Uncertainty Quantification in Wind Power Forecasting

Waleed Alhaddad* Ahmed Kebaier‡ Raúl Tempone*†

*CEMSE Division, KAUST, Saudi Arabia

‡Université Paris 13, Sorbonne Paris Cité, LAGA, CNRS (UMR 7539) , France

†Alexander von Humboldt Professor, RWTH Aachen University, Germany

October 4th, 2019

Introduction

Integration of renewable resources into the urban power grid is a challenge due to uncertainties in power production. We focus on wind power. Reliable wind power production forecasting is crucial to:

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

Current State of Affairs

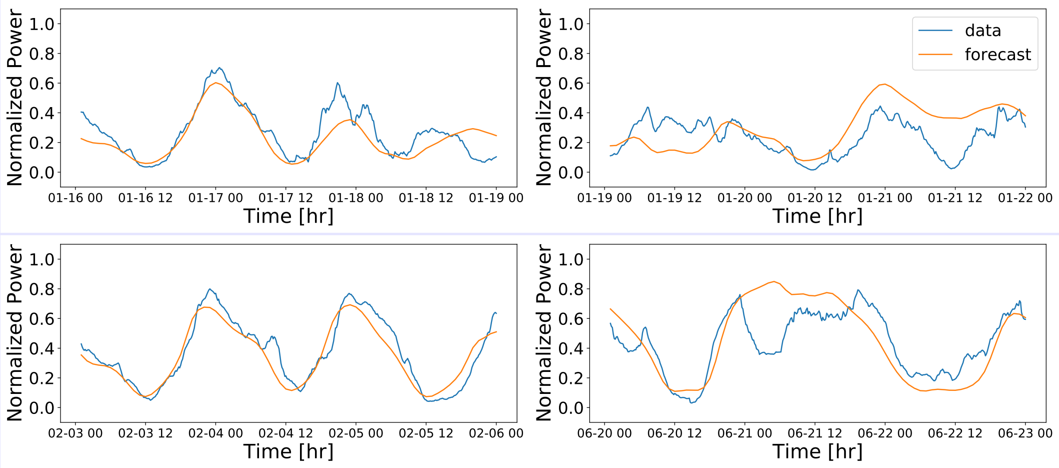
Wind power forecasts can be generally categorized as follows:

- physical models
- statistical methods
- artificial intelligence methods
- other 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 has data driven stochastic forecasts based on the real-world performance of forecasting models.

Data

This is a year long data set from Uruguay based on **1000 72-hour long paths with observations recorded every 10 min** **($\sim$ half a million data points)** recorded in 2018.



Data Skewness

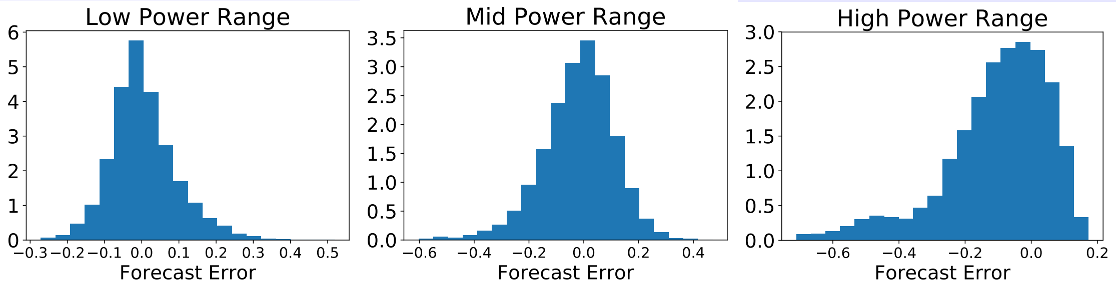


Figure 1: We see that forecast errors exhibit skewness near the boundries (i.e. low and high power production regimes.)

Model

Our goals are to produce a stochastic forecast of wind power production forecasting errors while:

- Capturing the dynamics of the forecast error process.
- Capturing the skew nature of forecast errors.
- Being forecasting-technology agnostic. Thus, compatible with past and future forecasting-technology.
- Learning from historical power production data.

Model

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 forecasting errors.

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

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

Question: How do we choose an appropriate drift and diffusion functions?

How do we choose an appropriate drift and diffusion functions?

