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

February 21, 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

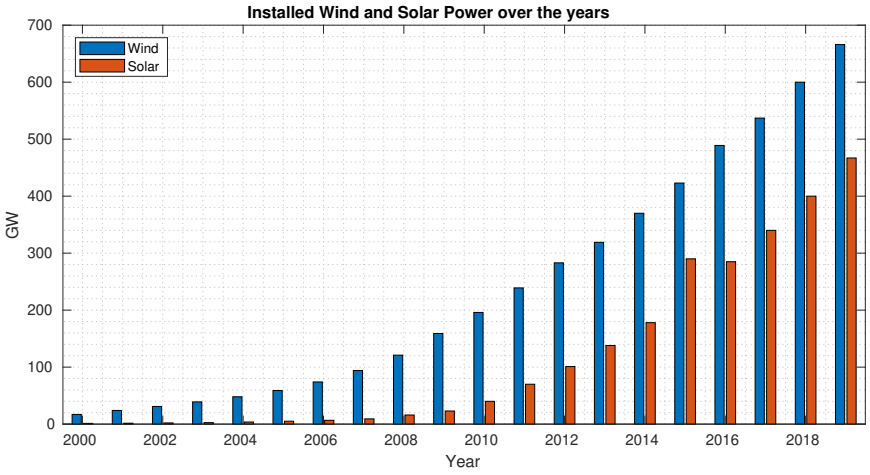

Figure 1: Installed wind and solar power over the years **sultana2017review**
We recall the importance of accurate forecasts to use green energies optimally.

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). The data is normalized with respect to the maximuim 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 a low, meduim 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.

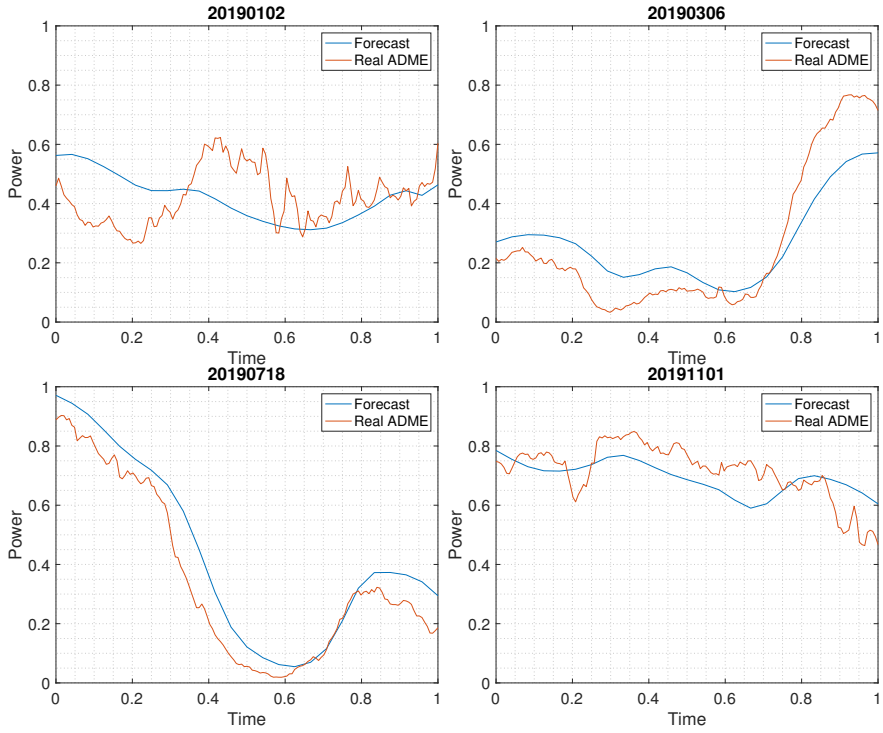

Figure 2: Four samples from the Uruguayan data of 2019. Each sample compromises of two 24-hour paths. In blue is an hourly wind power production forecast. In orange is the actual wind power production recorded in 10 minute intervals.

