Stochastic Wind Power Forecasting

Renzo Caballero[1], Ahmed Kebaier[2], Marco Scavino[3], and Raúl Tempone[4]

[1,4]CEMSE Division, King Abdullah University of Science and Technology (KAUST), Saudi Arabia
[2]Universit Paris 13, Sorbonne Paris Cit, LAGA, CNRS (UMR 7539), Villetaneuse, France
[3]Universidad de la República, Instituto de Estadística (IESTA), Montevideo, Uruguay
[4]Alexander von Humboldt Professor, RWTH Aachen University, Germany

April 14, 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 Uruguayan wind power production data and forecasts on the year 2019.

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

1 Introduction

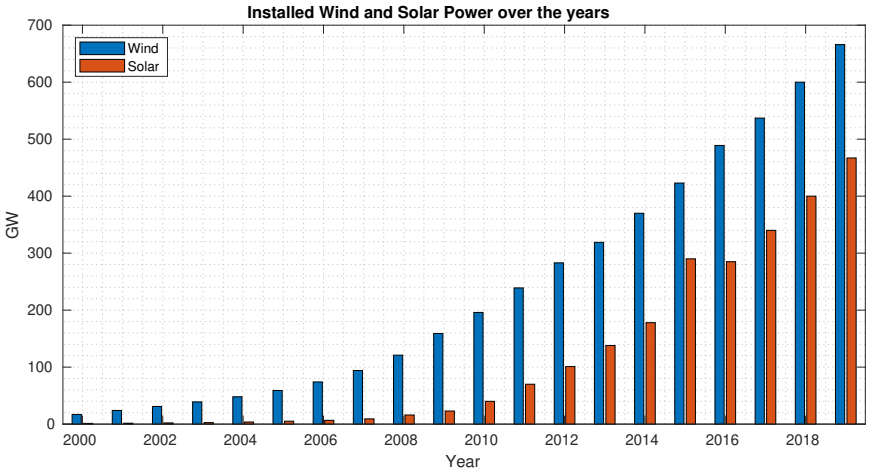

Figure 1: Installed wind and solar power over the years **sultana2017review**. We recall the importance of accurate forecasts to use green energies optimally.

Reliable wind power generation forecasting is crucial for the following applications (see, for example, **gieb**, **chang**, **zhbo**):

- Allocation of energy reserves such as water levels in dams or oil, and gas reserves.
- Operation scheduling of 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. Expand discussion about works on probabilistic forecasting. 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 (**mozuma**) considered stochastic wind power forecast models based on stochastic differential equations. 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 wind power forecasts together with historical wind power production data pertaining to the year 2019.

Change this paragraph. Discuss with the global dataset, the dataset after removing days when curtailing has been detected, the data set without curtailing and with Lamperti transform. We have available a year long data set from Uruguay based on 363 observation paths, each of which is 24-hours long with observations recorded every 10 min. In total, it is a data set of approximately fifty thousand data points recorded in 2019. See Figure (2). Data are normalized with respect to the maximum power capacity of wind power production in Uruguay, which is 1474 MW. However, sometimes the real production is artificially modifies due to curtailing. This effect becomes clear when we inspect for different power production levels. We split the available data into low, medium and high power range. See Figure (3).
We identify and remove the days with curtailing, removing a total of 108 days. From the resulting 255 days, we use 127 days to train the system and 128 to test it. Also, to avoid correlation between days, we intercalate the days we use for training and testing (Show that 24h is enough to ensure independence).

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.

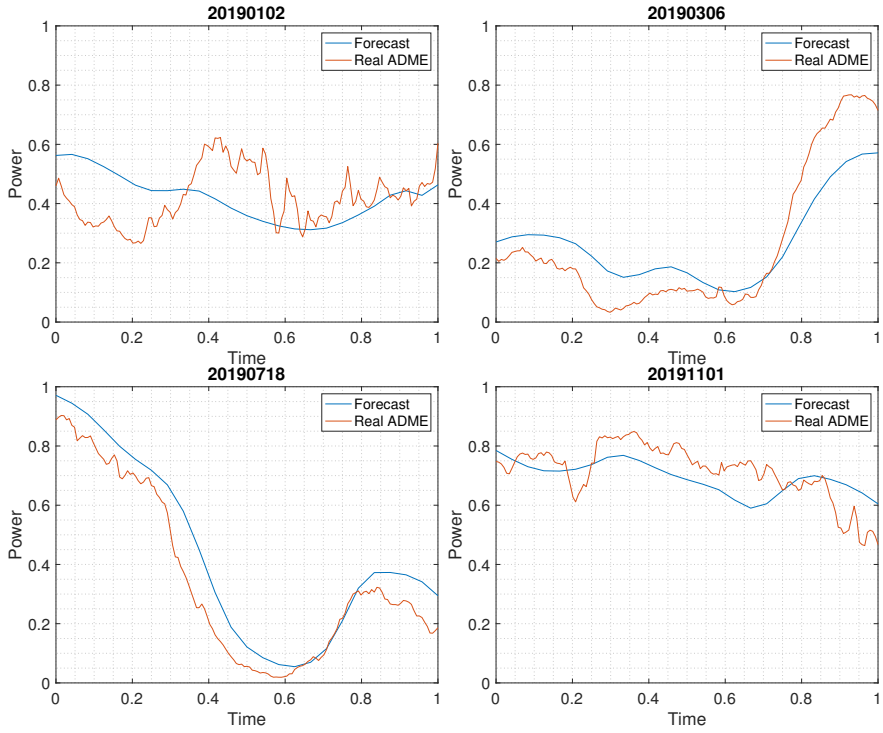

Figure 2: Four samples from the Uruguayan data of 2019. Each sample compromises of two 24-hour paths. In blue is an hourly wind power production forecast. In orange is the actual wind power production recorded in 10 minute intervals.

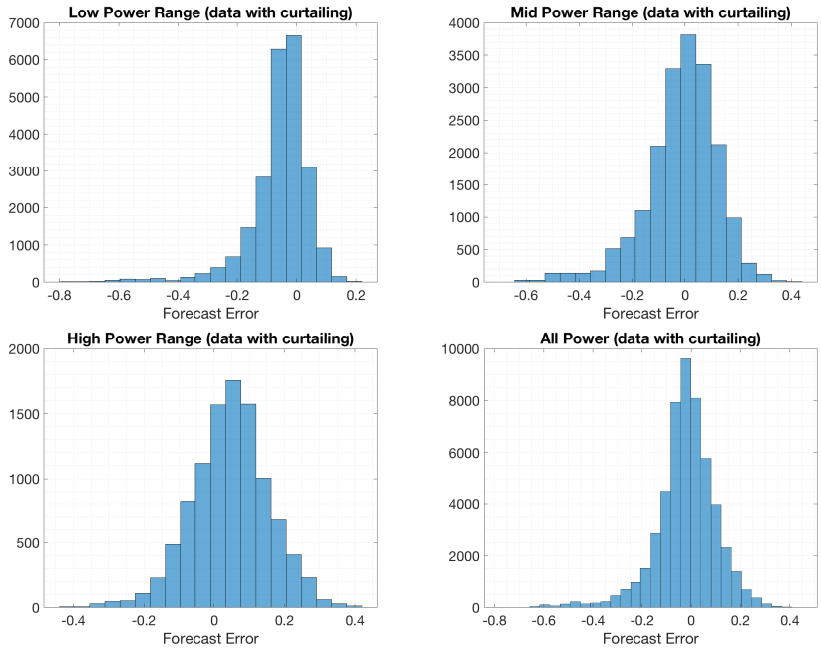

Figure 3: We see that forecast errors exhibit skewness, which is exaggerated as a consequence of curtailing in the production. Low power is when produced power is in $[0, 0.3]$, mid-power is when it is in $(0.3, 0.6]$, and high power when it is in $(0.6, 1]$. We have a total of 363 days to use.

