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

May 29, 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.

2020 AMS Subject Classification: 60H10

1 Introduction

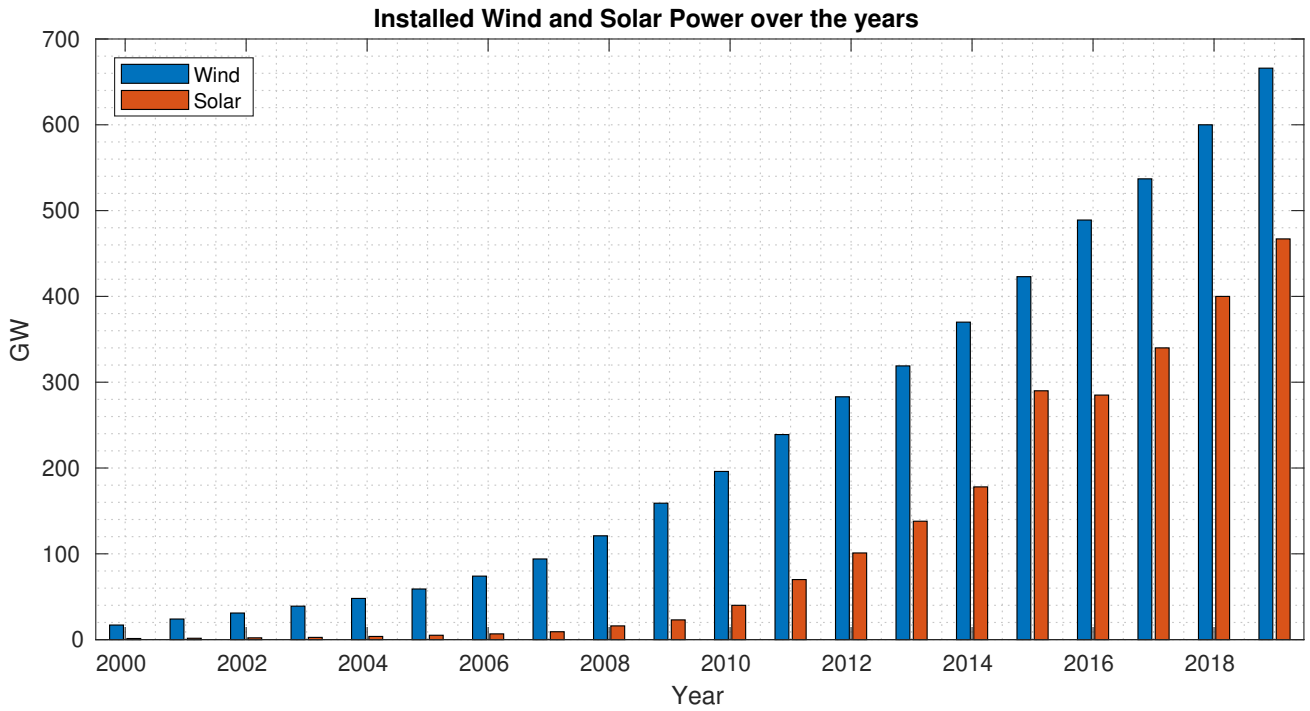

Figure 1: Installed wind and solar power over the years **sultana2017review**. We recall the importance of accurate forecasts to use green energies optimally.

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 the forecast's actual performance as it has been known that different forecasting technologies exhibit different behavior for different wind farms and seasons (**chang**). 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 (**schicker2017short**). 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 maximum 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 (**tsitsiklis2015pricing**).

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.

In Section 2, we introduce the main characteristics of a real data set encompassing the normalized wind power production in Uruguay during the period April-December 2019, joint with the most accurate predictions, as highlighted in our posterior analysis, performed by one out of the three sources of forecast providers. The significant steps for constructing phenomenological models of the normalized wind power production and the forecast error based on stochastic differential equations are described in Section 3. After reformulating the SDE for the forecast error in Section 4, by applying the Lamperti transform that leads to a unit diffusion coefficient, we write down in Section 5 the expressions for the likelihood functions of the forecast error in its original space and in the Lamperti space. We also derive simple approximations of the likelihood functions. Section 5 concludes with the description of the optimization algorithm to compute approximate maximum likelihood estimates, including the case where we expand the model comprising an initial transition from the time the forecast is performed to the

time of the first forecast. In Section 6, we apply our proposed parametric estimation procedure to the Uruguay wind and forecast dataset, comparing two alternative models and assessing, for the best candidate model, the performance of three different forecast providers. The proofs of the existence, strong uniqueness, and boundedness of the solutions of the SDE used to model normalized wind power production and its forecast error are given in the Appendix.

2 Wind power production data in Uruguay and forecast providers

In recent years, Uruguay has triggered a remarkable change in its energy matrix. In (**irena**, p.23) Uruguay is among those countries showcasing innovation, like Denmark, Ireland, Germany, Portugal, and Spain, with proven feasibility of managing annual variable renewable energy (VRE) higher than 25% in power systems.

According to (**ren21**, pp.118-119), in 2018, Uruguay achieved 36% of its electricity production from variable wind energy and solar PV, raising the share of generation from wind energy more than five-fold in just four years, from 6.2% in 2014 to 33% in 2018.

At present, Uruguay is fostering even higher levels of wind penetration by boosting regional power trading with Argentina and Brazil. In this rapidly evolving scenario, it is essential to analyze national data on wind power production joint with wind power short-term forecastings to orientate and assess the strategies and decisions of wind energy actors and businesses.

Our study is based on publicly available data (source: Administrator of Electric Market) on the wind power production in Uruguay for the period April-December 2019, that we adequately normalized with respect to the present 1474 MW maximum installed wind power capacity. Each day, wind power production recordings are available every ten minutes. In this work, we have also considered data from three different forecast providers, available each day starting at 1 pm.

The next Figure 2 shows the wind power real production during four segments 24-hour long selected from the observation period together with their corresponding hourly short-term forecast, computed by a forecast provider. For the sake of visualization clarity, this Section relies only on forecasts from one provider, called "provider A" from now on, ranked as the most accurate forecast provider, as it emerged from our posterior analysis.

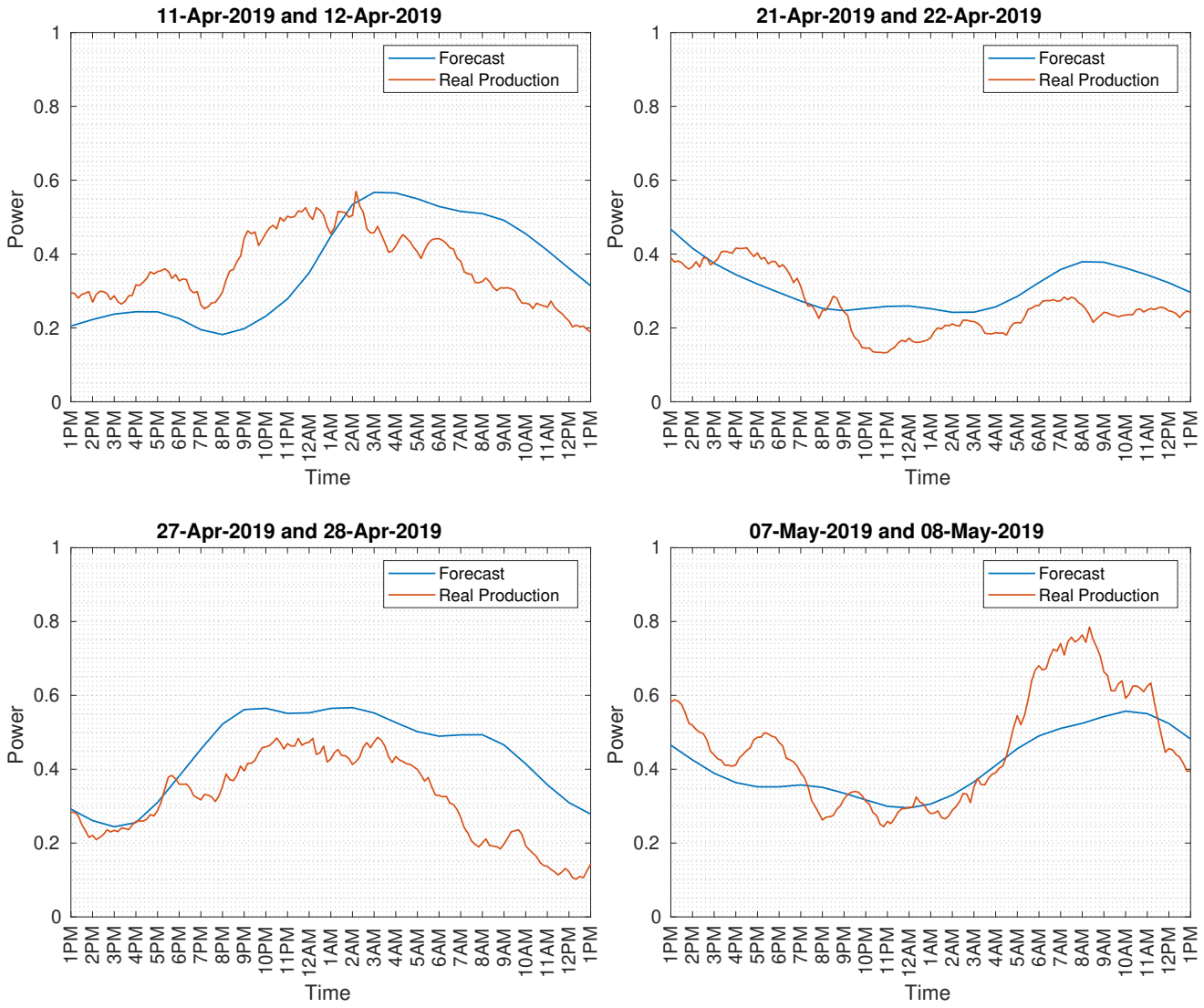

Figure 2: Four 24-hour segments with the wind power real production in Uruguay (orange line) recorded every ten minutes, and the hourly wind power production forecasted by provider A (blue line).