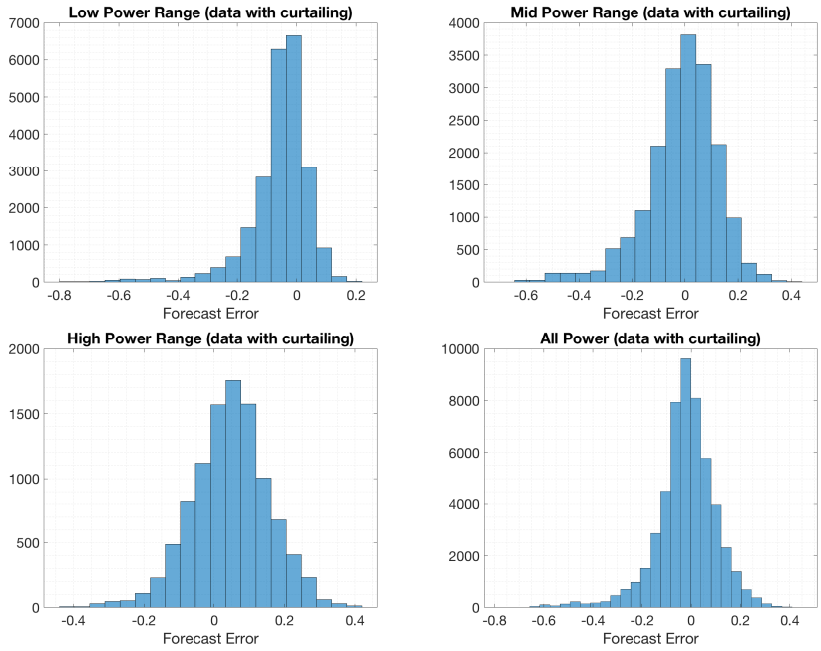

Figure 3: We see that forecast errors exhibit skewness, which is exaggerated as a consequence of curtailing in the production. Low power is when produced power is in $[0, 0.3]$, mid-power is when it is in $(0.3, 0.6]$, and high power when it is in $(0.6, 1]$. We have a total of 363 days to use.

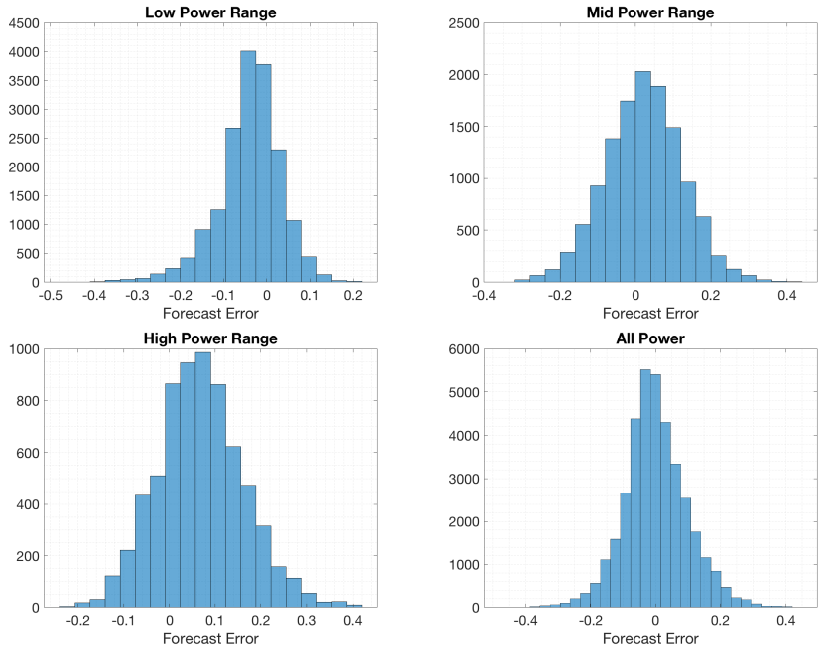

Figure 4: We observe that skewness has been reduced after removing the days with curtailing. We have a total of 255 days with no curtailing.