Let $\boldsymbol{\theta} = (\theta_0, \alpha)$

1. We want the process to follow the wind forecast, thus we choose a drift term that is mean reverting and tracks the derivative of the deterministic forecast p_t, which is an input to our model.
$$a(x; p_t, \boldsymbol{\theta}) = \dot{p}_t - \theta_t (x - p_t) \tag{2}$$
where θ_t is a time-dependent parameter that controls the speed of reversion.
2. We want a diffusion term that vanishes at the boundaries to prevent the process from escaping the region $[0, 1]$.
$$b(x; p_t, \boldsymbol{\theta}) = \sqrt{2\theta_t \alpha x(1-x)} \tag{3}$$
where α is a constant parameter that controls the path variability.
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,
$$\theta_t = \max\left(\theta_0\ ,\ \frac{|\dot{p}_t|}{\min(p_t, 1-p_t)}\right) \tag{4}$$

Model

Thus, our SDE becomes

$$
\begin{aligned}
dX_t &= \dot{p}_t \ dt - \theta_t(X_t - p_t) \ dt + \sqrt{2\theta_t \alpha X_t(1 - X_t)} \ dW_t \quad t > 0 \\
X_0 &= x_0
\end{aligned}
\tag{5}
$$

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

$$V_t = X_t - p_t$$

The model becomes,

$$
\begin{aligned}
dV_t &= -\theta_t V_t \ dt + \sqrt{2\theta_t \alpha (V_t + p_t)(1 - V_t - p_t)} \ dW_t \\
V_0 &= v_0
\end{aligned}
\tag{6}
$$

Note that this model is Markovian.

Model

Since V_t defined by our SDE is Markovian, the likelihood function can be written as a product of transition densities. 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.

bold theta in red

$$\mathcal{L}(\boldsymbol{\theta}; V) = \prod_{j=1}^{M}\prod_{i=1}^{N} \rho(V_{j,i+1}|V_{j,i}, \boldsymbol{\theta})\rho(V_{j,0}) \tag{7}$$

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

theta_0

$$\begin{aligned} \frac{\partial f}{\partial t}(y, t|x, s, \theta_t, \alpha) = &-\frac{\partial}{\partial y}(a(y; \dot{p}_t, p_t, \theta_t) f(y, t|x, s, \theta_t, \alpha)) \\ &+ \frac{1}{2}\frac{\partial^2}{\partial y^2}(b(y; \theta_t, \alpha) f(y, t|x, s, \theta_t, \alpha)) \quad t < s \end{aligned} \tag{8}$$

2

p_t

This is a parametric PDE which is computationally expensive to solve and optimize for every transition.

Moment Matching

Instead of solving for exact transition densities by the Fokker-Planck, we propose a proxy transition density. We match the moments of our SDE model with that of the proxy density. Using Ito, we arrive at the following iterative ODEs.

double check b^2

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

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

$$\begin{aligned} \frac{dm_1(t)}{dt} &= -m_1(t)\theta_t \\ \frac{dm_2(t)}{dt} &= -2m_2(t)\theta(1+\alpha) + 2\alpha\theta_t m_1(t)(1-2p_t) + 2\alpha\theta_t p_t(1-p_t) \end{aligned} \tag{10}$$

with initial conditions, $m_1(t_{n-1}) = v_{n-1}$ and $m_2(t_{n-1}) = v_{n-1}^2$ where t_{n-1} is the time of the previous obseravtion and v_{n-1} is its value.

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

Algorithm

Execute the following until an accuracy threshold is met:

1. initialize.
2. optimize the log-likelihood function.
   - 2.1 For every evaluation of the log-liklihood function:
     - 2.1.1 Sample a mini-batch of transitions randomly with their associated forecast and parameters.
     - 2.1.2 Solve the ODE system for every transition to obtain the moments.
     - 2.1.3 Match the resulting moments with the parameters of the chosen proxy distribution for every transition.
3. go to step 1 (i.e. re-initialize the optimization with the most recent result).

In the above:

- Choose your favorite deterministic optimization algorithm.
- Choose your favorite integrator to solve the ODE system.

Inference Results

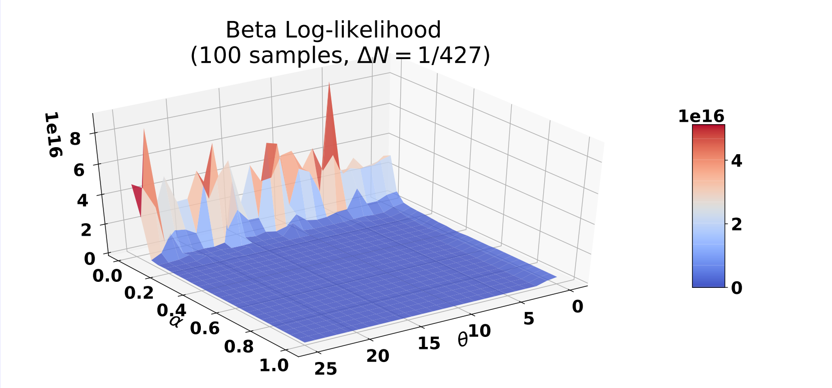


Figure 2: 3-D view of the inverted beta log-likelihood function of 100 sample paths. That is a total of $\sim 42,700$ data points

