A SDE model with derivative tracking for wind power forecast error: model building, inference and application

Ahmed Kebaier

Université d'Evry, Université Paris-Saclay

Séminaire du laboratoire Finance des Marchés de l'énergie

March 18, 2022

Joint work

A Derivative Tracking Model for Wind Power Forecast Error
Stat. Comput. 31 (2021), no. 5, Paper No. 64, 19 pp.

- **Renzo Caballero** King Abdullah University of Science and Technology (KAUST), Saudi Arabia
- **Marco Scavino**, Instituto de Estadística, Universidad de la República, Montevideo, Uruguay
- **Raúl Tempone**, RWTH Aachen University Germany and King Abdullah University of Science and Technology (KAUST), Saudi Arabia.

Introduction and motivation

Wind and solar energy are expanding renewable generation capacity, experiencing record growth in the last years.

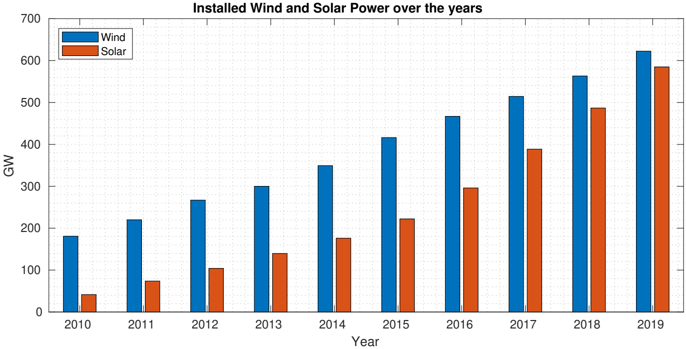


Figure 1: Worldwide installed wind and solar power 2010–2019 IRENA, 2020. We recall the importance of accurate forecasts to use green energies optimally.

Reliable wind power generation forecasting is crucial for the following applications (see, for example, Giebel et al., 2011, Chang, 2014, Zhou et al., 2013):

- Allocation of energy reserves such as water levels in dams or oil, and gas reserves.
- Operation scheduling of controllable 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.

Introduction and motivation

- In recent years, Uruguay has triggered a remarkable change in its energy matrix. In (IRENA, 2019, p.23), Uruguay was among those countries showcasing innovation, like Denmark, Ireland, Germany, Portugal, Greece and Spain, with proven feasibility of managing annual variable renewable energy (VRE) higher than 25% in power systems.
- According to (REN21, 2019, 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 **36%** in 2018.
- Including hydropower, Uruguay now produces more than **97%** of its electricity from renewable energy sources.

Introduction and motivation

- 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 with wind power short-term forecasting to orientate and assess the strategies and decisions of wind energy actors and businesses.

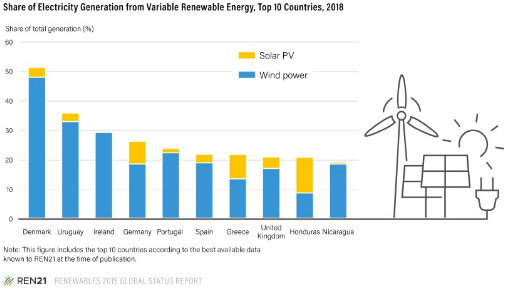


Figure 2: Renewables: Top Ten countries according to REN21 in 2018.

Data description

Wind power production data in Uruguay between April and December 2019, normalized with respect to the maximum installed wind power capacity (1474 MW). Each day, wind power production recordings are available every ten minutes. Data from three different forecast providers, available each day starting at 1 pm.