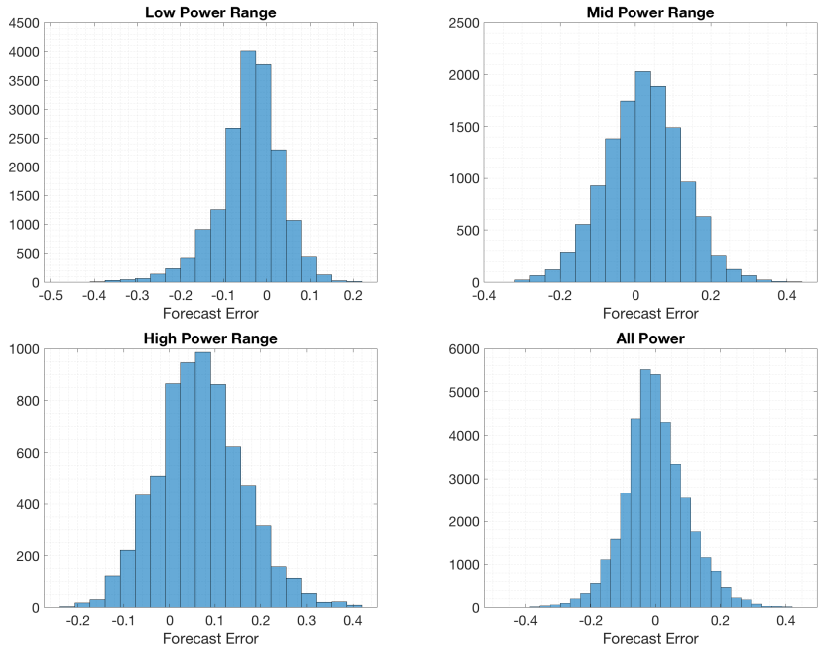

Figure 4: We observe that skewness has been reduced after removing the days with curtailing. We have a total of 255 days with no curtailing.

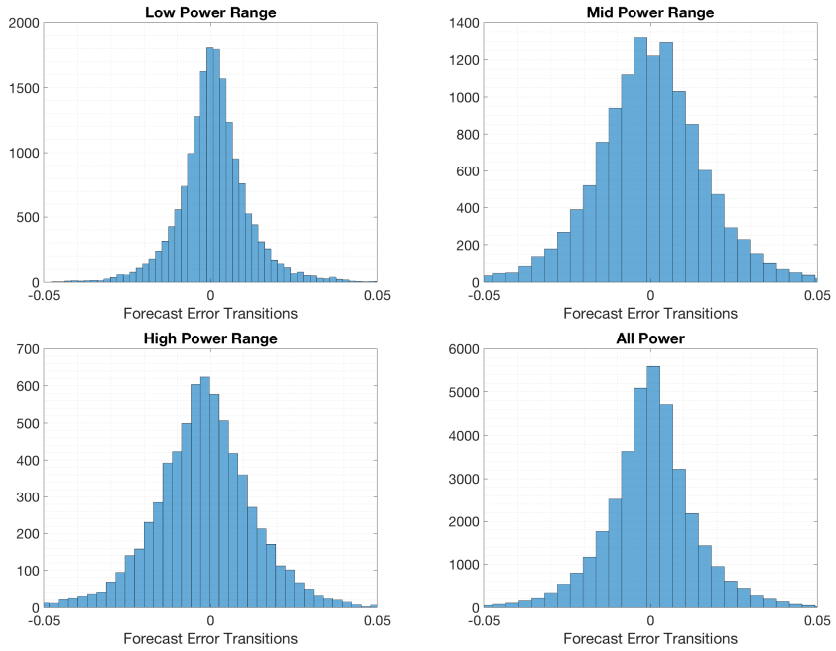

Figure 5: Transitions in error between forecast and real production. We can observe that they are not totally centered around zero.

2 Phenomenological Model

In this section, once we analyzed the available dataset, we consider a type of phenomenological model for the normalized wind power generation forecasts which is given, in its most general form, as a stochastic process $X = \{X_t,\, t \in [0,T]\}$ defined by the following stochastic differential equation (SDE):

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

where

- $a(\cdot; p_t, \dot{p}_t, \boldsymbol{\theta}) : [0,1] \to \mathbb{R}$ denotes a drift function,
- $b(\cdot; p_t, \dot{p}_t, \boldsymbol{\theta}) : [0,1] \to \mathbb{R}_+$ a diffusion function,
- $\boldsymbol{\theta}$ is a vector of unknown parameters,

- $(p_t)_{t\in[0,T]}$ is a time-dependent deterministic function $[0,1]$-valued and $(\dot{p}_t)_{t\in[0,T]}$ is its time derivative,
- $\{W_t,\, t\in[0,T]\}$ is a standard real-valued Wiener process.

In this work, $(p_t)_{t\in[0,T]}$ is to be considered as a deterministic forecast for the normalized wind power generation, which is provided by an official source.

Our goal is to achieve a specification of the model (1) to follow closely the available normalized wind power forecasts while ensuring its unbiasedness with respect to the forecast.

2.1 Physical Constraints

Let $(p_t)_{t\in[0,T]}$ be the available prediction function for the normalized wind power, which is an input to this approach. To start with the model specification, first we introduce a time-dependent drift function that features the mean-reverting property as well as derivative tracking:

$$a(X_t; p_t, \dot{p}_t, \boldsymbol{\theta}) = \dot{p}_t - \theta_t(X_t - p_t)\,, \tag{2}$$

where $(\theta_t)_{t\in[0,T]}$ is a positive deterministic function, whose range depends on $\boldsymbol{\theta}$, as will be explained shortly.

Now, looking at the normalized wind power generation forecast process X, modeled as solution to the Itô stochastic differential equation (1) with the drift specified in (2), it is straightforward to check that $\mathbb{E}X_t = p_t$, given $\mathbb{E}X_0 = p_0$. The application of Itô's lemma on $g(X_t,t) = X_t e^{\int_0^t \theta_s ds}$, leads to

$$d\left(X_t e^{\int_0^t \theta_s ds}\right) = e^{\int_0^t \theta_s ds}(\dot{p}_t + \theta_t p_t)dt + e^{\int_0^t \theta_s ds}\, b(X_t; p_t, \dot{p}_t, \boldsymbol{\theta})\, dW_t\,,$$

whose integral form is

$$e^{\int_0^t \theta_s ds} X_t - X_0 = \int_0^t (\dot{p}_s + \theta_s p_s) e^{\int_0^s \theta_u du} ds + \int_0^t b(X_s; p_s, \dot{p}_s, \boldsymbol{\theta}) e^{\int_0^s \theta_u du}\, dW_s\,. \tag{3}$$

Taking expectation on (3) we obtain

$$\begin{aligned}
\mathbb{E}X_t &= e^{-\int_0^t \theta_s ds}\left(\mathbb{E}X_0 + \int_0^t (\dot{p}_s + \theta_s p_s) e^{\int_0^s \theta_u du} ds\right) \\
&= e^{-\int_0^t \theta_s ds}\left(\mathbb{E}X_0 + \int_0^t \dot{p}_s e^{\int_0^s \theta_u du} ds + \left[p_s\, e^{\int_0^s \theta_u du}\right]_0^t - \int_0^t \dot{p}_s e^{\int_0^s \theta_u du} ds\right) \\
&= e^{-\int_0^t \theta_s ds}\left(\mathbb{E}X_0 + p_t\, e^{\int_0^t \theta_s ds} - p_0\right) = p_t\,.
\end{aligned} \tag{4}$$

At this stage, the process defined by (1) with drift (2) satisfies the two following properties:

- it reverts to its mean p_t, with a time-varying speed θ_t that is proportional to the deviation of the process X from its mean,
- it tracks the time derivative $\dot{p}_t$.