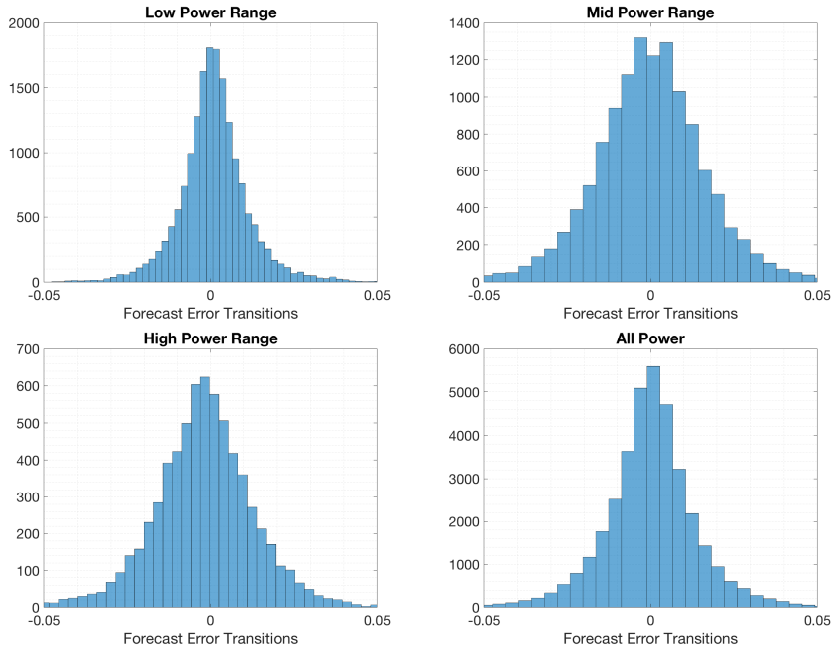

Figure 5: Transitions in error between forecast and real production. We can observe that they are not totally centered around zero.

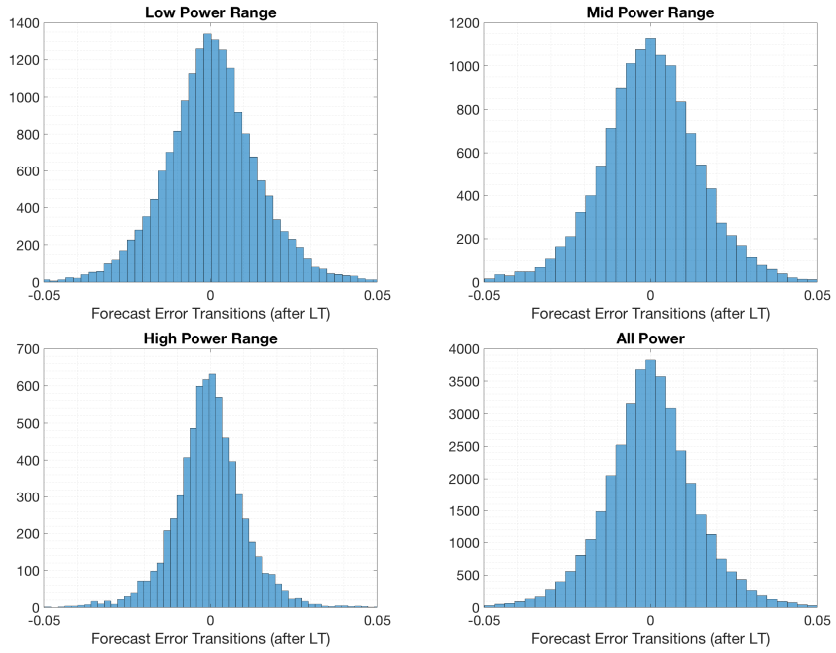

Figure 6: Transitions after the Lamperti transform. The transitions now present a normal Gaussian shape. This effect motivates the use of Gaussian proxies to approximate the process after using the Lamperti transform.

2 Phenomenological Model

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

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

where

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

Figure 7: insert figure showing the difference with and without derivative tracking

- $\{W_t,\, t > 0\}$ is a standard Wiener process in $\mathbb{R}$.

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

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

2.1 Physical Constrains

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

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

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

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

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

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

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

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

where $\theta_t > 0$ is a time-dependent parameter that controls the speed of reversion. Derivative tracking does not assure that the paths of the process X take values in $[0, 1]$.