A view of the global discrepancy between the real production and the forecasted production, during the nine months observation period, is summarized through the forecast error histograms, Figure 3, where we also partitioned the forecast errors according to three contiguous categories of normalized generated power. Low normalized generated power corresponds to the range $[0, 0.3]$, mid-power refers to the range $(0.3, 0.6]$, and high-power to the range $(0.6, 1]$.

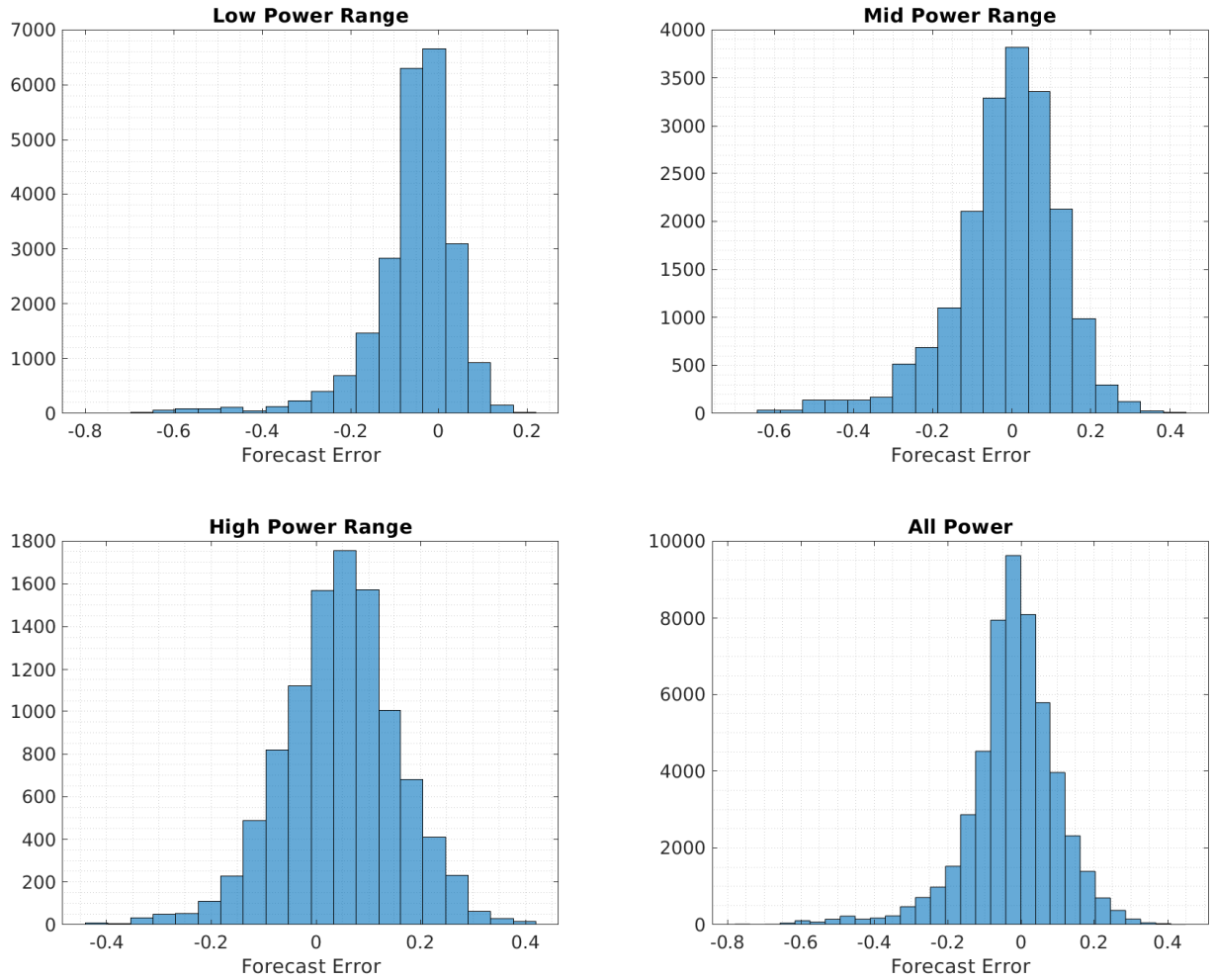

Figure 3: Wind production forecast error histograms during the period April-December 2019: low-power (upper-left graph), mid-power (upper-right graph), high-power (lower-left graph), and the global range of power (lower-right graph).

We may observe that all the histograms in the Figure 3 exhibit skewed patterns, to a different extent, as well as extreme observations. The presence of these features can be partly explained. The analysis of the dataset highlighted that, during several 24-hour segments, the system operators decided to reduce, or even cease the wind power production. Indeed, as recalled in (**irena2**, p.3), "Uruguay experiences high curtailment levels because generation exceeds demand." Despite the large countrys interconnection capacity with Argentina and Brazil, there is no active cross-border market, the energy is traded via ad hoc short-term agreements. "Even with interconnection capacity exceeding peak demand, the power system experiences high VRE curtailment, mostly at night when wind generation exceeds demand."

The curtailment of the wind power production imposed by the system operators has a strong influence on the forecast error. To build a model that, driven by the available forecast, allows to include the true power production with a prescribed degree of uncertainty, it is necessary to remove the data segments affected by wind curtailment.

Once we removed all the 24-hour segments showing wind curtailment, we set up a dataset containing 147 daily segments. In the absence of the curtailment intervention, the forecast error histograms are shown below in Figure 4, where it can be appreciated skewness reduction, except for low power forecast errors histogram.

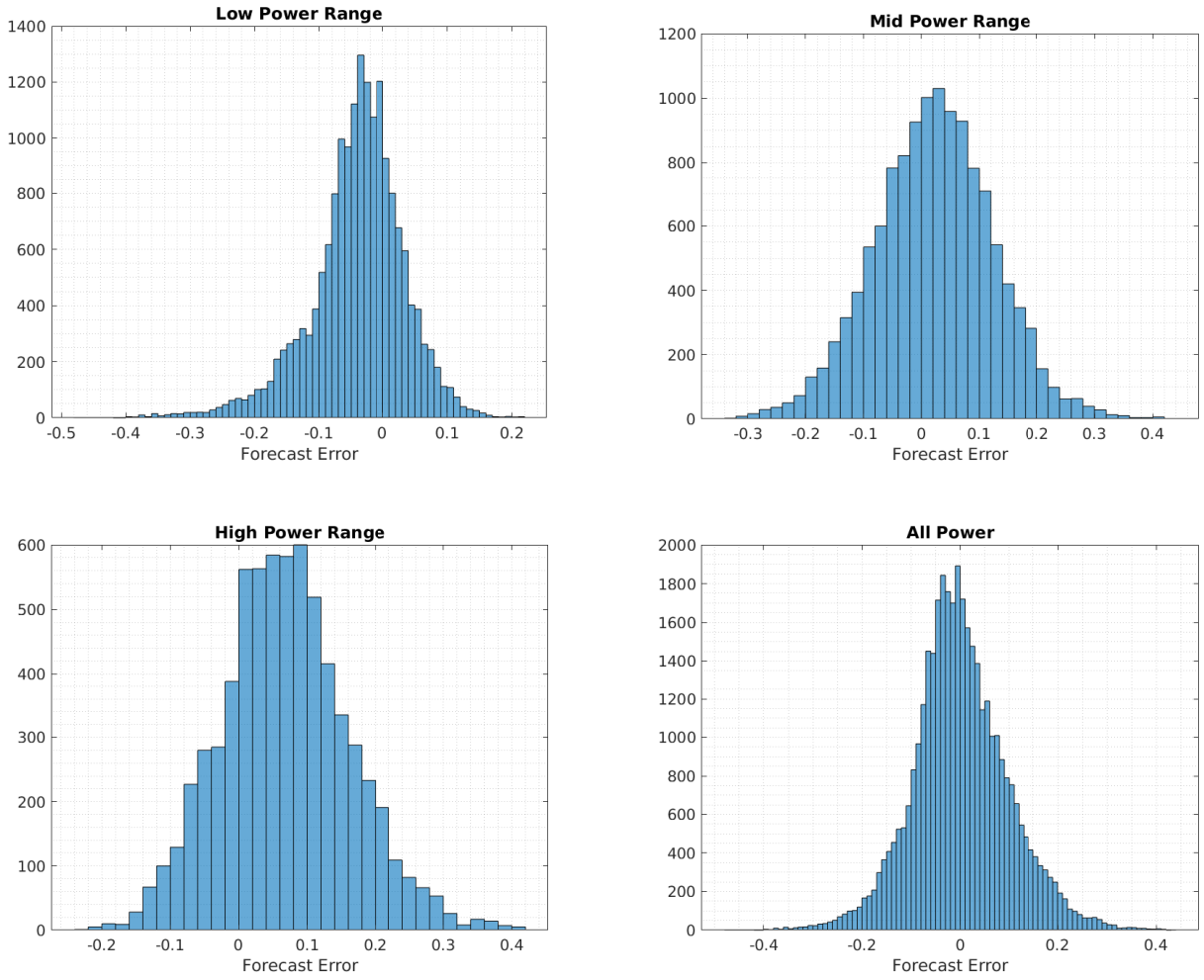

Figure 4: Wind production forecast error histograms during the period April-December 2019 after removing 24-hour segments with artificial wind curtailment: low-power (upper-left graph), mid-power (upper-right graph), high-power (lower-left graph), and the global range of power (lower-right graph).