Remark: Observe that a mean-reverting model without derivative tracking shows a delayed path behavior. For instance, consider the diffusion model (1) with $a(X_t; p_t, \boldsymbol{\theta}) = -\theta_0(X_t - p_t), \theta_0 > 0$. In this case, given $\mathbb{E}X_0 = p_0$, the diffusion has mean $\mathbb{E}X_t = p_t - e^{-\theta_0 t}\int_0^t \dot{p}_s e^{\theta_0 s} ds$. The next Figure (6) illustrates paths of two diffusions with and without derivative tracking.

Figure 6: Add Figure showing the difference with and without derivative tracking

The forecast and production wind power data of Uruguay are normalized with respect to the installed power capacity during the period of observation. Thus, the mean-reverting level lies in $[0, 1]$, and the process X must take values in the same interval, a requirement that is not automatically fulfilled through the derivative tracking. To impose that the state space of X is $[0, 1]$, we may choose a convenient diffusion term, and require that the time-varying parameter θ_t satisfies an ad-hoc condition.

Let $\boldsymbol{\theta} = (\theta_0, \alpha)$, and choose a state-dependent diffusion term that avoids the process exiting from the range $[0, 1]$ as follows:

$$b(X_t; \boldsymbol{\theta}) = \sqrt{2\alpha\theta_0 X_t(1 - X_t)} \tag{5}$$

where $\alpha > 0$ is a unknown parameter that controls the path variability. This diffusion term belongs to the Pearson diffusion family and, in particular, it defines a Jacobi type diffusion. It is useful to recall that (**foso**) a Pearson diffusion is a stationary solution to a stochastic differential equation of the form

$$dX_t = -\theta(X_t - \mu)dt + \sqrt{2\theta(aX_t^2 + bX_t + c)}dW_t \tag{6}$$

where $\theta > 0$, and a, b and c are parameters such that the square root is well defined when X_t is in the state space. These parameters, together with μ, the mean of the invariant distribution, determine the state space of the diffusion as well as the shape of the invariant distribution.

An exhaustive classification of the (stationary) Pearson diffusions is presented in (**foso**) where, in particular, it is discussed the case $a < 0$ and $b(x;\boldsymbol{\theta}) = \sqrt{2a\theta x(x-1)}$, where the invariant distribution is a Beta distribution with parameters $\left(\frac{\mu}{-a}, \frac{1-\mu}{-a}\right)$, that leads to the well-known Jacobi diffusions, so-called because the eigenfunctions of the infinitesimal generator of these processes are the Jacobi polynomials (see, for example, **leph**).

It is worth mentioning that Jacobi diffusions have been successfully applied in several disciplines, among them finance (see **vago** and references therein) and neuroscience (**dotala**).

However, a distinctive feature in our proposed model

$$\begin{cases} dX_t &= (\dot{p}_t - \theta_t(X_t - p_t))dt + \sqrt{2\alpha\theta_0 X_t(1-X_t)}dW_t\,, \quad t \in [0,T] \\ X_0 &= x_0 \in [0,1]\,, \end{cases} \tag{7}$$

is that the drift term contains the time-varying parameter θ_t, rendering the solution X to (7) a non-stationary and time-inhomogeneous process. To ensure that the process X be the unique strong solution to (7) for all $t \in [0,T]$ with state space $[0,1]$ a.s., the mean-reversion time-varying parameter must satisfy the following condition:

$$\theta_t \geq \max\left(\frac{\alpha\theta_0 + \dot{p}_t}{1-p_t}, \frac{\alpha\theta_0 - \dot{p}_t}{p_t}\right). \tag{C}$$

The proof of this theoretical statement is presented in Section 9.

Remark: Condition (C) shows that the time-varying parameter θ_t becomes unbounded when $p_t = 0$ or $p_t = 1$. Therefore, we consider the following truncated prediction function

$$p_t^\epsilon = \begin{cases} \epsilon & \text{if} \quad p_t < \epsilon \\ p_t & \text{if} \quad \epsilon \leq p_t < 1-\epsilon \\ 1-\epsilon & \text{if} \quad p_t \geq 1-\epsilon \end{cases} \tag{8}$$

that satisfies $p_t^\epsilon \in [\epsilon, 1-\epsilon]$ for any $0 < \epsilon < \frac{1}{2}$ and $t \in [0,T]$, providing that θ_t is bounded for every $t \in [0,T]$.
From now on, we will keep the notation p_t to denote the truncated prediction function (8), unless specified otherwise.

2.2 A model specification for the forecast error

After applying to (7) the simple change of variables

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

we may introduce the following model for the forecast error of the normalized wind power production:

$$\begin{cases} dV_t &= -\theta_t V_t\, dt + \sqrt{2\alpha\theta_0(V_t+p_t)(1-V_t-p_t)}\, dW_t\,, \quad t\in[0,T] \\ V_0 &= v_0 \in [-1+\epsilon, 1-\epsilon]\,. \end{cases} \tag{9}$$

3 State independent diffusion term: Lamperti transform

Our model (9) 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 (9) by applying the so-called Lamperti transform (see **iacus1**, **moma**, **saso**) to the process V to obtain a SDE for the transformed process whose diffusion term does not depend anymore on the state variable. To this purpose, we consider the following Lamperti transformation

$$\begin{aligned} Z_t = h(V_t,t) &= \frac{1}{\sqrt{2\alpha\theta_0}} \int \frac{1}{\sqrt{(v+p_t)(1-v-p_t)}}\, dv \Bigg|_{v=V_t} \\ &= -\sqrt{\frac{2}{\alpha\theta_0}}\, \arcsin(\sqrt{1-V_t-p_t}) \end{aligned} \tag{10}$$

that, after applying Itô's formula on $h(V_t,t)$, leads to the following SDE with state independent unit diffusion term

$$\begin{aligned} dZ_t = \Bigg[&\frac{\dot{p}_t}{\sqrt{2\alpha\theta_0(V_t+p_t)(1-V_t-p_t)}} \\ &+ \frac{-\theta_t V_t}{\sqrt{2\alpha\theta_0(V_t+p_t)(1-V_t-p_t)}} - \frac{1}{4}\frac{\sqrt{2\alpha\theta_0}\,\big(1-2(V_t+p_t)\big)}{\sqrt{(V_t+p_t)(1-V_t-p_t)}} \Bigg]\, dt + dW_t\,. \end{aligned} \tag{11}$$

After replacing $V_t = 1 - p_t - \sin^2\left(-\sqrt{\frac{\alpha\theta_0}{2}}Z_t\right)$ in (11), we obtain that the process Z satisfies the SDE

$$
\begin{aligned}
dZ_t = &\left[\frac{\dot{p}_t - \theta_t\left(1 - p_t - \sin^2\left(-\sqrt{\frac{\alpha\theta_0}{2}}Z_t\right)\right)}{\sqrt{2\alpha\theta_0}\cos\left(-\sqrt{\frac{\alpha\theta_0}{2}}Z_t\right)\sin\left(-\sqrt{\frac{\alpha\theta_0}{2}}Z_t\right)}\right. \\
&\left. - \frac{1}{4}\frac{\sqrt{2\alpha\theta_0}\left(1 - 2\cos^2\left(-\sqrt{\frac{\alpha\theta_0}{2}}Z_t\right)\right)}{\cos\left(-\sqrt{\frac{\alpha\theta_0}{2}}Z_t\right)\sin\left(-\sqrt{\frac{\alpha\theta_0}{2}}Z_t\right)}\right] dt + dW_t \\
= &\left[\frac{2\dot{p}_t - \theta_t(1-2p_t) + (\alpha\theta_0 - \theta_t)\cos(-\sqrt{2\alpha\theta_0}Z_t)}{\sqrt{2\alpha\theta_0}\sin\left(-\sqrt{2\alpha\theta_0}Z_t\right)}\right] dt + dW_t\,. \qquad (12)
\end{aligned}
$$

We can see in Figure (7) the effect of the Lamperti transformation upon the forecast error data.