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

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

where $\alpha > 0$ is a constant parameter that controls the path variability. This diffusion term belongs to the Pearson diffusion family and , in particular, it is a Jacobi type diffusion. Add expressions and references for these class of diffusions, for example **iacus1**

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

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

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

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

The expression for θ_t is important: θ_t depens on t through $\dot{p}_t$ and p_t. Besides it is parsimonious because only one unkonwn parameter appear in 6. Recall it is an upped bound for the real inequality.

Change of Variables:

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

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

The model becomes,

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

3 State independent diffusion: Lamperti transform

Our model (7) for the forecast error has a diffusion term that depends on the state variable V_t. To estimate the unknown model parameters, a recommended technique is to modify the SDE (7) by applying the so-called

Lamperti transform (see **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.

We consider the following Lamperti transformation (Remind that in continuous-time models with different diffusion term are singular and the likelihood cannot be written down. Besides, it can be unstable to work with low frequency. If the frequency of the data is too low, the Lamperti transform helps to detect this problem (in principle, it is not clear how to determine the right frequency for estimation.)

$$Z_t = h(V_t) = \frac{1}{\sqrt{2\theta_t \alpha}} \int_{\xi}^{x} \frac{1}{\sqrt{(u+p_t)(1-u-p_t)}}\, du \Bigg|_{x=V_t}, \tag{8}$$

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

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

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

$$dZ_t = \left[\frac{-\theta_t V_t}{\sqrt{2\theta_t \alpha (V_t + p_t)(1 - V_t - p_t)}} - \frac{1}{4} \frac{\sqrt{2\theta_t \alpha}\,(1 - 2(V_t + p_t))}{\sqrt{(V_t + p_t)(1 - V_t - p_t)}} \right] dt + dW_t\,. \tag{10}$$

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

$$\begin{aligned} dZ_t &= \left[\frac{-\theta_t(\sin(\sqrt{2\theta_t \alpha} Z_t) + 1 - 2p_t)}{\sqrt{2\theta_t \alpha}\cos(\sqrt{2\theta_t \alpha} Z_t)} + \frac{\sqrt{2\theta_t \alpha}\sin(\sqrt{2\theta_t \alpha} Z_t)}{2\cos(\sqrt{2\theta_t \alpha} Z_t)} \right] dt + dW_t \\ &= \left[\frac{\sqrt{2\theta_t}\left(\sin(\sqrt{2\theta_t \alpha} Z_t)(\alpha - 1) + (2p_t - 1)\right)}{2\sqrt{\alpha}\cos(\sqrt{2\theta_t \alpha} Z_t)} \right] dt + dW_t\,. \end{aligned} \tag{11}$$

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

4 Likelihood in V space

4.1 Likelihood

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

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

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

$$\mathcal{L}(\boldsymbol{\theta}, \delta; V^{M,N}) = \prod_{j=1}^{M} \prod_{i=1}^{N} \rho(V_{j,t_i}|V_{j,t_{i-1}}; p_{[t_{i-1},t_i]}, \boldsymbol{\theta})\, \rho_0(V_{j,t_0}|V_{j,t_{-\delta}}; \boldsymbol{\theta}, \delta)\,. \quad (12)$$

In 12 modify the expression for ρ_0 contemplating the lag, say δ, between the time when the forecast is done and the first time of forecasting. ρ_0 can be written as a conditional density between these two times. Once θ_0 and α have been estimated, one could estimate δ.

The exact computation of the likelihood (12) relies on the availability of a closed-form expression for the transition densities of V that, on the basis of the Markovian property of V, are characterized, for $t_{j,i-1} < t < t_{j,i}$, as solutions of the Fokker-Planck-Kolmogorov equation (**iacus1 saso**):

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

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

Closed-form solutions to initial-boundary value problem for SDEs can be obtained only in a few cases. Besides, in our case solving numerically (13) 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. 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

A common choice is performing a Gaussian approximation of the transition densities, but this is inappropriate here due to physical constraints which give rise to asymmetric forecasting errors as seen in figure (**??**).