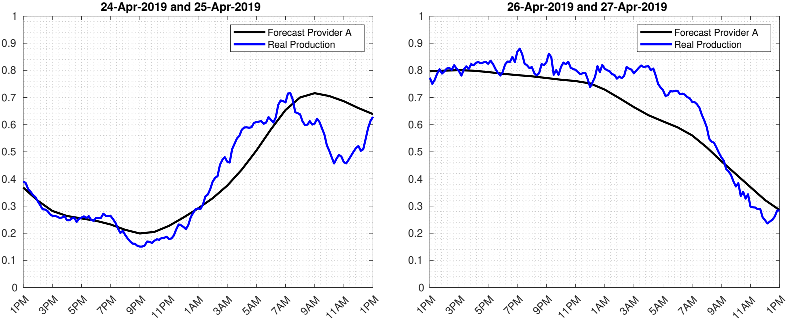
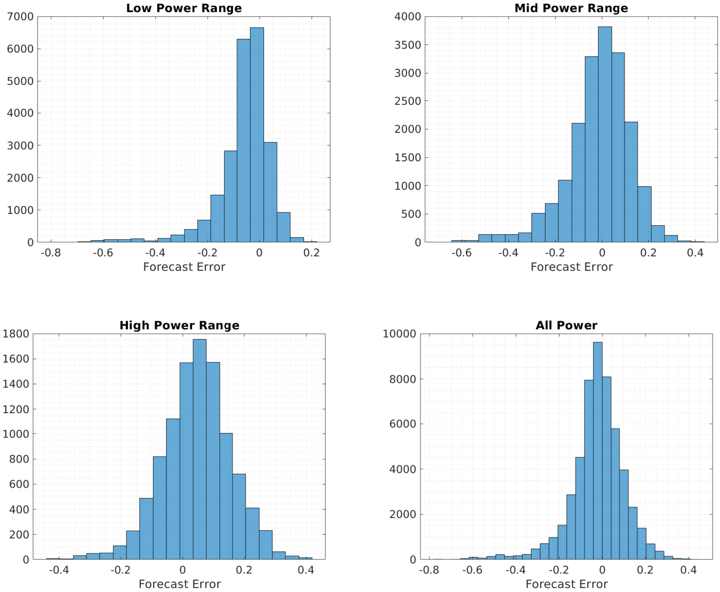


Figure 3: Two 24-hour segments with the normalized wind power real production in Uruguay (blue line) recorded every ten minutes, and the hourly wind power production forecasted by provider A (black line).

Wind production forecast error histograms



Curtailment

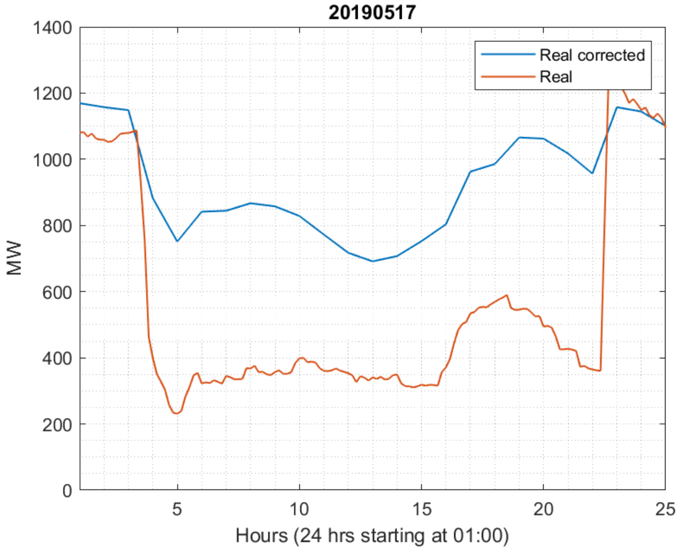
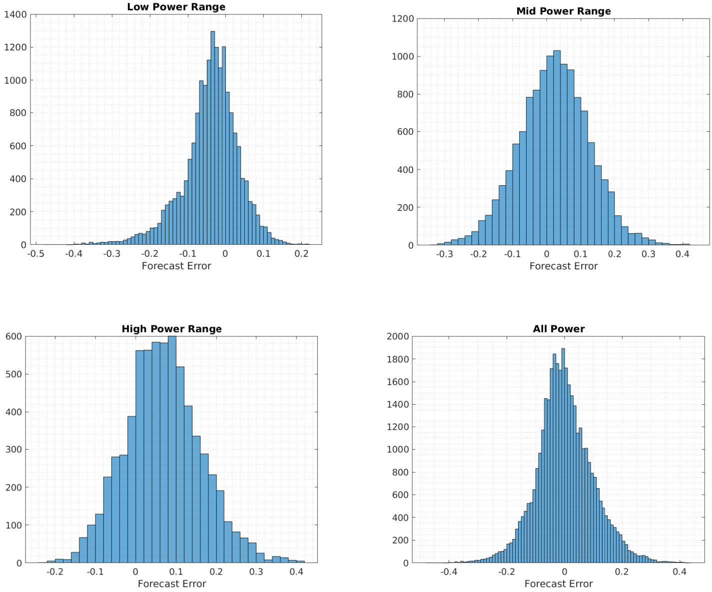