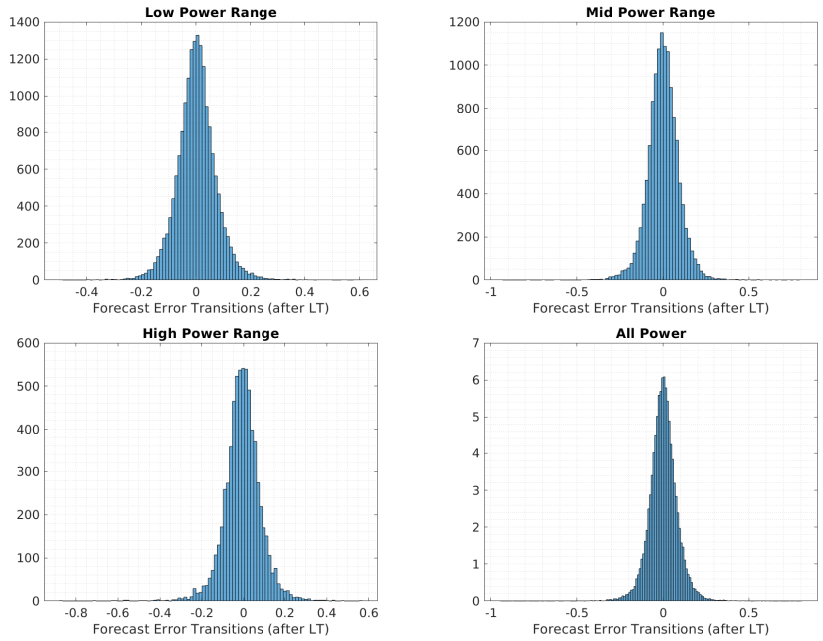

Figure 7: Transitions after the Lamperti transform. The transitions now present a normal Gaussian shape. This effect motivates the use of Gaussian proxies to approximate the process after using the Lamperti transform.

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 $V = \{V_t, t \in [0, T]\}$ is sampled at $N+1$ equispaced discrete points with given length interval Δ, and let $V^{M,N+1} = \{V_{t_1^{N+1}}, V_{t_2^{N+1}}, \ldots, V_{t_M^{N+1}}\}$ denote this random sample, with $V_{t_j^{N+1}} = \{V_{t_j+i\Delta}, i = 0, \ldots, N\}$, $j = 1, \ldots, M$.

Let $\rho(v|v_{j,i-1}; \boldsymbol{\theta})$ be the conditional probability density of $V_{t_j+i\Delta} \equiv V_{j,i}$ given $V_{j,i-1} = v_{j,i-1}$ evaluated at v, where $\boldsymbol{\theta} = (\theta_0, \alpha, \epsilon)$ are the unknown model parameters.

The Itô process V defined by the SDE (9) is Markovian, then the likelihood function of the sample $V^{M,N+1}$ can be written as the following product of transition densities:

$$\mathcal{L}(\boldsymbol{\theta}; V^{M,N+1}) = \prod_{j=1}^{M} \left\{ \prod_{i=1}^{N} \rho(V_{j,i}|V_{j,i-1}; p_{[t_{j,i-1}, t_{j,i}]}, \boldsymbol{\theta}) \right\}, \tag{13}$$

where $t_{j,i} \equiv t_j + i\Delta$ for any $j = 1, \ldots, M$ and $i = 0, \ldots, N$.

Remark: The statistical model (13) can be generalized incorporating the transition that occurs during the time interval, say δ, between the time when the forecast is done and the first date of forecasting. In this case, the inner product in (13) must include for any of the M paths an additional factor, say $\rho_0(V_{j,0}|V_{j,-\delta}; \boldsymbol{\theta}, \delta)$, expressing the conditional density of this early transition. The parameter δ can be calibrated after the estimation of $\boldsymbol{\theta}$, suggesting an optimal time for the scheduling of the forecasts.

The exact computation of the likelihood (13) 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 (**iacus1**, **saso**):

$$\frac{\partial f}{\partial t}\rho(v,t|v_{j,i-1},t_{j,i-1};\boldsymbol{\theta}) = -\frac{\partial}{\partial v}(-\theta_t v\,\rho(v,t|v_{j,i-1},t_{j,i-1};\boldsymbol{\theta})) + \frac{1}{2}\frac{\partial^2}{\partial v^2}(2\theta_0\alpha(v+p_t)(1-v-p_t)\,\rho(v,t|v_{j,i-1},t_{j,i-1};\boldsymbol{\theta}))\,, \tag{14}$$

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

Closed-form solutions to initial-boundary value problem for time-inhomogeneous diffusions can be obtained only in a few cases. Besides, in our case solving numerically (14) for the transition densities of the process V 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. Rewrite: Parametric estimation problems for diffusion processes sampled at discrete times are presented in (**iacus1**), and a survey of estimation methods for the parameter vector of the general one-dimensional, time-homogeneous SDE from a single sample of observations at discrete times is presented in (**hurn**).

4.2 Approximate Likelihood

Gaussian approximations to the transition densities of nonlinear time-inhomogeneous SDEs are available through different algorithms (**saso**). However, as Figure (5) may suggest at a first glance, the choice of a Gaussian density could be inadequate when straightly applied to approximate the transition density of the forecast error V of the normalized wind power production.

Therefore, we propose to use a surrogate transition density for V other than Gaussian. The moments of the SDE model (9) are then matched to the moments of the surrogate density.

From (4), we have $m_1(t) \equiv \mathbb{E}V_t = e^{-\int_{t_{j,i-1}}^{t}\theta_s ds}\,\mathbb{E}V_{t_{j,i-1}}$, for any $t \in [t_{j,i-1}, t_{j,i})$, $j = 1,\ldots,M$ and $i = 1,\ldots,N$.

For $m \geq 2$, using Itô's lemma on $g(V_t) = V_t^m$, we obtain

$$d(V_t^m) = \left(-m\theta_t V_t^m + \frac{1}{2}m(m-1)V_t^{m-2}2\alpha\theta_0(V_t+p_t)(1-V_t-p_t)\right)dt + mV_t^{m-1}2\alpha\theta_0(V_t+p_t)(1-V_t-p_t)dW_t\,,$$

from which we derive

$$\frac{d\mathbb{E}[V_t^m]}{dt} = -m\theta_t\mathbb{E}[V_t^m]+m(m-1)\alpha\theta_0\mathbb{E}[-V_t^m+(1-2p_t)V_t^{m-1}+p_t(1-p_t)V_t^{m-2}]\,. \tag{15}$$

For any $t \in [t_{j,i-1}, t_{j,i})$, the first two moments of V, $m_1(t)$ and $m_2(t) \equiv \mathbb{E}[V_t^2]$, can be computed by solving the following system

$$\begin{cases} \frac{dm_1(t)}{dt} &= -m_1(t)\theta_t \\ \frac{dm_2(t)}{dt} &= -2(\theta_t+\alpha\theta_0)m_2(t) + 2\alpha\theta_0(1-2p_t)m_1(t) + 2\alpha\theta_0 p_t(1-p_t) \end{cases} \tag{16}$$

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

Moment Matching

A suitable candidate for a surrogate transition density of V is a Beta distribution on a compact interval parameterized by two positive shape real parameters, ξ_1, ξ_2.