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

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

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

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

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

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

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

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

Moment Matching

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

The shape parameters are given by

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

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

4.3 Optimization

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.

$$\theta_0 \approx \arg\min_{\theta} \sum_{j}^{M} \sum_{i}^{N} \Big(v_{i+1,j} - v_{i,j} - \left(-\theta_t v_{i,j}\right)\left(t_{i+1,j} - t_{i,j}\right)\Big)^2\,, \tag{18}$$

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

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

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

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

$$\delta \approx \arg\min_{\delta} \mathcal{L}_\delta(\boldsymbol{\theta}, \delta; V^{M,1}) = \arg\min_{\delta} \prod_{j=1}^{M} \rho_0(V_{j,t_0} | V_{j,t_{-\delta}}; \boldsymbol{\theta}, \delta)\,. \tag{20}$$

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.

$$\begin{aligned} dX_t &= -\theta_0(p_t - X_t)\ dt + \sqrt{2\theta\alpha X_t(1-X_t)}\ dW_t \\ X_0 &= x_0\,, \end{aligned} \tag{21}$$

  with $\theta_t = \theta_0$.

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

$$\begin{aligned} dV_t &= -\theta_t V_t\ dt + \sqrt{2\theta_t\alpha(V_t+p_t)(1-V_t-p_t)}\ dW_t \\ V_0 &= v_0\,, \end{aligned} \tag{22}$$

  with θ_t given by (6).

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 $(\theta_0, \alpha, \delta) \approx (1.2, 0.1, 0.6)$.

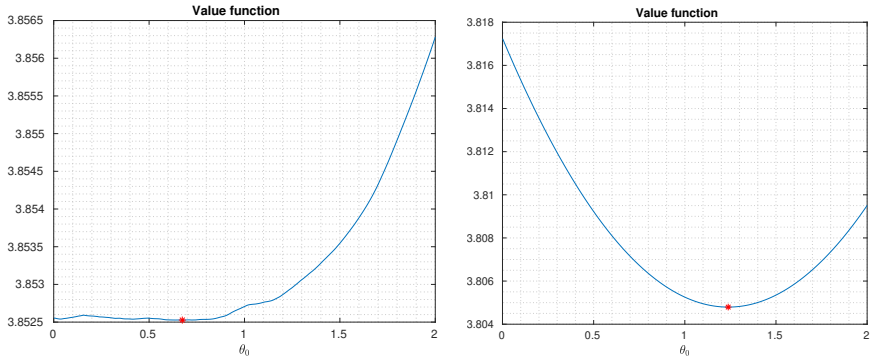

Figure 8: Value function from (18) as a function of different θ_0 values. We can observe that the function if flat for a wide range.

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.