Inference Results

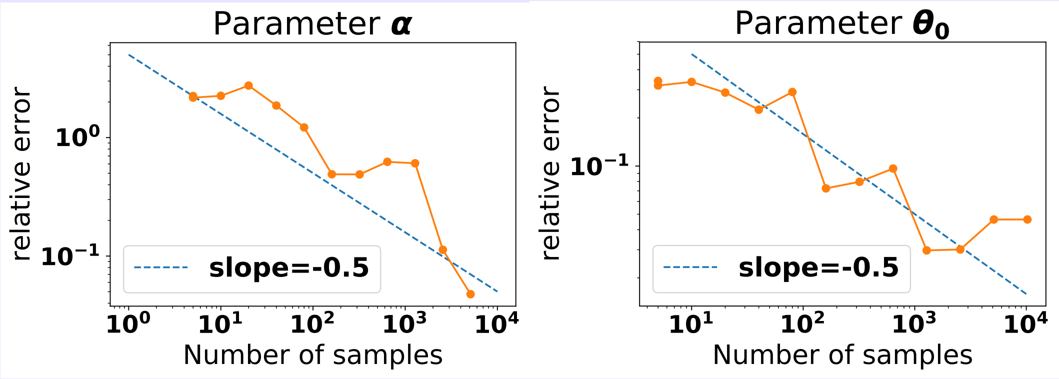


Figure 3: We have self-convergence of our algorithm at a rate that matches the convergence rate of Monte Carlo.

Future Wind Power Simulation

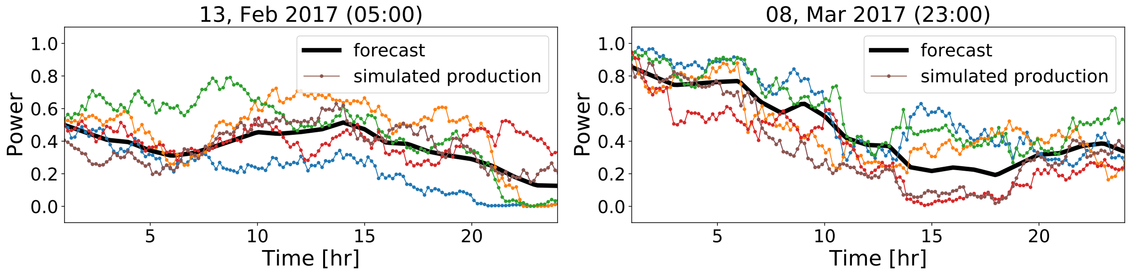


Figure 4: We simulate five possible future wind power production paths using the obtain optimal parameters $(\theta_0, \alpha) = (12, 0.3)$

Confidence Bands

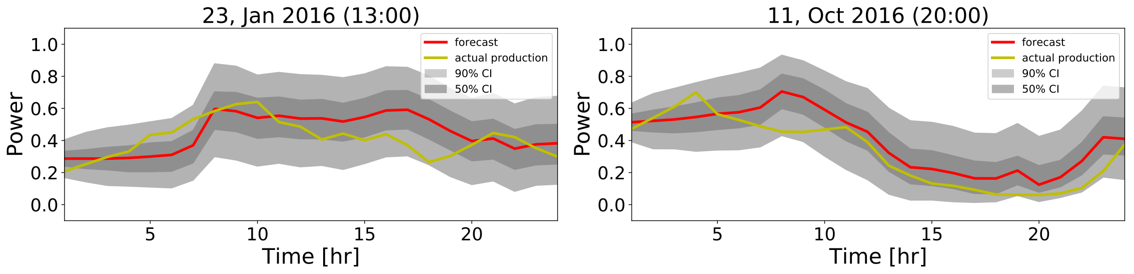


Figure 5: We obtain confidence intervals for future wind power production using the obtain optimal parameters $(\theta_0, \alpha) = (12, 0.3)$. Actual production plotted in retrospect.

State-Independent diffusion formulation

The model we have demonstrated previously is state-dependent, that is the diffusion of the SDE depends on the state.

$$\begin{aligned} dV_t &= -\theta_t V_t \; dt + \sqrt{2\theta_t \alpha (V_t + p_t)(1 - V_t - p_t)} \; dW_t \\ V_0 &= v_0 \end{aligned} \tag{11}$$

Why are we interested in a state-independent diffusion formulation? Because it's more tractable and numerically stable.
We apply a **Lamperti transform** to obtain the following state-independent diffusion SDE,

$$\begin{aligned} dZ_t &= \frac{-\theta_t(1 + \sin(Z_t) - 2p_t) + \alpha\theta_t \sin(Z_t)}{\cos(Z_t)} \; dt + \sqrt{2\alpha\theta_t} dW_t \\ Z_0 &= z_0 \end{aligned} \tag{12}$$

where $Z_t = \arcsin\left(\frac{1}{2}(V_t + p_t) - 1\right)$.