In this stage of data preprocessing, we obtain another useful result by applying the first-order difference operator to the forecast errors. The forecast error transition histograms, displayed in the next Figure 5, will constitute later a reference for the visual assessment of the global fit of the proposed models.

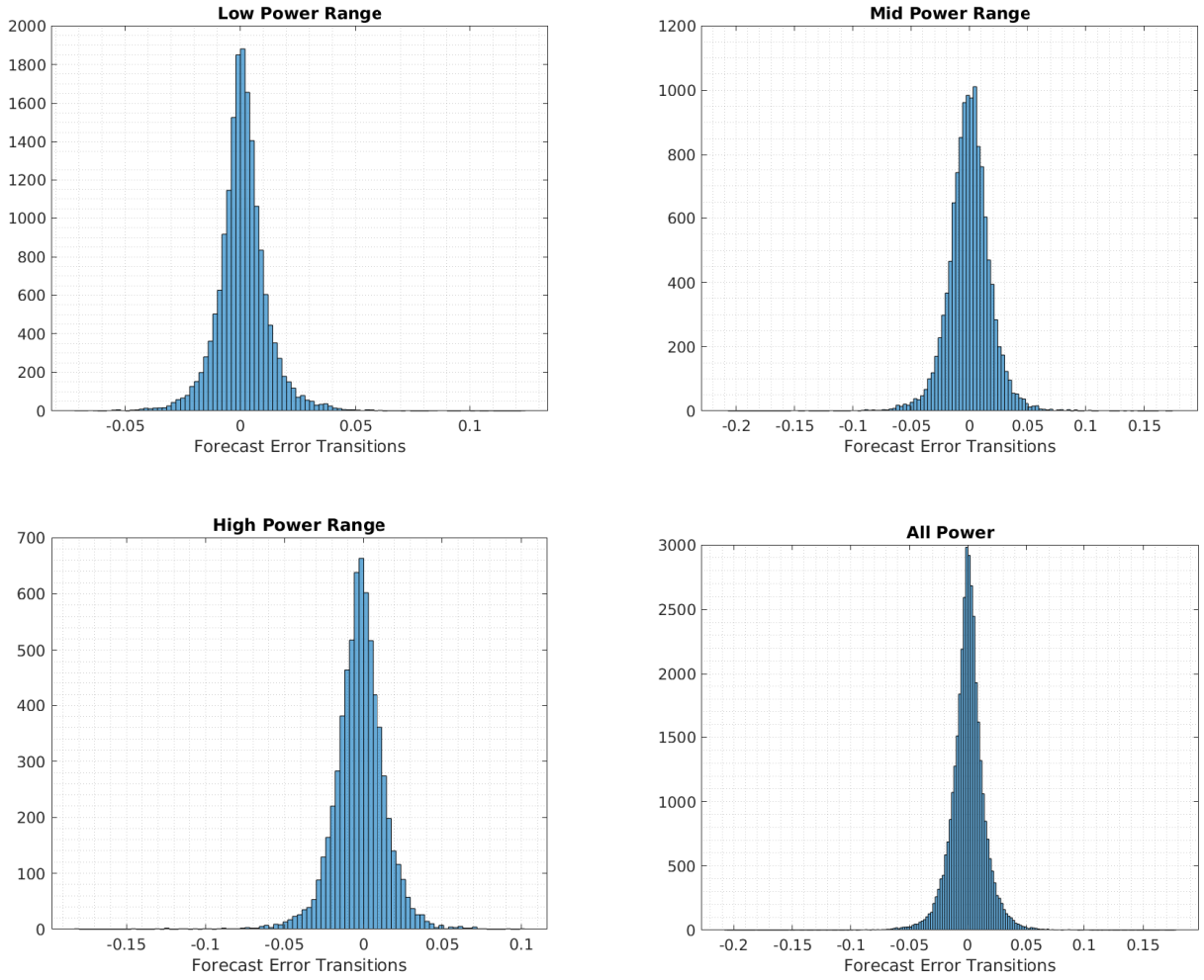

Figure 5: Forecast error transition histograms during the period April-December 2019 without wind power production curtailment: low-power (upper-left graph), mid-power (upper-right graph), high-power (lower-left graph), and the global range of power (lower-right graph).

The histograms in Figure 5 feature a non-Gaussianity trait and provide initial input for the model-building stage, which is described in Section next. Later, in Section 4, guided from inferring the unknown model parameters, we will also propose transforming data as a strategy that naturally leads to an alternative model.

3 Phenomenological Model

After analyzing the available dataset, we are now in the condition to build a type of phenomenological model for the normalized wind power generation forecasts that, in its most general form, is a stochastic process $X = \{X_t,\, t \in [0,T]\}$ defined by the following stochastic differential equation (SDE):

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

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.

3.1 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 $g(X_t, t) = X_t e^{\int_0^t \theta_s \,\mathrm{d}s}$, leads to

$$\mathrm{d}\Big(X_t e^{\int_0^t \theta_s \,\mathrm{d}s}\Big) = e^{\int_0^t \theta_s \,\mathrm{d}s}(\dot{p}_t + \theta_t p_t)\,\mathrm{d}t + e^{\int_0^t \theta_s \,\mathrm{d}s} b(X_t; p_t, \dot{p}_t, \boldsymbol{\theta})\,\mathrm{d}W_t,$$

whose integral form is

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

Taking expectation on (3) we obtain

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

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 $\mathbb{E}[X_t] = p_t - e^{-\theta_0 t}\int_0^t \dot{p}_s e^{\theta_0 s}\,\mathrm{d}s$. The next Figure (6) illustrates how different behave the estimated confidence bands for two diffusion models with and without derivative tracking, fitting the same daily segment.

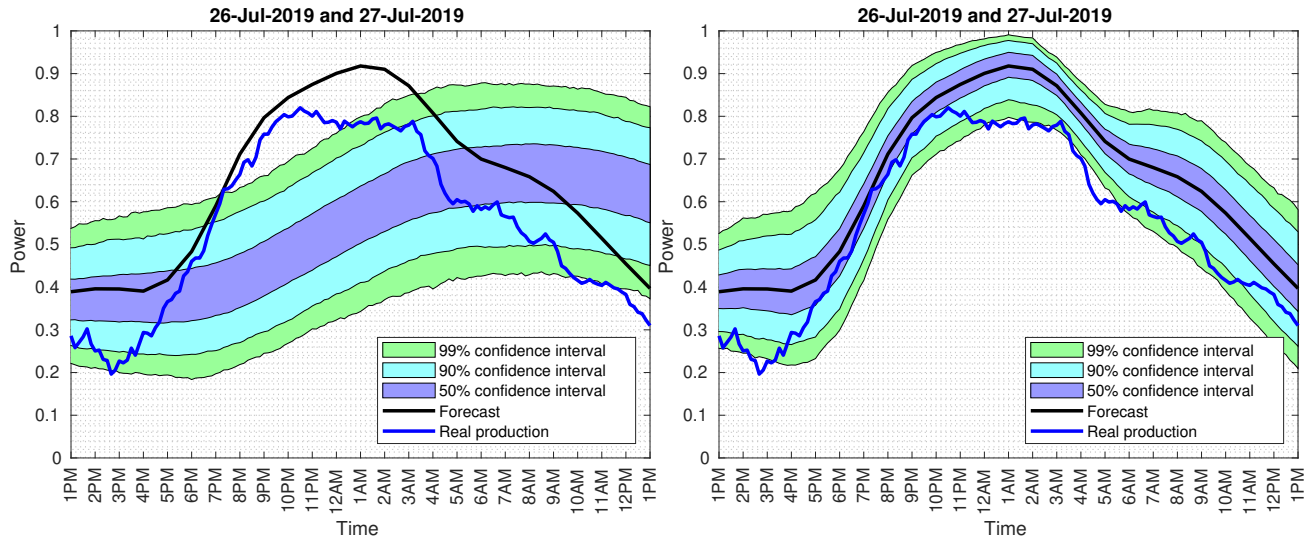

Figure 6: Pointwise confidence bands fitted, for the same daily segment, through diffusion models without derivative tracking (graph on the left) and with derivative tracking (graph on the right).