To approximate the transition densities of the process V using a Beta distribution, we equal the first two central moments of V with the corresponding moments of the Beta surrogate distribution on $[-1+\epsilon, 1-\epsilon]$ with shape parameters ξ_1, ξ_2 .

The shape parameters are given by

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

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

The approximate log-likelihood $\tilde{\ell}(\cdot; v^{V,N+1})$ of the observed sample $v^{V,N+1}$ can be expressed as

$$\tilde{\ell}(\boldsymbol{\theta}; v^{M,N+1}) = \sum_{j=1}^{M}\sum_{i=1}^{N} \log\left\{ \frac{1}{2(1-\epsilon)} \frac{1}{B(\xi_1,\xi_2)} \Big(\frac{z_{j,i}+1-\epsilon}{2(1-\epsilon)}\Big)^{\xi_1-1} \Big(\frac{1-\epsilon-z_{j,i}}{2(1-\epsilon)}\Big)^{\xi_2-1} \right\}\,, \tag{18}$$

where the shape parameters ξ_1 and ξ_2, according to (17), depend on the quantities $\mu(t_{j,i};\boldsymbol{\theta})$ and $\sigma^2(t_{j,i};\boldsymbol{\theta})$ that are computed solving numerically the initial-value problem (16).

The transition density of the process Z, which has been defined through the Lamperti transformation (10) of V, can be conveniently approximated by a Gaussian surrogate density.

The drift coefficient $a(Z_t; p_t, \dot{p}_t, \boldsymbol{\theta})$ of the process Z that satisfies (12) is nonlinear. After linearizing the drift around the mean of Z, $\mu_Z(t) \equiv \mathbb{E}Z_t$, we obtain the following system of ODEs to compute, for any $t \in [t_{j,i-1}, t_{j,i})$, the approximations of the first two central moments of Z, say $\tilde{\mu}_Z(t) \approx \mathbb{E}Z_t$ and $\tilde{v}_Z(t) \approx \mathrm{Var}Z_t$:

$$\begin{cases} \frac{d\tilde{\mu}_Z(t)}{dt} &= a\big(\tilde{\mu}_Z(t); p_t, \dot{p}_t, \boldsymbol{\theta}\big) \\ \frac{d\tilde{v}_Z(t)}{dt} &= 2a'\big(\tilde{\mu}_Z(t); p_t, \dot{p}_t, \boldsymbol{\theta}\big)\tilde{v}_Z(t) + 1 \end{cases} \tag{19}$$

with initial conditions $\tilde{\mu}_Z(t_{j,i-1}) = z_{j,i-1}$ and $\tilde{v}_Z(t_{j,i-1}) = 0$, and where

$$a'\big(\tilde{\mu}_Z(t); p_t, \dot{p}_t, \boldsymbol{\theta}\big) = \frac{(\alpha\theta_0 - \theta_t) - \cos(\sqrt{2\alpha\theta_0}Z_t)[\theta_t(1-2p_t) - 2\dot{p}_t]}{\sin^2\left(\sqrt{2\alpha\theta_0}Z_t\right)} .$$

The approximate log-likelihood $\tilde{\ell}(\cdot; z^{V,N+1})$ of the observed sample $z^{V,N+1}$ is given by

$$\tilde{\ell}(\boldsymbol{\theta}; z^{M,N+1}) = \sum_{j=1}^{M}\sum_{i=1}^{N} \log\left\{ \frac{1}{\sqrt{2\pi\tilde{v}_Z(t_{j,i};\boldsymbol{\theta})}} \exp\left(-\frac{(z_{j,i} - \tilde{\mu}_Z(t_{j,i};\boldsymbol{\theta}))^2}{2\tilde{v}_Z(t_{j,i};\boldsymbol{\theta})} \right) \right\} , \tag{20}$$

where $\tilde{\mu}_Z(t_{j,i};\boldsymbol{\theta})$ and $\tilde{v}_Z(t_{j,i};\boldsymbol{\theta})$ are computed solving numerically the initial-value problem (19).

4.3 Optimization

Rewrite the discussion to obtain fast initial estimate for the two unknow parameters.

To initialize the optimization process for the likelihood function of the process V, we solve the following least-squares problem which gives us first estimates of the mean reversion parameter θ_0: Add explanation regarding the motivation for this objective function.

$$\theta_0 \approx \arg\min_{\theta} \sum_{j}^{M}\sum_{i}^{N} \Big(v_{i+1,j} - v_{i,j} - \big(-\theta_t v_{i,j}\big)\left(t_{i+1,j} - t_{i,j}\right) \Big)^2 , \tag{21}$$

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 not ergodicity, instead simply for the **quadratic variation formula**., we can obtain a first estimate on the product of the parameters as follows,

$$\theta_0\alpha \approx \frac{1}{2M\Delta t}\sum_{j}^{M}\frac{\sum_{i}^{N}(x_{i+1,j}-x_{i,j})^2}{\sum_{i}^{N}x_{i,j}(1-x_{i,j})}\,. \tag{22}$$

Solving for α, we have both first estimates and we can also estimate δ solving the problem

$$\delta \approx \arg\min_{\delta}\mathcal{L}_{\delta}(\boldsymbol{\theta},\delta;V^{M,1}) = \arg\min_{\delta}\prod_{j=1}^{M}\rho_0(V_{j,t_0}|V_{j,t_{-\delta}};\boldsymbol{\theta},\delta)\,. \tag{23}$$

With our estimations of $\boldsymbol{\theta}$, we can start the inference processes 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 1: This model does not feature derivative tracking.

$$\begin{cases} dX_t &= -\theta_0(X_t-p_t)dt + \sqrt{2\alpha\theta_0 X_t(1-X_t)}dW_t\,, \quad t\in[0,T] \\ X_0 &= x_0\in[0,1]\,, \end{cases} \tag{24}$$

  with $\theta_0>0$, $\alpha>0$.

- Model 2: This model features derivative tracking and time-varying mean-reversion parameter:

$$\begin{cases} dX_t &= \big(\dot{p}_t-\theta_t(X_t-p_t)\big)dt + \sqrt{2\alpha\theta_0 X_t(1-X_t)}dW_t\,, \quad t\in[0,T] \\ X_0 &= x_0\in[0,1]\,, \end{cases} \tag{25}$$

  with $\theta_0>0$, $\alpha>0$ and .

Model	parameters $(\theta_0, \alpha, \delta)$	AIC	BIC
Model 1	([*insert*], [*insert*])	[insert]	[insert]
Model 2	$(1.2, 0.1, 0.6)$	[insert]	[insert]

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

6 Forecast Provider Comparison

We compare forecasts from two different companies for the same period.

Forecast Provider	parameters $(\theta_0, \alpha, \delta)$	AIC	BIC
Provider A	([*insert*], [*insert*])	[insert]	[insert]
Provider B	([*insert*], [*insert*])	[insert]	[insert]

Table 2: [insert]

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

7.1 Initial guesses

Following the instruction for the initial guesses from section 4.3, we get the initial guesses $(\theta_0, \alpha, \delta) \approx (1.2, 0.1, 0.6)$.