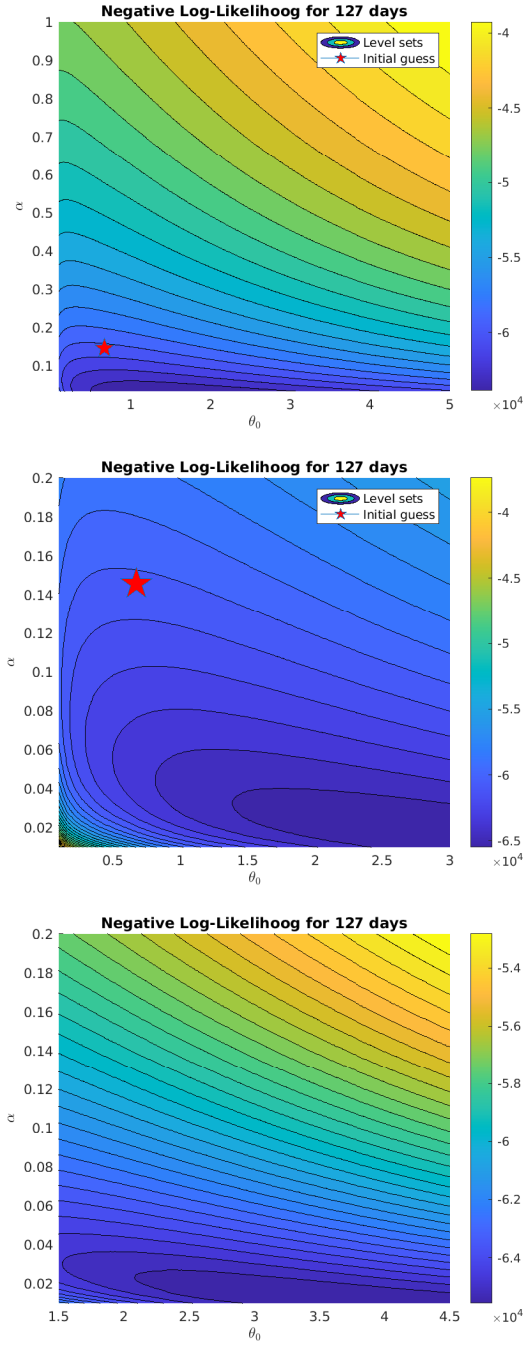

Figure 9: Negative Log-Likelihood for different domains.

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.

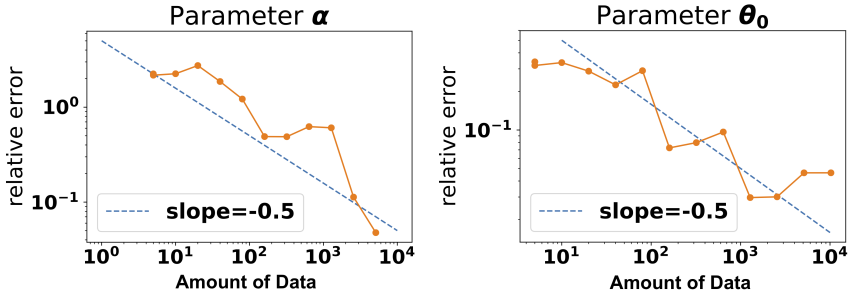

Figure 10: We show self-convergence of our algorithm applied to model 2. We conlcude that the rate matches that of of Monte Carlo. Data is from Uruguay pertaining to year 2019.

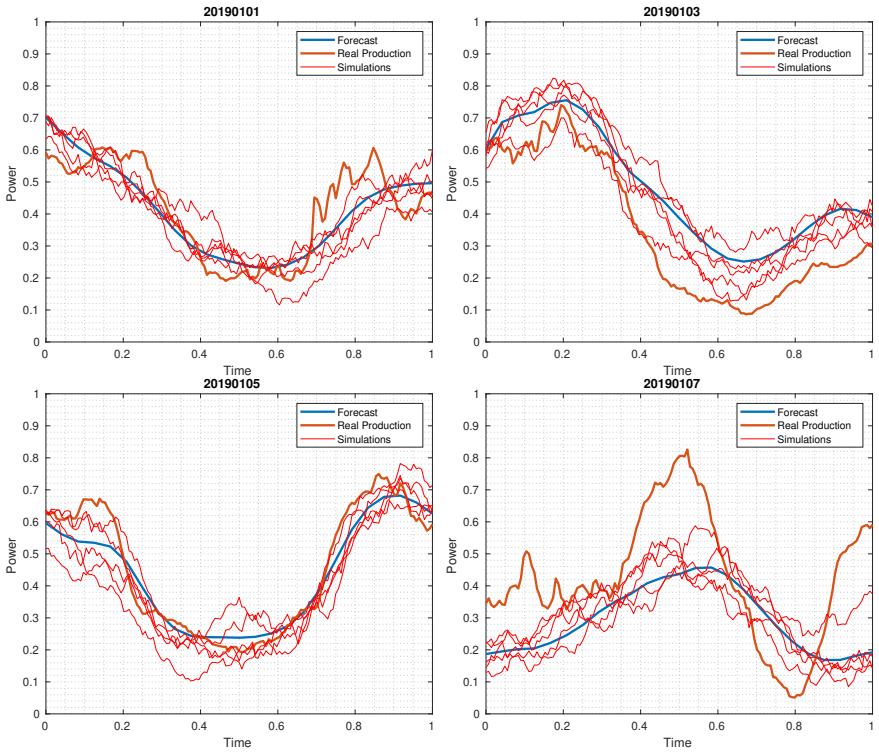

Figure 11: We simulate following model 2 five possible future wind power production paths using the obtained optimal parameters $(\theta_0, \alpha, \delta) = (1.2, 0.1, 0.6)$. Forecast is from Uruguay pertaining to year 2019.