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

$$\mathrm{d}X_t = -\theta(X_t - \mu)\,\mathrm{d}t + \sqrt{2\theta\left(aX_t^2 + bX_t + c\right)}\,\mathrm{d}W_t \tag{6}$$

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

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

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

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

The proof of this theoretical statement is presented in Section 8.

Remark: Condition (B) 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

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

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 $t \in [0,T]$.

Remark: In this work, the analysis of the three different forecast datasets shows that there exists, for any forecast provider, a small $\epsilon > 0$ to define the truncated prediction function fulfilling the above condition.

From now on, we will keep the notation p_t to denote the truncated prediction function (8), unless specified otherwise.

3.2 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:

$$\begin{cases} \mathrm{d}V_t &= -\theta_t V_t\,\mathrm{d}t + \sqrt{2\alpha\theta_0(V_t+p_t)(1-V_t-p_t)}\,\mathrm{d}W_t, \quad t \in [0,T] \\ V_0 &= v_0 \in [-1+p_0, 1-p_0]. \end{cases} \tag{9}$$

4 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. Under the conditions that permit to use Itô's formula on a well-chosen transformation of the process V, John Lamperti (**lamp**) first showed that the transformed process is again a diffusion that is solution to a SDE with unit coefficient for the diffusion term. A vast literature refers nowadays to this result as the so-called Lamperti transform (see, for example, **iacus1**, **moma**, **pani**, **saso**), which is a basic tool to obtain a SDE for the transformed process whose diffusion term does not depend anymore on the state variable. A remarkable effect of removing the state dependency from the random noise term is to increase the numerical stability of the simulated paths of the transformed process. For this reason, some estimation methods of the unknown parameters of non-linear SDE models incorporated the use on the Lamperti's change of variable as part of a more complex approximation procedure (for example, the local linearization method in **shoz**, or the expansion method in **ait**, later extended to multivariate diffusion in **ait2**).

We consider the following Lamperti transformation

$$Z_t = h(V_t, t; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\alpha\theta_0}} \int \frac{1}{\sqrt{(v+p_t)(1-v-p_t)}}\,\mathrm{d}v \bigg|_{v=V_t} = -\sqrt{\frac{2}{\alpha\theta_0}}\arcsin(\sqrt{1-V_t-p_t}) \tag{10}$$

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

$$\mathrm{d}Z_t = \left[\frac{\dot{p}_t}{\sqrt{2\alpha\theta_0(V_t+p_t)(1-V_t-p_t)}} + \frac{-\theta_t V_t}{\sqrt{2\alpha\theta_0(V_t+p_t)(1-V_t-p_t)}} - \frac{1}{4}\frac{\sqrt{2\alpha\theta_0}\,\big(1-2(V_t+p_t)\big)}{\sqrt{(V_t+p_t)(1-V_t-p_t)}}\right]\mathrm{d}t + \mathrm{d}W_t\,. \tag{11}$$

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

$$\begin{aligned} \mathrm{d}Z_t &= \left[\frac{\dot{p}_t - \theta_t\left(1 - p_t - \sin^2\left(-\sqrt{\frac{\alpha\theta_0}{2}}Z_t\right)\right)}{\sqrt{2\alpha\theta_0}\cos\left(-\sqrt{\frac{\alpha\theta_0}{2}}Z_t\right)\sin\left(-\sqrt{\frac{\alpha\theta_0}{2}}Z_t\right)} - \frac{1}{4}\frac{\sqrt{2\alpha\theta_0}\left(1 - 2\cos^2\left(-\sqrt{\frac{\alpha\theta_0}{2}}Z_t\right)\right)}{\cos\left(-\sqrt{\frac{\alpha\theta_0}{2}}Z_t\right)\sin\left(-\sqrt{\frac{\alpha\theta_0}{2}}Z_t\right)}\right]\mathrm{d}t + \mathrm{d}W_t \\ &= \left[\frac{2\dot{p}_t - \theta_t(1-2p_t) + (\alpha\theta_0 - \theta_t)\cos(-\sqrt{2\alpha\theta_0}Z_t)}{\sqrt{2\alpha\theta_0}\sin\left(-\sqrt{2\alpha\theta_0}Z_t\right)}\right]\mathrm{d}t + \mathrm{d}W_t. \end{aligned} \tag{12}$$

We can see in Figure (7) how the Lamperti transformation modifies the forecast error transition histogram without curtailment displayed in Figure (5).

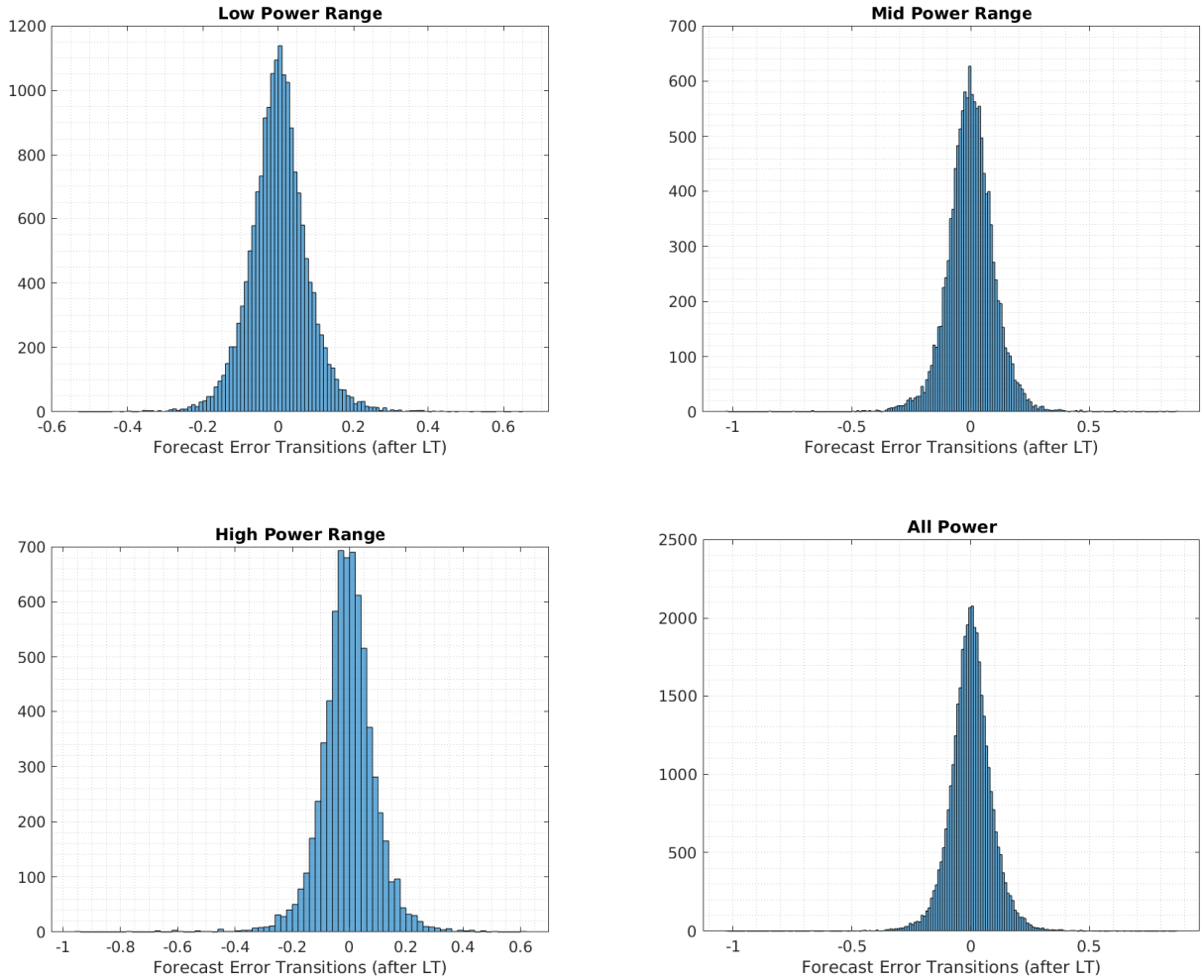

Figure 7: Lamperti transformed forecast error transition histograms during the period April-December 2019 without wind power production curtailment: low-power (upper-left graph), mid-power (upper-right graph), high-power (lower-left graph), and the global range of power (lower-right graph).

The shape of the forecast error transition histograms after the Lamperti transformation has much resemblance with a Gaussian distribution, motivating toward the use of Gaussian-like approximations of the unknown density transition functions of the process Z.

5 Likelihood functions of the forecast errors data and optimization algorithm

5.1 Likelihood in the $V-$space

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 $V^{M,N+1} = \left\{V_{t_1^{N+1}}, V_{t_2^{N+1}}, \ldots, V_{t_M^{N+1}}\right\}$ denote this random sample, with $V_{t_j^{N+1}} = \left\{V_{t_j+i\Delta}\,, i = 0, \ldots, N\right\}$, $j = 1, \ldots, M$.

