}[V_t^2] - \mathbb{E}[V_t]^2$ and m_i being the i-th moment of the process V_t. Using Itô's formula, the moments of the process V_t are iteratively given by,

$$\frac{d\mathbb{E}[V_t^k]}{dt} = -k\theta_t \mathbb{E}[V_t^k] + \frac{k(k-1)}{2} \mathbb{E}[V_t^{k-2} b(y; \theta_t, \alpha)] \tag{9}$$

with initial conditions, $\mathbb{E}[V_t^k] = v_s^k$ where $t < s$.

4.3. Optimization

Let ℓ be the natural logarithm of the likelihood in (6). Given the data set $V^{M,N}$, we optimize the the log-likelihood ℓ using L-BFGS algorithm for the optimal parameters θ_0^*, α^* as follows,

$$(\theta_0^*, \alpha^*) = \arg \min_{\theta_0, \alpha > 0} -\ell(\theta_0, \alpha | V^{M,N}) \tag{10}$$

Model	parameters (θ_0, α)	AIC	BIC
Model 0	(,)		
Model 1	(,)		
Model 2	(,)		

Table 1. We compare the different models based on information criterion. ¡discuss results¿

The contours of the log-likelihood can be seen in Figure (**??**). To further inspect the log-likelihood at the point of optimality (θ_0^*, α^*), we check the ellipse defined by the Hessian of the log-likelihood at the point of optimality. We observe that the ellipse shrinks at a fast rate as shown in Figure (**??**). We see that the convergence rate of the parameters θ_0 and α is slightly faster than the expected rate of $1/\sqrt{M}$. This is due to the correlation structure of the process V_t, thus a path may act as more than one uncorrelated sample. We note that the uncertainty in determining the parameter α is much higher than in θ_0.

5. Model Comparison

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

- Model 0: This model is the most basic model without derivative tracking.

$$dX_t = -\theta_0(p_t - X_t)\, dt + \sqrt{2\theta\alpha X_t(1-X_t)}\, dW_t$$
$$X_0 = x_0 \tag{11}$$

with $\theta_t = \theta_0$.

- Model 1: This model features derivative tracking, i.e. it's equivalent to 2 with a diffusion term that is forecast dependent by including the term $p_t(1-p_t)$.

$$dV_t = -\theta_t V_t\, dt + \sqrt{2\theta_t \alpha p_t(1-p_t)(V_t+p_t)(1-V_t-p_t)}\, dW_t$$
$$V_0 = v_0 \tag{12}$$

with θ_t given by (4).

- Model 2: This model features derivative tracking, i.e. it's equivalent to 2

$$dV_t = -\theta_t V_t\, dt + \sqrt{2\theta_t \alpha (V_t+p_t)(1-V_t-p_t)}\, dW_t$$
$$V_0 = v_0 \tag{13}$$

with θ_t given by (4).

6. Forecast Provider Comparison

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

Forecast Provider	parameters (θ_0, α)	AIC	BIC
Provider A			
Provider B			

Table 2. ¡discuss results¿

Model	parameters (θ_0, α)	AIC	BIC
low frequency data (hourly)			
high frequency data (every 10 minutes)			

Table 3. ¡discuss results¿

Forecast Provider A	parameters (θ_0, α)	AIC	BIC
Wind Farm 1			
Wind Farm 2			
Wind Farm 3			
Wind Farm 4			
Wind Farm 5			
Wind Farm 6			

Table 4. ¡discuss results¿

Forecast Provider B	parameters (θ_0, α)	AIC	BIC
Wind Farm 1			
Wind Farm 2			
Wind Farm 3			
Wind Farm 4			
Wind Farm 5			
Wind Farm 6			

Table 5. <discuss results>

7. Model on Disaggregated Data

We apply Model 2 to disaggregated wind power production data and forecasts. The data is available in this case on an hourly basis. Results are shown in Table (4).

8. Results

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.

9. Conclusions

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 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

- 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. doi:10.1002/for.2367
- G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl, The State-Of-The-Art in Short-Term Prediction of Wind Power: A Literature Overview, 2nd edition. ANEMOS.plus, 2011, project funded by the European Commission under the 6th Framework Program, Priority 6.1: Sustainable Energy Systems.
- Ravn, H. F. (2006). Short term wind power prognosis with different success criteria. 2006 International Conference on Probabilistic Methods Applied to Power Systems. doi:10.1109/pmaps.2006.360291
- Chang, W. (2014). A Literature Review of Wind Forecasting Methods. Journal of Power and Energy Engineering, 02(04), 161-168. doi:10.4236/jpee.2014.24023
- Elkantassi, S., Kalligiannaki, E., & Tempone, R. (2017). Inference And Sensitivity In Stochastic Wind Power Forecast Models. Proceedings of the 2nd International Conference on Uncertainty Quantification in Computational Sciences and Engineering (UNCECOMP 2017). doi:10.7712/120217.5377.16899