Figure 4: A real headache: Example of a day with curtailment

Forecast error, no curtailment (147 daily segments)



Forecast error *transition* histograms

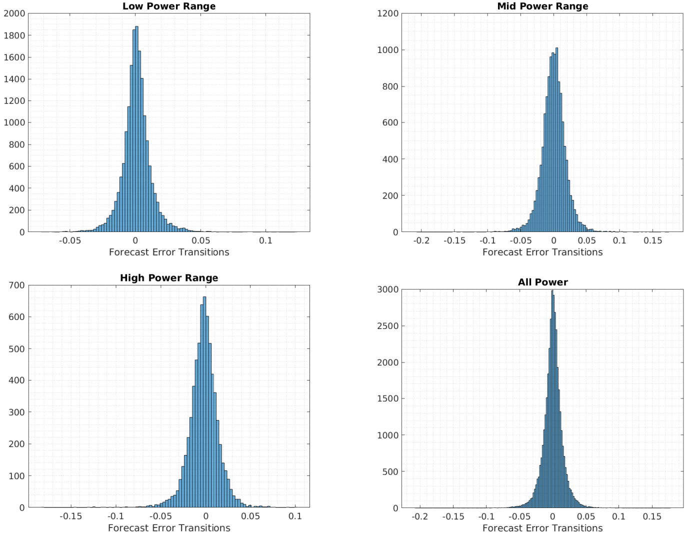


Figure 5: Forecast error transition histograms, applying the first-order difference operator to the forecast errors.

Phenomenological model

Let $X = \{X_t,\ t \in [0,\ T]\}$ be a $[0,1]$-valued stochastic process that represents the normalized wind power production, defined by the following Itô stochastic differential equation (SDE):

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

where:

- $a\left(\cdot, p_t, \dot{p}_t, \boldsymbol{\theta}\right) : [0,1] \to \mathbb{R}$ denotes a drift function,
- $b\left(\cdot; p_t, \dot{p}_t, \boldsymbol{\theta}\right) : [0,1] \to \mathbb{R}_+$ is a diffusion function,
- $\boldsymbol{\theta}$ is a vector of unknown parameters,
- $(p_t)_{t\in[0,T]}$ is the given forecast, taking values in [0,1] 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.

Specification of the drift function

Time-dependent drift function that features the mean-reverting property as well as derivative tracking:

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

where $(\theta_t)_{t\in[0,T]}$ is a positive deterministic function, whose range depends on $\boldsymbol{\theta}$, that controls the speed of reversion.

Observe: Given $\mathbb{E}\left[X_0\right] = p_0$, apply Itô's lemma on the forecast error, $V_t = X_t - p_t$, yielding

$$dV_t = dX_t - \dot{p}_t dt = -\theta_t V_t dt + b_t dW_t,$$

and taking expectations yields, for $t > 0$,

$$\frac{dE[V_t]}{dt} = -\theta_t E[V_t]$$

implying $E[V_t] = 0$ for $t > 0$. [Centering property]

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 parameter θ_t,
- it tracks the time derivative $\dot{p}_t$.

Obs: A mean-reverting model without derivative tracking shows a delayed path behavior.

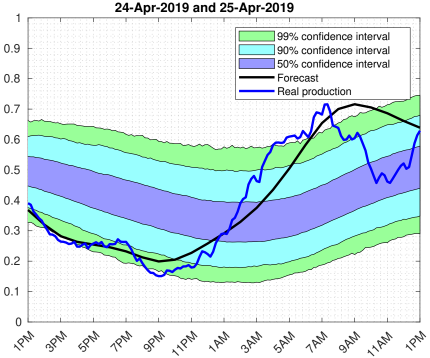
Example: Consider the diffusion model (1) with

$$a(X_t; p_t, \boldsymbol{\theta}) = -\theta_0(X_t - p_t)\,, \theta_0 > 0.$$

Then, given $\mathbb{E}[X_0] = p_0$, this diffusion has mean

$$\mathbb{E}[X_t] = p_t - e^{-\theta_0 t} \int_0^t \dot{p}_s e^{\theta_0 s} ds \neq p_t. \text{ [Not Centered]}$$