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 $\boldsymbol{\theta} = (\theta_0, \alpha)$ 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: In the last subsection of this Section, we will extend the statistical model (13) by adding the transition that occurs during the time interval, say of length δ, between the epoch when the forecast is done and the first epoch (1 pm) of each day-ahead forecast. To this purpose, the likelihood function (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 the 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**):

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

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 (see, for example, (**eglix**)). 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**), consider Preston and Wood, then Egorov, and then Sorensen, remove the 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**), refer to time-homogeneous.

5.2 Approximate likelihood in the $V-$space

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 $m_1(t) \equiv \mathbb{E}[V_t] = e^{-\int_{t_{j,i-1}}^{t} \theta_s \, \mathrm{d}s} \, \mathbb{E}\left[V_{t_{j,i-1}}\right]$, 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

$$\begin{aligned} \mathrm{d}(V_t^m) = &\left(-m\theta_t V_t^m + \frac{1}{2} m(m-1) V_t^{m-2} 2\alpha\theta_0 (V_t + p_t)(1 - V_t - p_t) \right) \mathrm{d}t \\ &+ m V_t^{m-1} 2\alpha\theta_0 (V_t + p_t)(1 - V_t - p_t) \, \mathrm{d}W_t \,, \end{aligned}$$

from which we derive

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

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

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

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

5.2.1 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

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

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).

5.3 Approximate likelihood in the $Z-$space

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 $\tilde{v}_Z(t) \approx \mathrm{Var} Z_t$:

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

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

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

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

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

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).

5.4 Algorithm for the approximate maximum likelihood estimations

5.4.1 Initial guess

To guarantee the good behave for our optimization algorithm, we aim to start the optimization as close as we can from the optimal parameters. We use **Least Square Minimization** and **Quadratic Variation** over the data to find an initial guess.

- **Least square minimization:** We consider the observed data $v^{M,N+1}$ with length between observations Δ, where $i \in \{0,\dots,N\}$ and $j \in \{1,\dots,M\}$. $(V_{j,i+1}|v_{j,i})$ is a random variable which conditional mean can be approximated for any $t \in [t_{j,i-1}, t_{j,i})$ by the solution of the system

$$\begin{cases} \mathrm{d}\mathbb{E}[V] & = -\theta_t \mathbb{E}[V]\,\mathrm{d}t \\ \mathbb{E}[V](t_{j,i}) & = v_{j,i}, \end{cases}$$

  evaluated in $t_{j,i+1}$, i.e., $\mathbb{E}[V](t_{j,i+1})$. Then, the random variable $(V_{j,i+1} - \mathbb{E}[V](t_{j,i+1}))$ has zero mean. If we assume that $\theta_t = c \in \mathbb{R}^+$ for all $t \in [t_{j,i}, t_{j,i+1}]$, then $\mathbb{E}\left[V(t_{j,i+1})\right] = v_{j,i}e^{-c\Delta}$. If we have a total of $M \times N$ transitions, we can write the regression problem for the conditional mean with L^2 loss function as

$$\begin{aligned} c^* &\approx \arg\min_{c\,\geq\,0} \left[\sum_{j=1}^{M}\sum_{i=0}^{N-1}\left(v_{j,i+1} - \mathbb{E}[V](t_{j,i+1})\right)^2\right] \\ &= \arg\min_{c\,\geq\,0} \left[\sum_{j=1}^{M}\sum_{i=0}^{N-1}\left(v_{j,i+1} - v_{j,i}e^{-c\Delta}\right)^2\right] \\ &\approx \arg\min_{c\,\geq\,0} \left[\sum_{j=1}^{M}\sum_{i=0}^{N-1}\left(v_{j,i+1} - v_{j,i}(1-c\Delta)\right)^2\right]. \end{aligned} \tag{21}$$

  As Equation (21) is convex in c (composition of convex is convex), it is enough to verify the first order optimality conditions. From

$$\frac{\partial}{\partial c}\left[\sum_{j=1}^{M}\sum_{i=0}^{N-1}\left(v_{j,i+1} - v_{j,i}(1-c\Delta)\right)^2\right] = \sum_{j=1}^{M}\sum_{i=0}^{N-1} 2v_{j,i}\Delta\left(v_{j,i+1} - v_{j,i}(1-c\Delta)\right),$$

  it follows that

$$c^* = \frac{\sum_{j=1}^{M}\sum_{i=0}^{N-1} v_{j,i}(v_{j,i} - v_{j,i+1})}{\Delta \cdot \sum_{j=1}^{M}\sum_{i=0}^{N-1}(v_{j,i})^2}. \tag{22}$$

- **Quadratic variation:** We use the numerical Euler-Maruyama (E-M) scheme to approximate the solution to the SDE (9). In particular, we approximate the quadratic variation of the Itô's process V

$$[V]_t = \int_0^t b(V_s;\boldsymbol{\theta},p_s)^2\,\mathrm{d}s, \text{ where } \; b(V_s;\boldsymbol{\theta},p_s) = \sqrt{2\alpha\theta_0(V_s+p_s)(1-V_s-p_s)},$$

with the discrete sum $\sum_{0<t_{j,i-1}\leq t}(V_{t_{j,i}} - V_{t_{j,i-1}})^2$.

As initial guess for the diffusion variability coefficient $\theta_0\alpha$, we choose

$$\theta_0^*\alpha^* \approx \frac{\sum_{j=1}^{M}\sum_{i=1}^{N}(v_{j,i} - v_{j,i-1})^2}{2\Delta \cdot \sum_{j=1}^{M}\sum_{i=1}^{N}(v_{j,i} + p_{j,i})(1 - v_{j,i} - p_{j,i})}, \tag{23}$$

where Δ is the length of the time interval between two consecutive measurements.

5.4.2 Negative log-likelihood minimization in the $V-$space

To find the optimal parameters, we minimize the negative log-likelihood (negative version of (18)) using the derivative-free function *fminsearch* from MATLAB R2019b over the parameters (θ_0, α). At each step of the iteration, we:

- Use the training dataset to find the SDE's first and second moments as explained in subsection 5.2.
- Match the proxy distribution moments with the SDE's moments.
- Evaluate the negative log-likelihood using the training dataset.

5.4.3 Negative log-likelihood minimization in the $Z-$space

Let $v^{M,N+1}$ be the observed data, and $h(v_{j,i};\boldsymbol{\theta})$ the Lamperti transform of the observation $v_{j,i}$. The transformed observations $z^{M,N+1}$ depend on the parameter $\boldsymbol{\theta}$.
The problem of maximizing the approximated Lamperti log-likelihood (20), i.e.,

$$\max_{\boldsymbol{\theta}} \tilde{\ell}_Z\left(\boldsymbol{\theta}; z^{M,N+1}\right),$$

is not totally defined as the data $z^{M,N+1}$ depend on $\boldsymbol{\theta}$. However, we propose to find a fixed point $\boldsymbol{\theta}^*$ such that

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \tilde{\ell}_Z\left(\boldsymbol{\theta}; \{h(v_{j,i};\boldsymbol{\theta}^*)\}_{j=1,i=1}^{M,N}\right). \tag{24}$$

In this point, the likelihood has a maximum for the data set corresponding to that point. We are interested in finding solutions for (24). If $\boldsymbol{\theta}^*$ solves (24), then

$$\boldsymbol{\theta}^* - \arg\max_{\boldsymbol{\theta}} \tilde{\ell}_Z\left(\boldsymbol{\theta}; \{h(v_{j,i};\boldsymbol{\theta}^*)\}_{j=1,i=1}^{M,N}\right) = 0.$$

Given $\boldsymbol{\theta}^* = (\theta_0^*, \alpha^*)$, we call $\boldsymbol{\theta}^{**} = (\theta_0^{**}, \alpha^{**})$ to the solution of

$$\arg\max_{\boldsymbol{\theta}} \tilde{\ell}_Z\left(\boldsymbol{\theta}; \{h(v_{j,i};\boldsymbol{\theta}^*)\}_{j=1,i=1}^{M,N}\right).$$

For each $\boldsymbol{\theta}^*$, we define the relative error function $\Psi : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ as

$$\Psi(\boldsymbol{\theta}^*) = \frac{|\theta_0^* - \theta_0^{**}|}{|\theta_0^*|} + \frac{|\alpha^* - \alpha^{**}|}{|\alpha^*|}, \tag{25}$$

and remark that $\boldsymbol{\theta}^*$ is a fixed point if and only if $\Psi(\boldsymbol{\theta}^*) = 0$. To find minimizers for Ψ, we proceed in the same way as described in subsection (5.4.2), but using Gaussian surrogate densities.

5.5 Model specification with the additional parameter δ

We observe that for most days, the forecast error at time $t_{j,0} = 0$ is not zero. According to the forecasts procedure, we may assume that there is a time in the past $t_{j,-\delta} < t_{j,0}$, such that the forecast error $V_{j,-\delta} = 0$.