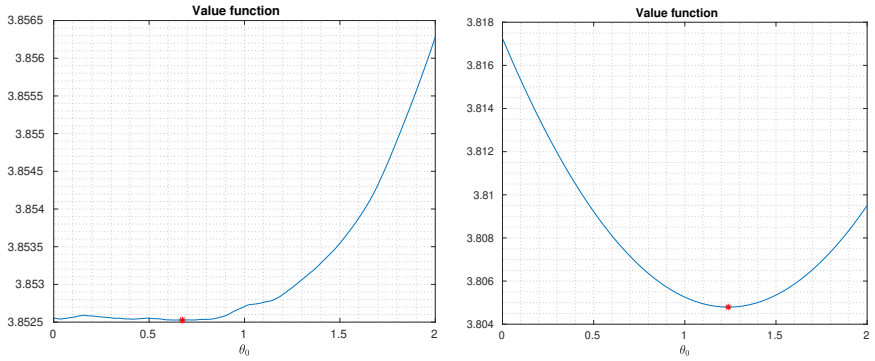

Figure 8: Value function from (21) as a function of different θ_0 values. We can observe that the function if flat for a wide range.

7.2 Negative Log-Lihelihood

We plot the Negative Log-Likelihood as a function of the parameters. The plots can be seen in Figure (9). We used all the training data (127 days of data) to construct and optimize the Negative Log-Likelihood.

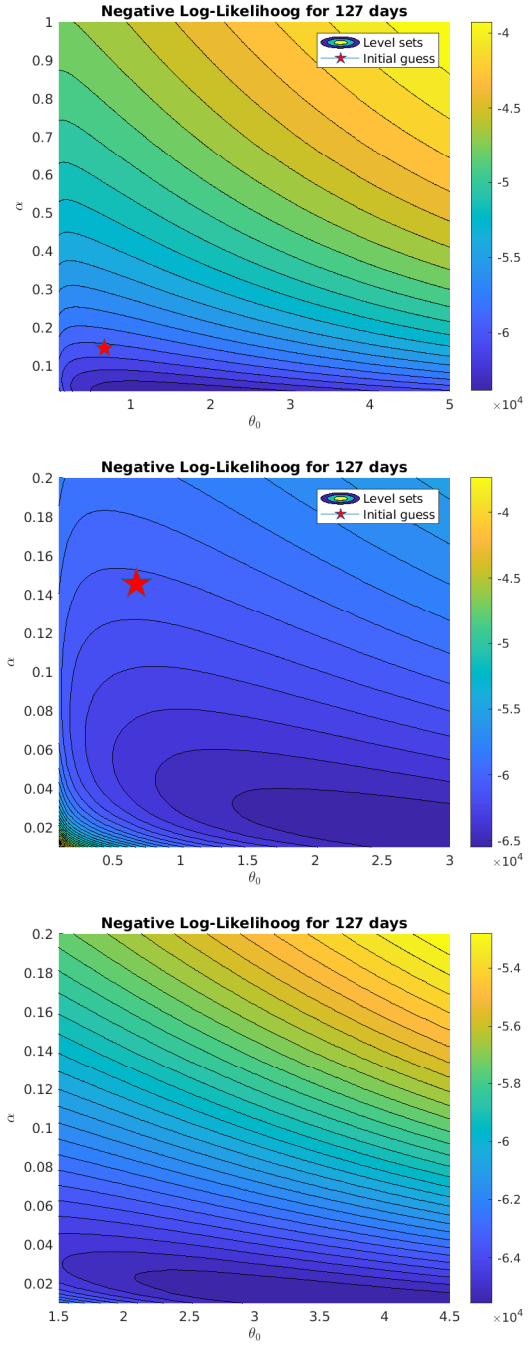

Figure 9: Negative Log-Likelihood for different domains.

7.3 Main guesses

We were able to obtain the following for

Formulation	parameters $(\theta_0, \alpha, \delta)$
Without Lamperti transform	$(1.2, 0.1, 0.6) \pm ([insert], [insert])$
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 2019[insert].

Remind that Lamperti transform has the purpose tho check the consistency of the models (if the estimation of $\boldsymbol{\theta}$ with V and Z differs it is alarming.)

Discuss to which extent it is needed more accuracy in the estimation of $\boldsymbol{\theta}$. How many data we need to achive enough accuracy for the estimates of the two parameters? Generally, such accuracy is application dependent

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.

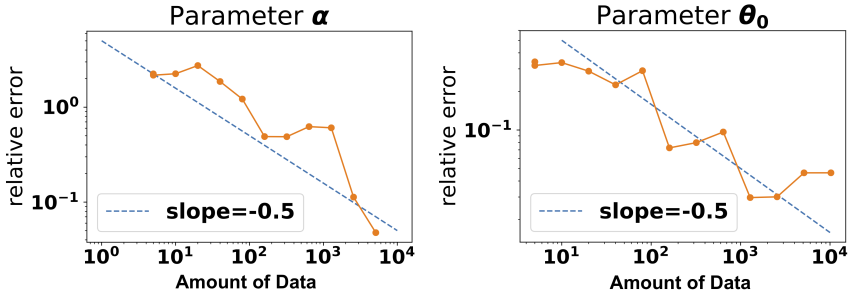

Figure 10: We show self-convergence of our algorithm applied to model 2. We conlcude that the rate matches that of of Monte Carlo. Data is from Uruguay pertaining to year 2019.

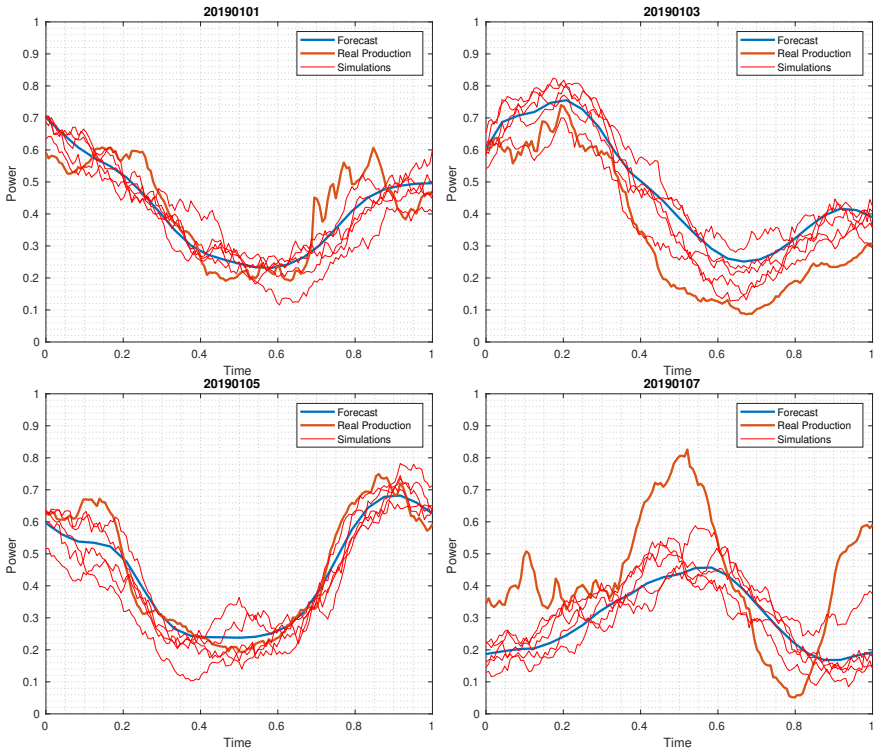

Figure 11: We simulate following model 2 five possible future wind power production paths using the obtained optimal parameters $(\theta_0, \alpha, \delta) = (1.2, 0.1, 0.6)$. Forecast is from Uruguay pertaining to year 2019.