Models with and without derivative tracking



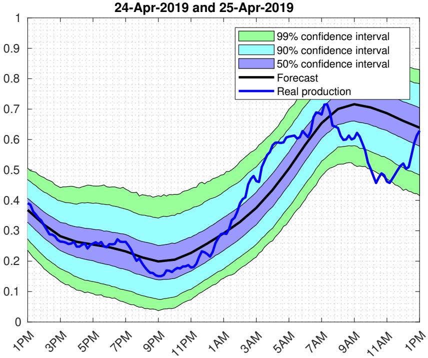


Figure 6: Pointwise confidence bands fitted, for the same daily segment, through diffusion models without derivative tracking (plot on the left) and with derivative tracking (plot on the right).

Specification of the diffusion function

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:

$$b(X_t; \boldsymbol{\theta}) = \sqrt{2\alpha\theta_0 X_t (1 - X_t)} \quad (3)$$

where $\theta_0 > 0$, $\alpha > 0$ is an unknown parameter that controls the path variability.

This diffusion term belongs to the Pearson diffusion family,
in particular, it defines a Jacobi type diffusion.
Recall (Forman and Sørensen, 2008) that a *Pearson diffusion* is a stationary solution to a stochastic differential equation of the form

$$dX_t = -\theta(X_t - \mu)dt + \sqrt{2\theta\left(aX_t^2 + bX_t + c\right)}dW_t \quad (4)$$

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, determine the state space of the diffusion as well as the shape of the invariant distribution.

Normalized wind power production model

Normalized wind power production model

$$\begin{cases} dX_t = (\dot{p}_t - \theta_t (X_t - p_t))dt + \sqrt{2\alpha\theta_0 X_t (1 - X_t)} dW_t, \quad t \in [0, T] \\ X_0 = x_0 \in [0, 1] \end{cases} \tag{5}$$

- To ensure that X_t is the unique solution to (5) $\forall t \in [0, T]$ with state space [0,1] a.s., the mean-reversion time-dependent function θ_t must satisfy the condition:

$$\theta_t \geq \max\left(\frac{\alpha\theta_0 + \dot{p}_t}{1 - p_t}, \frac{\alpha\theta_0 - \dot{p}_t}{p_t}\right). \tag{6}$$

Theorem (Existence and Uniqueness)

Assume that

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

Then, there is a unique strong solution to (5) *s.t. for all* $t \in [0,T]$, $X_t \in [0,1]$ *a.s.*

Truncated prediction function

- Issue: If we choose the equality in (6), then θ_t becomes unbounded when $p_t = 0$ or $p_t = 1$.
- Our approach: Introduce a truncation parameter, $0 < \epsilon << 1$. Consider the following truncated prediction function

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

that satisfies $p_t^\epsilon \in [\epsilon, 1 - \epsilon]$ for any $0 < \epsilon < \frac{1}{2}$ and $t \in [0, T]$, implying that θ_t is bounded for every $t \in [0, T]$.

Theorem

Take $0 < \epsilon < 1/2$ and let (6) hold. Once we truncate p into p_ϵ, the solution X to (5) does not reach the boundary of $[0, 1]$ a.s.

Forecast error of the normalized wind power production

Model for the forecast error of the normalized wind power production

The model for the forecast error of the normalized wind power production $V = \{V_t,\ t \in [0, T]\}$, $V_t = X_t - p_t$, $\forall t \in [0, T]$ is defined by the following Itô stochastic differential equation (SDE):

$$\begin{cases} dV_t = -\theta_t V_t dt + \sqrt{2\alpha\theta_0 (V_t + p_t)(1 - V_t - p_t)} dW_t, & t \in [0, T] \\ V_0 = v_0 \in [-p_0, 1 - p_0] \end{cases} \tag{7}$$

Lamperti transform

John Lamperti (Lamperti, 1964) first showed that the use of Itô's formula on a well-chosen transformation of a diffusion process is again a diffusion process solving a SDE with unit, constant diffusion coefficient.