For any $j = 1, \ldots, M$, we extrapolate backward linearly the truncated prediction function to get its value at time $-\delta$,$p_{j,-\delta}$, and set $v_{t_{j,-\delta}} = 0$. We assume that the initial transition $(V_{j,0}|v_{j,-\delta}; \boldsymbol{\theta}, \delta)$ has a Beta distribution and apply to it the same moment matching method used above. Given a starting value for $\boldsymbol{\theta}$, we obtain the initial estimation for δ solving the following problem

$$\delta \approx \arg\max_{\delta} \mathcal{L}_{\delta}\left(\boldsymbol{\theta}, \delta; v^{M,1}\right) = \arg\max_{\delta} \prod_{j=1}^{M} \rho_0\left(v_{j,0}|v_{j,-\delta}; \boldsymbol{\theta}, \delta\right), \tag{26}$$

where $\mathcal{L}_{\delta}$ is the $\delta-$likelihood. To solve this problem, we repeat the steps described in subsection (5.4.2), with the additional initial step of creating the linear extrapolation for $p_{j,-\delta}$.

As remarked, we extend the statistical model (13) to include the extra parameter δ. The complete likelihood, $\mathcal{L}_c$, is given by

$$\mathcal{L}_c\left(\boldsymbol{\theta}, \delta; v^{M,N+1}\right) = \mathcal{L}\left(\boldsymbol{\theta}; v^{M,N+1}\right) \mathcal{L}_{\delta}\left(\boldsymbol{\theta}, \delta; v^{M,1}\right). \tag{27}$$

As we may provide initial guesses for $\boldsymbol{\theta}$ and δ, we have a starting point for the numerical optimization of the complete likelihood (27).

6 Application to the April-December 2019 Uruguay wind and forecast dataset

Our statistical analysis starts with partitioning the 147 segments of normalized wind power production, each 24-hour long. We select 73 non-contiguous segments for the models' calibration procedure, assigning them to the training set. The other 74 non-contiguous segments compose the test set. Such an allocation mechanism guarantees independence among the segments, matching the assumption we did in Section 5 to formulate the statistical models.

Show that 24h is enough to ensure independence? All the following results involving a single provider refer to Provider A. Also, all calibrations involve the training sets and all simulations, the test sets. Following the instruction for the initial guesses from section 5.4, we obtain the initial guess $(\theta_0, \alpha, \delta) \approx (1.54, 0.072, 073)$.

6.1 Calibration of the approximate negative log-likelihood in the $V-$space

As an auxiliary verification, we plot the negative log-likelihood (negative version of (18)) as a function of the parameters, and we use additional minimization functions from MATLAB R2019b.

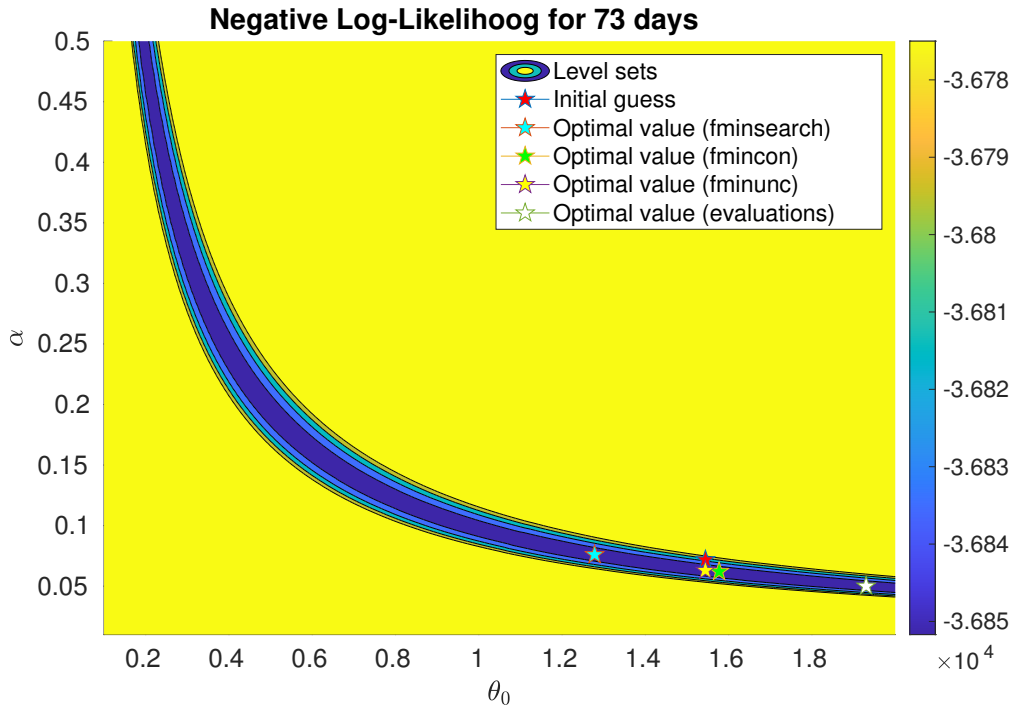

Figure 8: Negative log-likelihood's level sets. All optimal values are located over the curve $\theta_0\alpha = 0.097$.

On Figure (8), we can see the level sets for the negative log-likelihood. The numerical values of each important point can be seen in Table (1). We set $\boldsymbol{\theta}^V = (\theta_0^V, \alpha^V) = (1.93, 0.050)$, as it is where the negative log-likelihood reaches its minimum value.

	θ_0	α	$\theta_0\alpha$
Initial guess	1.54	0.070	0.111
fminsearch	1.14	0.076	0.097
fmincon	1.58	0.062	0.097
fminunc	1.54	0.063	0.097
evaluations	1.93	0.050	0.097

Table 1: Value points from Figure 8.

Observe that all the local (possibly global) minimizers are located in curve $\theta_0\alpha = 0.097$. We will explain this effect later, but it is a consequence of Markovianity and the indistinguishable observed transitions.

6.2 Calibration of the approximate negative log-likelihood in the $Z-$space

We obtain $(\theta_0^Z, \alpha^Z) = (1.87, 0.043)$ as a minimizer for the relative error function (25).

To verify and compare this two vector of parameters, (i.e., (θ_0^V, α^V) and (θ_0^Z, α^Z)), we simulate error paths in the $V-$space. We simulate five error paths for each day in the test set and construct histograms with the transitions. The histograms can be seen in Figure (9). We observe a slightly better approximation using (θ_0^Z, α^Z).

6.3 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:

$$\begin{cases} \mathrm{d}X_t &= -\theta_0(X_t - p_t)\,\mathrm{d}t + \sqrt{2\alpha\theta_0 X_t(1-X_t)}\,\mathrm{d}W_t, \quad t \in [0,T] \\ X_0 &= x_0 \in (0,1), \end{cases}. \tag{28}$$

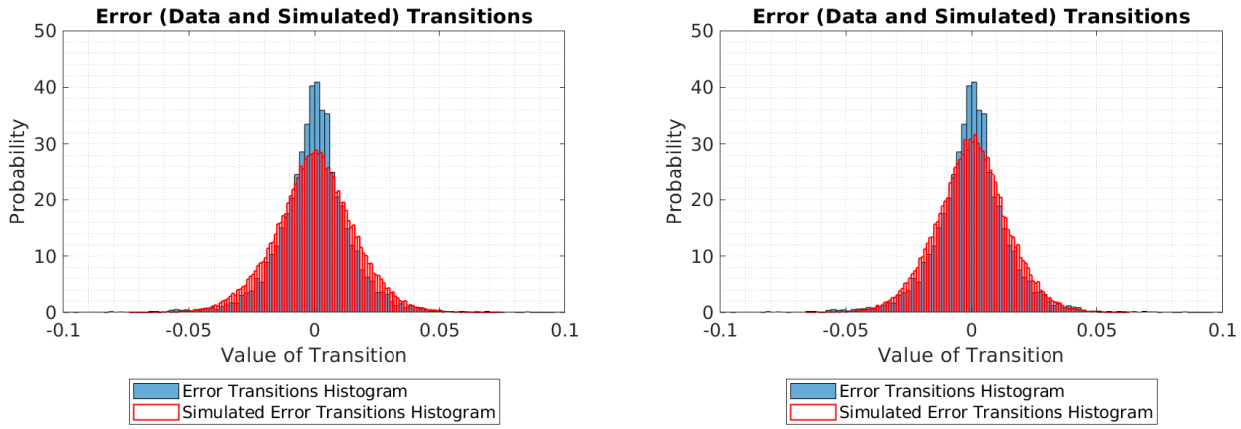

Figure 9: Probability histograms for the error transitions. Using the Provider A, we overlap the real transitions from the test set with the simulated ones from the $V-$space SDE. On the left, simulations using (θ_0^V, α^V). On the right, simulations using (θ_0^Z, α^Z).

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

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

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

with $\theta_0 > 0$, $\alpha > 0$ and θ_t satisfying the condition (B) .

To show the better performance of Model 2, we calculate the Akaike information criterion (AIC) and Bayesian information criterion (BIC) for each provider, each model, and different proxies densities. Table (2) summarizes this computations. We can observe that the best fitting is achieved with Model 2 and Beta proxies.