Data Skewness after Lamperti transformation

As stated before, the state-independent diffusion SDE follows a Lamperti transformed process Z_t given by $Z_t = \arcsin\left(\frac{1}{2}(V_t + p_t) - 1\right)$.

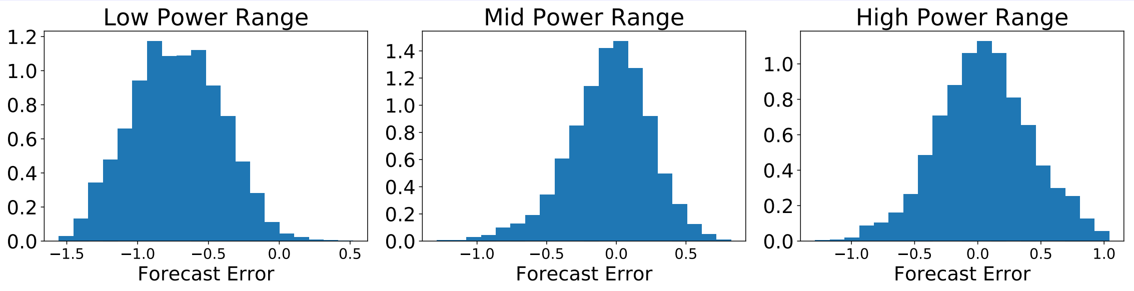


Figure 6: We observe that skewness has been greatly reduced after the Lamperti transformation. This motivates us to use a Gaussian transition density as a proxy density.

State-Independent diffusion formulation

Similarly, we try to 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,

$$\begin{aligned} \frac{dm_1(t)}{dt} &= -m_1(t)\theta_t(1-\alpha) - \theta(1-2p_t) \\ \frac{dvar(t)}{dt} &= 2var(t)\theta_t(2p_t - 1)\tan(m_1(t))\sec(m_1(t)) + \theta_t(\alpha - 1)\sec^2(m_1(t)) + 2\theta_t\alpha \end{aligned} \tag{13}$$

with initial conditions, $m_1(t_{n-1}) = v_{n-1}$ and $var(t_{n-1}) = v_{n-1}^2 - v_{n-1}$ where t_{n-1} is the time of the previous obseravtion and v_{n-1} is its value.

These are not exact ODEs for the centered moments, however they are accurate enough for small time intervals.

Inference Results

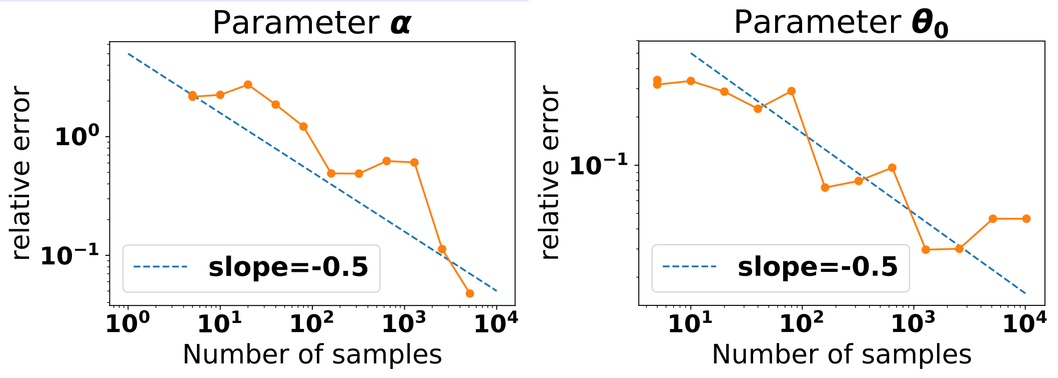


Figure 7: We have self-convergence of our algorithm in the Lamperti space at a rate that matches the convergence rate of Monte Carlo.

Results comparison in the different space

Formulation	parameters (θ_0, α)
Without Lamperti transform	$(12, 0.3)$
With Lamperti transform	$(,)$

Table 1: We compare the parameters obtained in both the original and Lamperti space.

Add slide for conference bands for the two different formulation

Concluding remarks

We were able to:

- simulate future wind power production based on real data.
- obtain an analytical description of the uncertainty of wind power forecasts in the form of an SDE.
- develop a forecasting technology agnostic method.
- capture skewness of the error process and its dynamics.

References

- Alhaddad, W. , Kebaier, A. , & Tempone, R. (2019). Stochastic Wind Power Forecasting. In preparation.
- 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
- 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
- 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, 2nd edition. ANEMOS.plus. https://doi.org/10.11581/DTU:00000017
- Chang, W.-Y. (2014) A Literature Review of Wind Forecasting Methods. Journal of Power and Energy Engineering, 2, 161-168. http://dx.doi.org/10.4236/jpee.2014.24023