(Nonlinear) Lamperti transform with unknown parameters:

$$
\begin{aligned}
Z_t = h(V_t, t; \boldsymbol{\theta}) &= \int \frac{dv}{\sigma(v)}\Bigg|_{v=V_t} \\
&= \frac{1}{\sqrt{2\alpha\theta_0}} \int \frac{1}{\sqrt{(v + p_t)(1 - v - p_t)}} dv \Bigg|_{v=V_t} \\
&= -\sqrt{\frac{2}{\alpha\theta_0}} \arcsin(\sqrt{1 - V_t - p_t})
\end{aligned}
\tag{8}
$$

By Itô's lemma, if $h(v,t)$ is $C^2([-p_t, 1-p_t])$ for v and $C^1([0,T])$ for t, then:

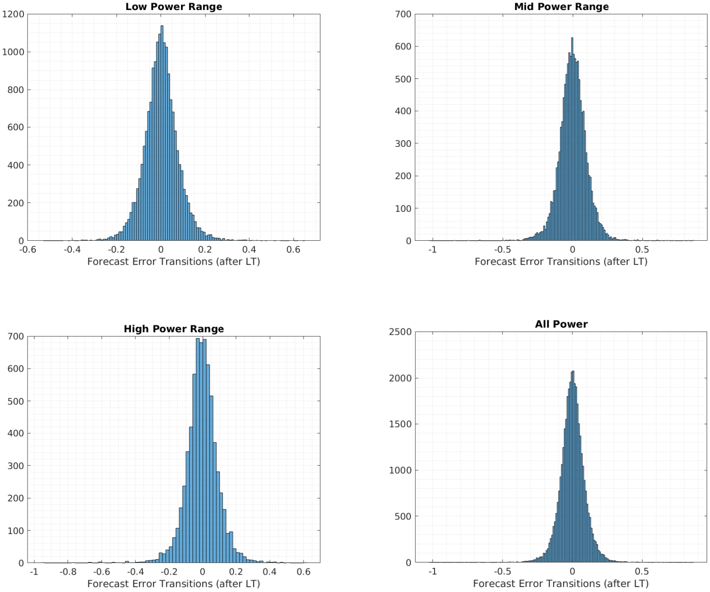
$$
dZ_t = \left(\partial_t h + \partial_v h\,(-\theta_t V_t) + \frac{1}{2}\partial_v^2 h\,\sigma^2\right) dt + \partial_v h\,\sigma\, dW_t .
$$

SDE with state independent unit diffusion term

Z_t satisfies the SDE with constant, unitary diffusion coefficient,

$$
\begin{aligned}
dZ_t = & \left[\frac{\dot{p}_t - \theta_t \left(1 - p_t - \sin^2 \left(-\sqrt{\frac{\alpha\theta_0}{2}} Z_t \right) \right)}{\sqrt{2\alpha\theta_0} \cos \left(-\sqrt{\frac{\alpha\theta_0}{2}} Z_t \right) \sin \left(-\sqrt{\frac{\alpha\theta_0}{2}} Z_t \right)} \right. \\
& \left. - \frac{1}{4} \frac{\sqrt{2\alpha\theta_0} \left(1 - 2\cos^2 \left(-\sqrt{\frac{\alpha\theta_0}{2}} Z_t \right) \right)}{\cos \left(-\sqrt{\frac{\alpha\theta_0}{2}} Z_t \right) \sin \left(-\sqrt{\frac{\alpha\theta_0}{2}} Z_t \right)} \right] dt + dW_t \qquad (9) \\
= & \left[\frac{2\dot{p}_t - \theta_t(1 - 2p_t) + (\alpha\theta_0 - \theta_t) \cos(-\sqrt{2\alpha\theta_0} Z_t)}{\sqrt{2\alpha\theta_0} \sin(-\sqrt{2\alpha\theta_0} Z_t)} \right] dt + dW_t .
\end{aligned}
$$

Z-Forecast error *transition* histograms after Lamperti T.