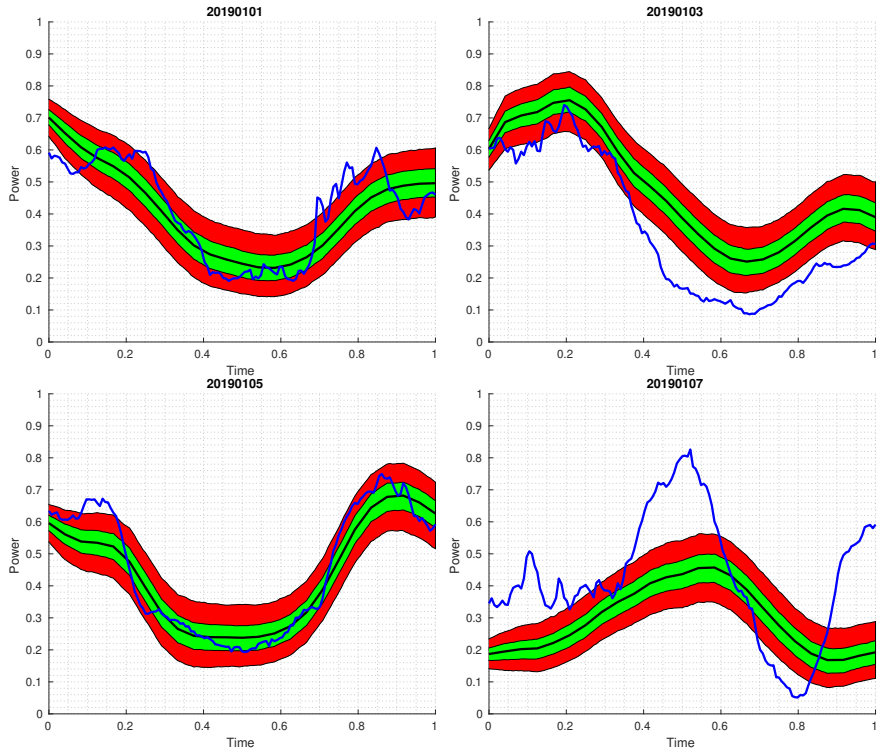

Figure 12: Pending add legend and choose better colors.We obtain confidence intervals following model 2 for future wind power production using the obtained optimal parameters $(\theta_0, \alpha, \delta) = (1.2, 0.1, 0.6)$. Actual production plotted in retrospect. Forecast and data is from Uruguay pertaining to year 2019.

8 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.

9 Appendix

10 The model

For a time horizon $T > 0$ and a parameter $\alpha > 0$ and $(\theta_t)_{t\in[0,T]}$ a positive deterministic function, let us consider the model given by

$$\begin{cases} dX_t &= \big(\dot{p}_t - \theta_t(X_t - p_t)\big)dt + \sqrt{2\alpha\theta_t X_t(1-X_t)}dW_t\,, \quad t \in [0,T] \\ X_0 &= x_0 \in [0,1]\,, \end{cases} \tag{26}$$

where $(p_t)_{t\in[0,T]}$ denotes the prediction function that satisfies $0 \leq p_t \leq 1$ for all $t \in [0,T]$. This prediction function is assumed to be a smooth function of the time so that

$$\sup_{t\in[0,T]} \big(|p_s| + |\dot{p}_s|\big) < +\infty.$$

The following proofs are based on standard arguments for stochastic processes that can be found e.g. in **Alf** and **KarShr** that we adapted to the setting of our model (26).

Theorem 10.1. Assume that

$$\forall t \in [0,T],\;\; 0 \leq \dot{p}_t + \theta_t p_t \leq \theta_t, \quad \text{and} \quad \sup_{t\in[0,T]} |\theta_t| < +\infty. \tag{A}$$

Then, there is a unique strong solution to (26) s.t. for all $t \in [0,T]$, $X_t \in [0,1]$ a.s.

Proof. Let us first consider the following SDE for $t \in [0,T]$

$$X_t = x_0 + \int_0^t \big(\dot{p}_s - \theta_s(X_s - p_s)\big)ds + \int_0^t \sqrt{2\alpha\theta_s|X_s(1-X_s)|}dW_s, \quad x_0 > 0. \tag{27}$$

According to Proposition 2.13, p.291 of **KarShr**, under assumption (A) there is a unique strong solution X to (27). Moreover, as the diffusion coefficient is of linear growth, we have for all $p > 0$

$$\mathbb{E}[\sup_{t\in[0,T]} |X_t|^p] < \infty. \tag{28}$$

Then, it remains to show that for all $t \in [0,T]$, $X_t \in [0,1]$ a.s. For this aim, we need to use the so-called Yamada function ψ_n that is a $\mathcal{C}^2$ function that satisfies a bench of useful properties:

$$|\psi_n(x)| \underset{n\to+\infty}{\to} |x|, \quad x{\psi'}_n(x) \underset{n\to+\infty}{\to} |x|, \quad |\psi_n(x)| \wedge |x{\psi'}_n(x)| \leq |x|$$

$${\psi'}_n(x) \leq 1, \quad \text{and } {\psi''}_n(x) = g_n(|x|) \geq 0 \quad \text{with} \quad g_n(x)x \leq \frac{2}{n} \quad \text{for all } x \in \mathbb{R}.$$

See the proof of Proposition 2.13, p. 291 of **KarShr** for the construction of such function. Applying It's formula we get

$$\begin{aligned}\psi_n(X_t) = \psi_n(x_0) &+ \int_0^t {\psi'}_n(X_s)(\dot{p}_s + \theta_s p_s - \theta_s X_s)ds \\ &+ \int_0^t {\psi'}_n(X_s)\sqrt{2\alpha\theta_s|X_s(1-X_s)|}dW_s + \alpha\int_0^t \theta_s g_n(|X_s|)|X_s(1-X_s)|ds.\end{aligned}$$

Now, thanks to (A), (28) and to the above properties of ψ_n and g_n, we get

$$\mathbb{E}[\psi_n(X_t)] \leq \psi_n(x_0) + \int_0^t \big(\dot{p}_s + \theta_s p_s - \theta_s \mathbb{E}[{\psi'}_n(X_s)X_s]\big)ds + \frac{2\alpha}{n}\int_0^t \theta_s \mathbb{E}|1-X_s|ds.$$

Therefore, letting n tends to infinity we use Lebesgue's theorem to get

$$\mathbb{E}[|X_t|] \leq x_0 + \int_0^t \big(\dot{p}_s + \theta_s p_s - \theta_s \mathbb{E}|X_s|\big)ds.$$

Besides, taking the expectation of (27), we get

$$\mathbb{E}X_t = x_0 + \int_0^t \big(\dot{p}_s + \theta_s p_s - \theta_s \mathbb{E}X_s\big)ds$$

and thus we have

$$\mathbb{E}[|X_t| - X_t] \leq \int_0^t \theta_s \mathbb{E}[|X_s| - X_s]ds.$$

Then, Gronwall's lemma gives us $\mathbb{E}[|X_t|] = \mathbb{E}X_t$ and thus for any $t \in [0,T]$ $X_t \geq 0$ a.s. The same arguments work to prove that for any $t \in [0,T]$ $Y_t := 1 - X_t \geq 0$ a.s. since the process $(Y_t)_{t\in[0,T]}$ is solution to

$$dY_t = \big(\theta_t(1-p_t) - \dot{p}_t - \theta_t Y_t\big)dt - \sqrt{2\alpha\theta_t Y_t(1-Y_t)}dW_t,$$

Then similarly, we need to assume that $\dot{p}_t + \theta_t p_t \geq 0$. This completes the proof. □

Theorem 10.2. Assume that assumptions of Theorem 10.1 hold with $x_0 \in]0,1[$. Let $\tau_0 := \inf\{t \in [0,T],\ X_t = 0\}$ and $\tau_1 := \inf\{t \in [0,T],\ X_t = 1\}$ with the convention that $\inf \emptyset = +\infty$. Assume in addition that

$$0 < \alpha < \frac{1}{2} \quad \text{and} \quad \forall t \in [0,T],\ \ \alpha\theta_t \leq \dot{p}_t + \theta_t p_t \leq (1-\alpha)\theta_t. \tag{B}$$

Then, $\tau_0 = \tau_1 = +\infty$ a.s.