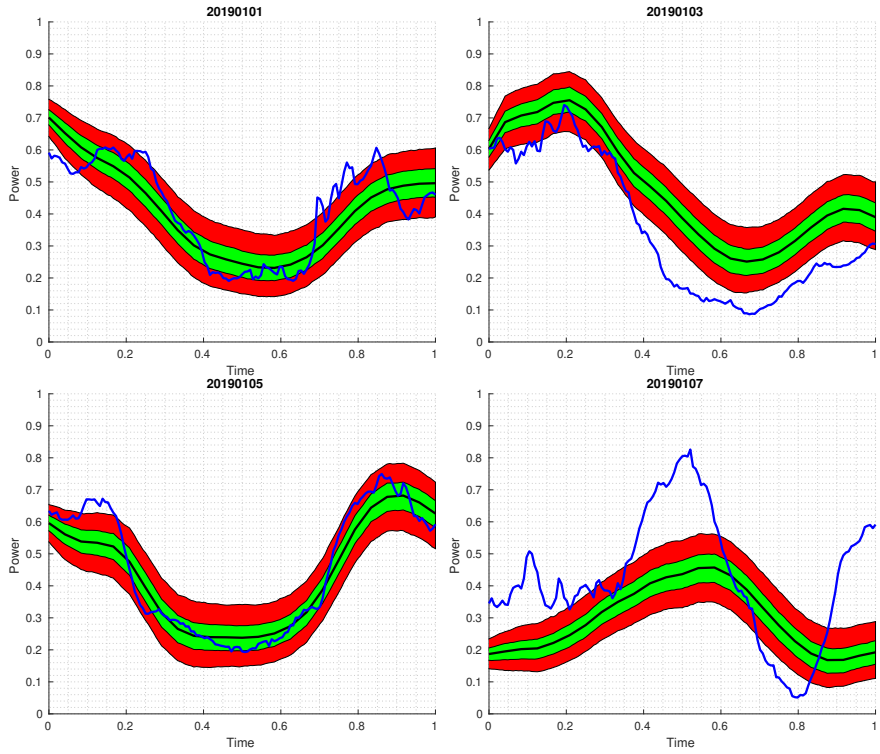

Figure 12: Pending add legend and choose better colors.We obtain confidence intervals following model 2 for future wind power production using the obtained optimal parameters $(\theta_0, \alpha, \delta) = (1.2, 0.1, 0.6)$. Actual production plotted in retrospect. Forecast and data is from Uruguay pertaining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

$$\begin{aligned} dX_t &= \big(\dot{p}_t - \theta_t(X_t - p_t)\big)dt + \sqrt{2\alpha\theta_t X_t(1-X_t)}dW_t \quad t \in [0,T] \\ X_0 &= x_0 \in [0,1], \end{aligned} \tag{23}$$

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 Alfonsi **Alf** and Karatzas and Shreve **KarShr** that we adapted to the setting of our model (23).

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 (23) s.t. for all $t \in [0,T]$, $X_t \in [0,1]$ a.s.

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

$$X_t = x_0 + \int_0^t \big(\dot{p}_s - \theta_s(X_s - p_s)\big)ds + \int_0^t \sqrt{2\alpha\theta_s|X_s(1-X_s)|}dW_s, \quad x_0 > 0. \tag{24}$$

According to Proposition 2.13, p. 291 of **KarShr** under assumption (A) there is a unique strong solution X to (24). Moreover, as the diffusion coefficient is of linear growth, we have for all $p > 0$

$$\mathbb{E}[\sup_{t\in[0,T]} |X_t|^p] < \infty. \tag{25}$$

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 fucntion. Applying It's formula we get

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

Now, thanks to (A), (25) 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 (24), 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 $\theta_t(1-p_t) - \dot{p}_t - \alpha\theta_t \geq 0$. □

When using Lamperti, the form