Model	Forecast Provider	Method	Product $\theta_0\alpha$	AIC	BIC
Model 1	Provider A	Gaussian Proxy	0.105	-58226	-58211
		Shoji-Ozaki	0.104	-58226	-58211
		Beta Proxy	0.104	-58286	-58271
	Provider B	Gaussian Proxy	0.105	-58226	-58211
		Shoji-Ozaki	0.104	-58226	-58211
		Beta Proxy	0.104	-58288	-58273
	Provider C	Gaussian Proxy	0.105	-58226	-58211
		Shoji-Ozaki	0.104	-58226	-58211
		Beta Proxy	0.104	-58286	-58271
Model 2	Provider A	Beta Proxy	0.097	-73700	-73685
	Provider B	Beta Proxy	0.098	-73502	-73487
	Provider C	Beta Proxy	0.108	-72518	-72503

Table 2: We compare the different models based on information criterion.

6.4 Forecast providers comparison

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

Forecast Provider	Parameters (θ_0, α)	Product $\theta_0\alpha$
Provider A	$(1.930, 0.050)$	0.097
Provider B	$(1.420, 0.069)$	0.098
Provider C	$(1.380, 0.078)$	0.108

Table 3: Optimal parameters for different providers using Model 2 with Beta proxies.

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

6.5 Calibration of Model 2 with additional parameter δ

After calibrating Model 2 on the training set using the complete likelihood (27), we can generate simulations of the wind power production for the time horizon of interest. Figure (10) shows five simulated paths of the wind power production for each day of interest.

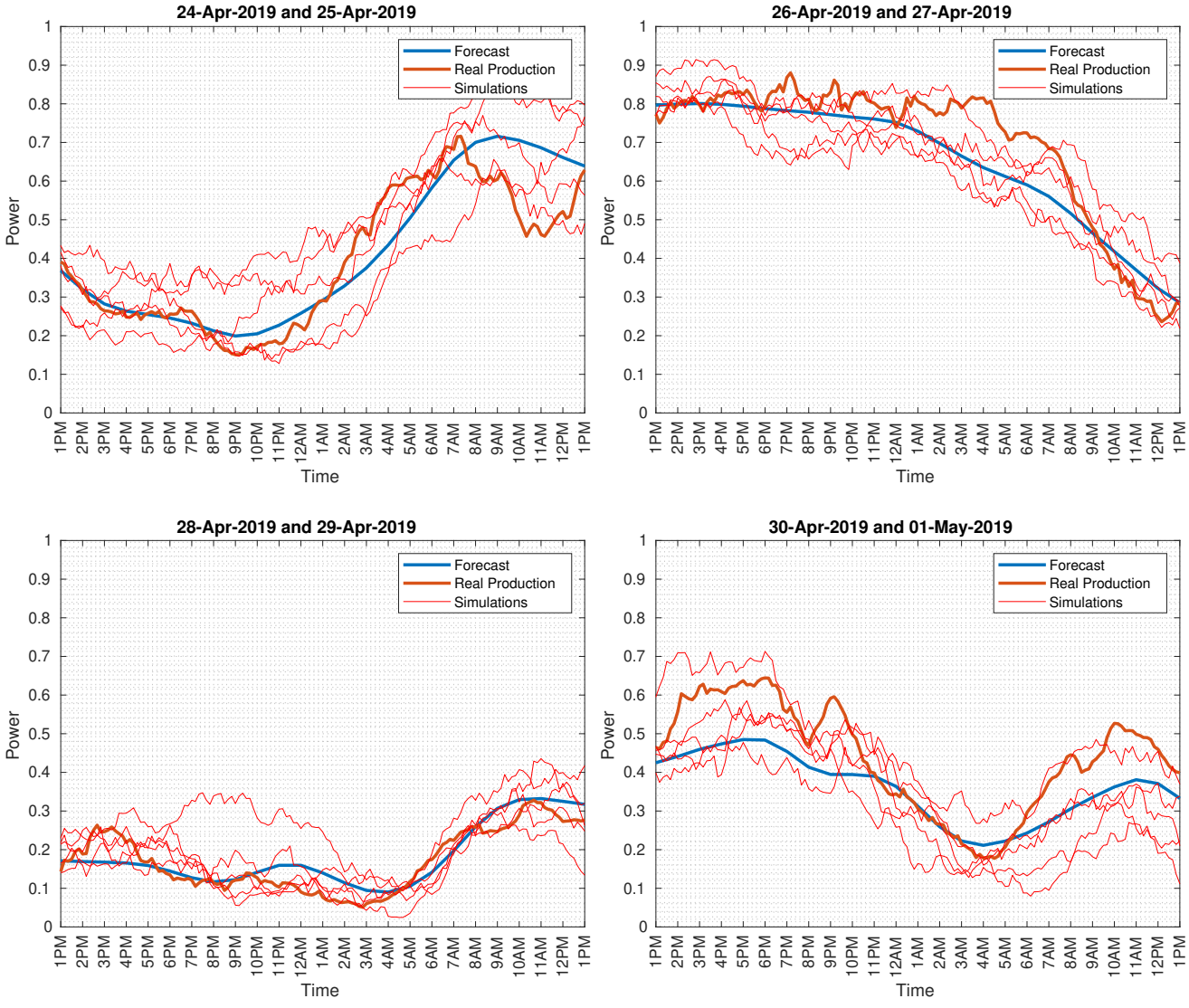

Figure 10: Four arbitrary days with five simulated wind power production paths each.

On the basis of the approximate MLEs for Model 2, we can obtain empirical pointwise confidence bands for the wind power production. Figure 11 shows the empirical pointwise confidence bands for the wind power production for each day of interest, assuming Model 2 specification and a given forecaster.

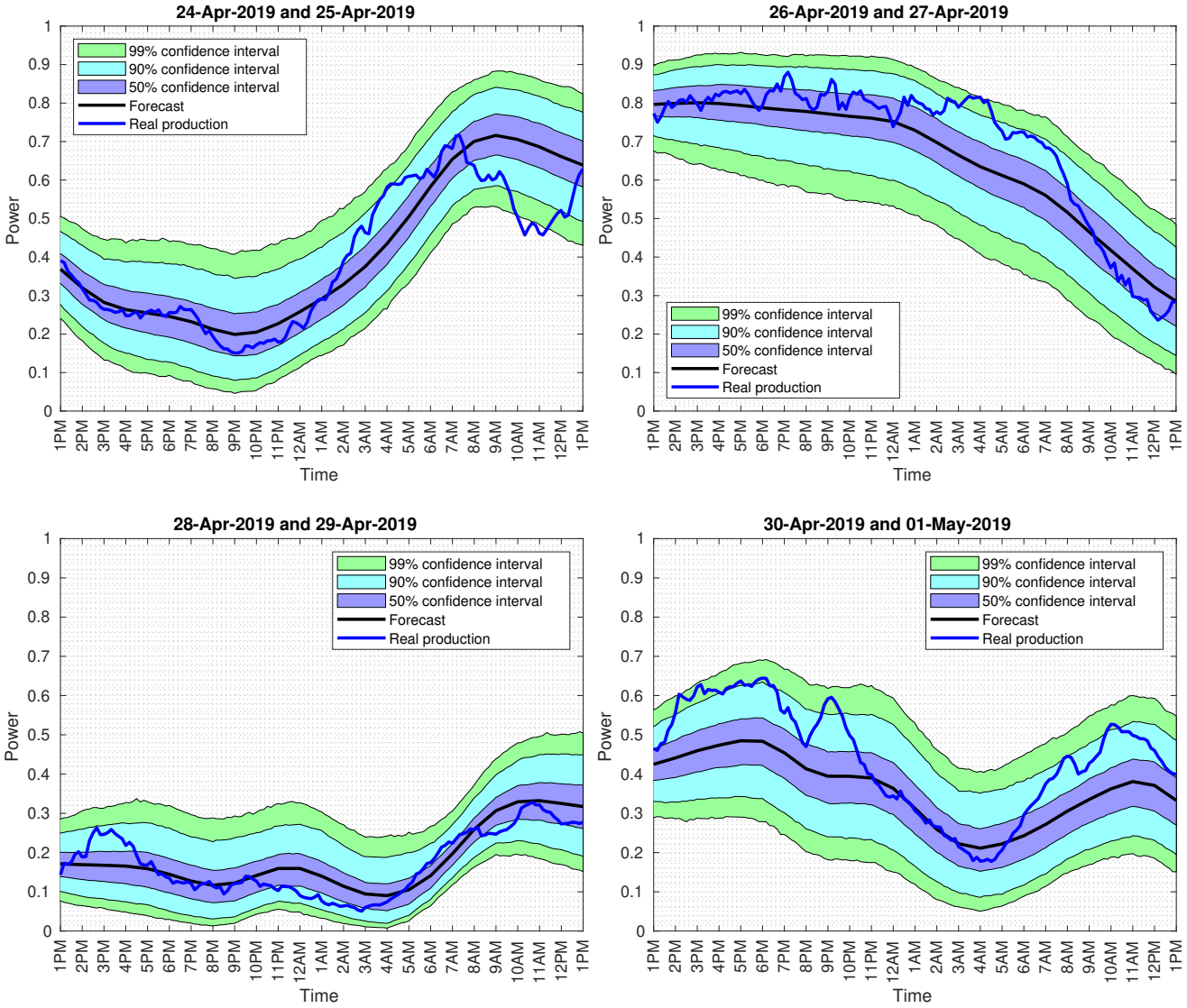

Figure 11: Empirical pointwise confidence bands for the wind power production using the approximate MLEs for Model 2 $(\theta_0, \alpha, \delta) = (1.93, 0.050, 0.090)$. Blue line: real production.