Proof. For $t \in [0, \tau_0[$, we have

$$\frac{dX_t}{X_t} = \left(\frac{\dot{p}_t + \theta_t p_t}{X_t} - \theta_t\right) \mathrm{d}t + \sqrt{\frac{2\alpha\theta_t(1-X_t)}{X_t}}\, \mathrm{d}W_t$$

so that

$$X_t = x_0 \exp\Big(\int_0^t \frac{\dot{p}_s + \theta_s(p_s - \alpha)}{X_s} ds - (1-\alpha)\int_0^t \theta_s ds + M_t \Big),$$

where $M_t = \int_0^t \sqrt{\frac{2\alpha\theta_s(1-X_s)}{X_s}} dW_s$ is a continuous martingale. Then as for all $t \in [0,T]$, we have $\dot{p}_t + \theta_t(p_t - \alpha) \geq 0$, we deduce that

$$X_t \geq x_0 \exp\Big(-(1-\alpha)\int_0^t \theta_s ds + M_t \Big).$$

By way of contradiction let us assume that $\{\tau_0 < \infty\}$, then letting $t \to \tau_0$ we deduce that $\lim_{t\to\infty} \mathbf{1}_{\{\tau_0<\infty\}} M_{t\wedge\tau_0} = -1_{\{\tau_0<\infty\}}\infty$ a.s. This leads to a contradiction since we know that continuous martingales likewise the Brownian motion cannot converge almost surely to $+\infty$ or $-\infty$. It follows that $\tau_0 = \infty$ almost surely. Next, recalling that the process $(Y_t)_{t\geq 0}$ given by $Y_t = 1 - X_t$ is solution to

$$dY_t = \big(\theta_t(1-p_t) - \dot{p}_t - \theta_t Y_t\big)dt - \sqrt{2\alpha\theta_t Y_t(1-Y_t)}dW_t$$

we deduce using similar arguments as above $\tau_1 = \infty$ a.s. provided that $\theta_t(1-p_t) - \dot{p}_t - \alpha\theta_t \geq 0$. □

11 New model

In what follows we adapt the proof of the above theorem for the new model

$$\begin{cases} dX_t &= \big(\dot{p}_t - \theta_t(X_t - p_t)\big)dt + \sqrt{2\alpha\theta_0 X_t(1-X_t)}dW_t\,, \quad t \in [0,T] \\ X_0 &= x_0 \in [0,1]\,, \end{cases} \tag{29}$$

Theorem 11.1. Assume that assumptions of Theorem 10.1 hold with $x_0 \in]0,1[$. Let $\tau_0 := \inf\{t \in [0,T],\ X_t = 0\}$ and $\tau_1 := \inf\{t \in [0,T],\ X_t = 1\}$ with the convention that $\inf \emptyset = +\infty$. Assume in addition that for all $t \in [0,T]$, $p_t \in]0,1[$ and that

$$\theta_t \geq \max\left(\frac{\alpha\theta_0 + \dot{p}_t}{1-p_t}, \frac{\alpha\theta_0 - \dot{p}_t}{p_t}\right). \tag{C}$$

Then, $\tau_0 = \tau_1 = +\infty$ a.s.

Proof. For $t \in [0, \tau_0[$, we have

$$\frac{dX_t}{X_t} = \left(\frac{\dot{p}_t + \theta_t p_t}{X_t} - \theta_t\right) dt + \sqrt{\frac{2\alpha\theta_0(1-X_t)}{X_t}} dW_t$$

so that

$$X_t = x_0 \exp\Big(\int_0^t \frac{\dot{p}_s + \theta_s p_s - \theta_0\alpha)}{X_s} ds + \alpha\theta_0 t - \int_0^t \theta_s ds + M_t\Big),$$

where $M_t = \int_0^t \sqrt{\frac{2\alpha\theta_s(1-X_s)}{X_s}} dW_s$ is a continuous martingale. Then as for all $t \in [0,T]$, we have $\dot{p}_t + \theta_t p_t - \theta_0\alpha \geq 0$, we deduce that

$$X_t \geq x_0 \exp\Big(\alpha\theta_0 t - \int_0^t \theta_s ds + M_t\Big).$$

By way of contradiction let us assume that $\{\tau_0 < \infty\}$, then letting $t \to \tau_0$ we deduce that $\lim_{t\to\infty} \mathbf{1}_{\{\tau_0<\infty\}} M_{t\wedge\tau_0} = -1_{\{\tau_0<\infty\}}\infty$ a.s. This leads to a contradiction since we know that continuous martingales likewise the Brownian motion cannot converge almost surely to $+\infty$ or $-\infty$. It follows that $\tau_0 = \infty$ almost surely. Next, recalling that the process $(Y_t)_{t\geq 0}$ given by $Y_t = 1 - X_t$ is solution to

$$dY_t = \big(\theta_t(1-p_t) - \dot{p}_t - \theta_t Y_t\big)dt - \sqrt{2\alpha\theta_0 Y_t(1-Y_t)} dW_t$$

we deduce using similar arguments as above $\tau_1 = \infty$ a.s. provided that $\theta_t(1-p_t) - \dot{p}_t - \alpha\theta_0 \geq 0$. □

12 Answers to Renzo's questions

So, Theorem 11.1 answers the second question in slide 11. The new condition (C) that guarantees for $(X_t)_{t\geq 0}$ to stay in $]0,1[$ depends now on α. I would

keep condition (C) as we have a proof for it. Moreover, if condition (C) is satisfied, then the orange condition in slide 8 is also satisfied since we have

$$\max\left(\frac{\alpha\theta_0+\dot{p}_t}{1-p_t},\frac{\alpha\theta_0-\dot{p}_t}{p_t}\right)\geq\max\left(\frac{\frac{\alpha\theta_0}{2}+\dot{p}_t}{1-p_t},\frac{\frac{\alpha\theta_0}{2}-\dot{p}_t}{p_t}\right)$$

which ensures that Z don't hit the boundaries too according to Renzo's intuitive arguments.

Concerning, the first question on slide 11: we can't apply It's formula, unless we are sure that the process don't hit the boundaries, that's why we have to assume (C) before using It's formula. In addition, as the diffusion coefficient of X given by $x\mapsto\sqrt{\frac{2\alpha\theta_0(1-x)}{x}}$ is strictly positive for all x in the state space of X, namely $]0,1[$ under assumption (C), this is enough to ensure that the transformation between Z and X is bijective so that we deduce the existence and uniqueness of Z from those of X. Otherwise the Lamperti transform is not valid if the state space of X is $[0,1]$.