6.5.1 Value of δ

As a final verification, we study the behavior of δ as function of $\boldsymbol{\theta}$. Given a vector of parameters, we calculate an initial guess for δ solving problem (26). Even when it is a guess, it helps us to understand qualitatively the meaning of this additional parameter.

We choose as domain the most significant values of θ_0 and $\theta_0\alpha$, regarding the previous numerical results. In Figure (12) we can observe that:

- The initial time δ decreases as $\theta_0\alpha$ increases. This is a consequence of the increment in the diffusion as $\theta_0\alpha$ increases. As there is more diffusion, it is needed less time for the initial transition density to cover the initial error observations.
- The initial time δ increases as θ_0 increases. As we increment θ_0, also the mean reversion becomes larger and reduces the variance for the initial transition density. Then, it is needed more time for the initial transition density to cover the initial error observations.

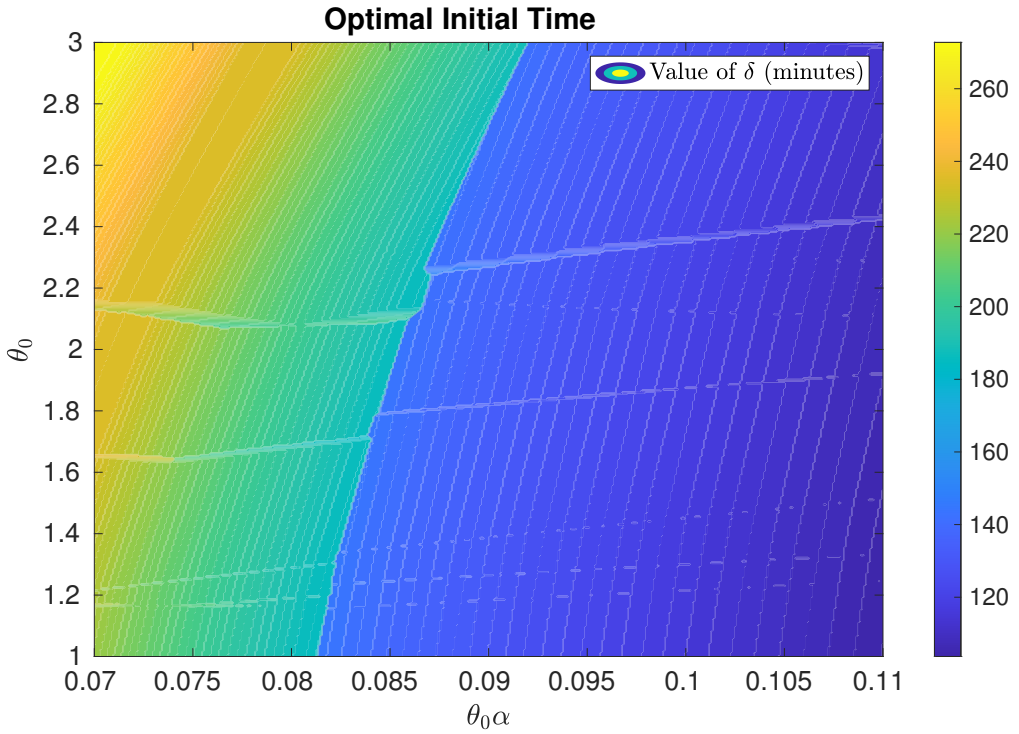

Figure 12: Initial value for δ as a function of the elements of the parameter vector $\boldsymbol{\theta}$.

7 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 three 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.

8 Appendix

8.1 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

$$\begin{cases} \mathrm{d}X_t &= \big(\dot{p}_t - \theta_t(X_t - p_t)\big)\,\mathrm{d}t + \sqrt{2\alpha\theta_0 X_t(1-X_t)}\,\mathrm{d}W_t, \quad t \in (0,T) \\ X_0 &= x_0 \in [0,1], \end{cases}. \tag{30}$$

where $(p_t)_{t\in[0,T]}$ denotes the prediction function that satisfies $0 \le p_t \le 1$ 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 (30).

Theorem 8.1. Assume that

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

Then, there is a unique strong solution to (30) s.t. for all $t \in [0,T]$, $X_t \in [0,1]$ a.s.

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

$$X_t = x_0 + \int_0^t \big(\dot{p}_s - \theta_s(X_s - p_s)\big)\,\mathrm{d}s + \int_0^t \sqrt{2\alpha\theta_0|X_s(1-X_s)|}\,\mathrm{d}W_s, \quad 0 < x_0 < 1. \tag{31}$$

According to Proposition 2.13, p.291 of **KarShr**, under assumption (A) there is a unique strong solution X to (31). Moreover, as the diffusion coefficient is of linear growth, we have for all $p > 0$

$$\mathbb{E}[\sup_{t\in[0,T]} |X_t|^p] < \infty. \tag{32}$$

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:

$$|\psi_n(x)| \underset{n\to+\infty}{\to} |x|,\ \ x{\psi'}_n(x) \underset{n\to+\infty}{\to} |x|,\ \ |\psi_n(x)| \wedge |x{\psi'}_n(x)| \le |x|$$
$${\psi'}_n(x) \le 1, \quad \text{and } {\psi''}_n(x) = g_n(|x|) \ge 0 \quad \text{with} \quad g_n(x)x \le \frac{2}{n} \quad \text{for all } x \in \mathbb{R}.$$

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

$$\begin{aligned}\psi_n(X_t) = \psi_n(x_0) &+ \int_0^t {\psi'}_n(X_s)(\dot{p}_s + \theta_s p_s - \theta_s X_s)\,\mathrm{d}s \\ &+ \int_0^t {\psi'}_n(X_s)\sqrt{2\alpha\theta_0|X_s(1-X_s)|}\,\mathrm{d}W_s + \alpha\theta_0\int_0^t g_n(|X_s|)|X_s(1-X_s)|\,\mathrm{d}s.\end{aligned}$$

Now, thanks to (A), (32) and to the above properties of ψ_n and g_n, we get

$$\mathbb{E}[\psi_n(X_t)] \le \psi_n(x_0) + \int_0^t \big(\dot{p}_s + \theta_s p_s - \theta_s \mathbb{E}[{\psi'}_n(X_s)X_s]\big)\,\mathrm{d}s + \frac{2\alpha\theta_0}{n}\int_0^t \mathbb{E}|1 - X_s|\,\mathrm{d}s.$$

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

$$\mathbb{E}[|X_t|] \le x_0 + \int_0^t \big(\dot{p}_s + \theta_s p_s - \theta_s \mathbb{E}|X_s|\big)\,\mathrm{d}s.$$

Besides, taking the expectation of (31), we get

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

and thus we have

$$\mathbb{E}[|X_t| - X_t] \le \int_0^t \theta_s \mathbb{E}[X_s - |X_s|]\,\mathrm{d}s.$$

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

$$\mathrm{d}Y_t = \big(\theta_t(1-p_t) - \dot{p}_t - \theta_t Y_t\big)\,\mathrm{d}t - \sqrt{2\alpha\theta_t Y_t(1-Y_t)}\,\mathrm{d}W_t,$$

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

Theorem 8.2. Assume that assumptions of Theorem 8.1 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

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

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

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

$$\frac{\mathrm{d}X_t}{X_t} = \left(\frac{\dot{p}_t + \theta_t p_t}{X_t} - \theta_t\right)\mathrm{d}t + \sqrt{\frac{2\alpha\theta_0(1-X_t)}{X_t}}\,\mathrm{d}W_t$$

so that

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

where $M_t = \int_0^t \sqrt{\frac{2\alpha\theta_0(1-X_s)}{X_s}}\,\mathrm{d}W_s$ 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

$$X_t \geq x_0 \exp\Big(\alpha\theta_0 t - \int_0^t \theta_s\,\mathrm{d}s + M_t\Big).$$

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

$$\mathrm{d}Y_t = \big(\theta_t(1-p_t) - \dot{p}_t - \theta_t Y_t\big)\,\mathrm{d}t - \sqrt{2\alpha\theta_0 Y_t(1-Y_t)}\,\mathrm{d}W_t$$

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

Remark: As the diffusion coefficient of X given by $x \mapsto \sqrt{2\alpha\theta_0 x(1-x)}$ is strictly positive for all $x \in]0,1[$, the condition (B) ensures that the transformation between Z and X is bijective, so that we deduce the properties of existence and uniqueness of Z from those of X. The application of Itô's formula in Section 4 is subjected to the condition (B) that avoids the process X hits the boundaries of the interval $]0,1[$, otherwise the Lamperti transform is not applicable.