Likelihood in the $V-$space (1/2)

- M non-overlapping paths of the continuous-time Itô process V.
- Each path is sampled at $N+1$ equispaced discrete points with a given interval length Δ.
- We denote this random sample by

$$V^{M,N+1} = \left\{ V_{t_1}^{N+1}, V_{t_2}^{N+1}, \ldots, V_{t_M}^{N+1} \right\},$$

where t_j is the start time of the path j and $V_{t_j}^{N+1} = \left\{ V_{t_j+i\Delta}, i = 0, \ldots, N \right\}$, $\forall j \in \{1, \ldots, M\}$.

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.

Likelihood in the $V-$space (2/2)

- The Itô process defined by the SDE (7) is Markovian.
- The likelihood function of the sample $V^{M,N+1}$ can be written as follows:

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

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

Obs: We have used an independence assumption over the index j in the likelihood above.

Moment matching technique

- Closed-form expression for the transition densities of V, $\rho\left(V_{j,i}|V_{j,i-1};\boldsymbol{\theta}\right)$ are rarely available (Egorov et al., 2003).
- Approximate likelihood methods (Särkkä and Solin, 2019, Chapter 9).
- Moment matching technique:
  - assume a surrogate transition density for V.
  - match the conditional moments of the surrogate density for V with the conditional moments of the SDE models (7).

$m_1(t) \equiv \mathbb{E}\left[V_t |\, V_{t_{j,i-1}} = v_{j,i-1}\right] = e^{-\int_{t_{j,i-1}}^{t} \theta_s ds}\, v_{j,i-1}$, for any $t \in [t_{j,i-1}, t_{j,i}[$, $j = 1, \ldots, M$ and $i = 1, \ldots, N$.

For $k \geq 2$, let $m_k(t) \equiv \mathbb{E}\left[V_t^k | V_{t_{j,i-1}} = v_{j,i-1}\right]$ apply Itô's lemma on $g(V_t) = V_t^k$, yielding

Moment matching technique

$$\frac{dm_k(t)}{dt} = -k(\theta_t + (k-1)\alpha\theta_0)m_k(t) + k(k-1)\alpha\theta_0(1-2p_t)m_{k-1}(t) + k(k-1)\alpha\theta_0 p_t(1-p_t)m_{k-2}(t). \quad (10)$$

with initial conditions $m_k(t_{j,i-1}) = v_{j,i-1}^k$.

For any $t \in [t_{j,i-1}, t_{j,i}[$, the first two moments of V, $m_1(t)$ and $m_2(t)$, solve the following ODE system

$$\begin{cases} \frac{dm_1(t)}{dt} &= -m_1(t)\theta_t \\ \frac{dm_2(t)}{dt} &= -2(\theta_t + \alpha\theta_0)m_2(t) + 2\alpha\theta_0(1-2p_t)m_1(t) \\ & \quad +2\alpha\theta_0 p_t(1-p_t) \end{cases} \quad (11)$$

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

Approximate log-likelihood in the $V-$space

- For any $t \in [t_{j,i-1}, t_{j,i}[$, approximate the transition densities of the process V using a Beta distribution (the invariant distribution of the Jacobi type processes) with parameters ξ_1 and ξ_2.

$$\xi_1(t) = -\frac{(\mu_t + 1 - \epsilon)(\mu_t^2 + \sigma_t^2 - (1-\epsilon)^2)}{2(1-\epsilon)\sigma_t^2},$$
$$\xi_2(t) = \frac{(\mu_t - 1 + \epsilon)(\mu_t^2 + \sigma_t^2 - (1-\epsilon)^2)}{2(1-\epsilon)\sigma_t^2}, \qquad (12)$$

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

- The approximate log-likelihood $\tilde{\ell}(\cdot; v^{M,N+1})$ of the observed sample $v^{M,N+1}$:

Approximate log-likelihood in the $V-$space

$$
\begin{aligned}
&\tilde{\ell}\left(\boldsymbol{\theta}; v^{M,N+1}\right) \\
&= \sum_{j=1}^{M}\sum_{i=1}^{N} \log\left\{ \frac{1}{2(1-\epsilon)} \frac{1}{B(\xi_1(t_{j,i}^-), \xi_2(t_{j,i}^-))} \left(\frac{v_{j,i}+1-\epsilon}{2(1-\epsilon)}\right)^{\xi_1(t_{j,i}^-)-1} \right. \\
&\qquad \left. \times \left(\frac{1-\epsilon-v_{j,i}}{2(1-\epsilon)}\right)^{\xi_2(t_{j,i}^-)-1} \right\}, \quad (13)
\end{aligned}
$$

where the shape parameters $\xi_1(t_{j,i}^-)$ and $\xi_2(t_{j,i}^-)$, according to (12), depend on the left limit moments, $\mu(t_{j,i}^-; \boldsymbol{\theta})$ and $\sigma^2(t_{j,i}^-; \boldsymbol{\theta})$, as $t \uparrow t_{j,i}$. These are computed solving numerically the initial-value problem (11). $B(\xi_1, \xi_2)$ denotes the Beta distribution with parameters ξ_1 and ξ_2.

Approximate likelihood in the $Z-$space

The transition density of the process Z, which has been defined through the Lamperti transformation (8) 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 (9) 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 $\tilde{v}_Z(t) \approx \text{Var}[Z_t]$:

$$\begin{cases} \frac{d\tilde{\mu}_Z(t)}{dt} & = a\big(\tilde{\mu}_Z(t); p_t, \dot{p}_t, \boldsymbol{\theta}\big) \\ \frac{d\tilde{v}_Z(t)}{dt} & = 2a'\big(\tilde{\mu}_Z(t); p_t, \dot{p}_t, \boldsymbol{\theta}\big)\tilde{v}_Z(t) + 1 \end{cases} \tag{14}$$

Approximate likelihood in the $Z-$space

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 $\tilde{\ell}_Z\left(\cdot; z^{M,N+1}\right)$ for the observed sample $z^{M,N+1}$ is given by

$$\tilde{\ell}_Z\left(\boldsymbol{\theta}; z^{M,N+1}\right) = \sum_{j=1}^{M}\sum_{i=1}^{N} \log\left\{\frac{1}{\sqrt{2\pi\tilde{v}_Z(t_{j,i}^-; \boldsymbol{\theta})}} \exp\left(-\frac{(z_{j,i} - \tilde{\mu}_Z(t_{j,i}^-; \boldsymbol{\theta}))^2}{2\tilde{v}_Z(t_{j,i}^-; \boldsymbol{\theta})}\right)\right\}, \quad (15)$$

where the limits $\tilde{\mu}_Z(t_{j,i}^-; \boldsymbol{\theta})$ and $\tilde{v}_Z(t_{j,i}^-; \boldsymbol{\theta})$ are computed solving numerically the initial-value problem (14).

Initial guess for (θ_0, α)

We use least square minimization and quadratic variation over the data to find an initial guess (θ_0^*, α^*).
We consider the observed data $v^{M,N+1}$ with length between observations Δ, where $i \in \{0, \dots, N-1\}$ and $j \in \{1, \dots, M\}$.

- For any $t \in [t_{j,i}, t_{j,i+1}[$, the random variable $(V_{j,i+1}|v_{j,i})$ has a conditional mean that can be approximated by the solution of the following system:

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

  in the limit $t \uparrow t_{j,i+1}$, i.e., $\mathbb{E}[V](t^-_{j,i+1})$.
- If we assume that $\theta_t = c \in \mathbb{R}^+$ for all $t \in [t_{j,i}, t_{j,i+1}[$, then $\mathbb{E}[V](t^-_{j,i+1}) = v_{j,i}e^{-c\Delta}$.

Initial guess for (θ_0, α)

- Given $M \times N$ transitions, we can write the regression problem for the conditional mean with L^2 loss function as:

$$
\begin{aligned}
c^* &= \arg\min_{c \geq 0} \left[\sum_{j=1}^{M} \sum_{i=0}^{N-1} \left(v_{j,i+1} - \mathbb{E}[V] \left(t_{j,i+1}^{-} \right) \right)^2 \right] \\
&= \arg\min_{c \geq 0} \left[\sum_{j=1}^{M} \sum_{i=0}^{N-1} \left(v_{j,i+1} - v_{j,i} e^{-c\Delta} \right)^2 \right] \\
&\approx \arg\min_{c \geq 0} \left[\sum_{j=1}^{M} \sum_{i=0}^{N-1} (v_{j,i+1} - v_{j,i}(1 - c\Delta))^2 \right] \qquad (16)
\end{aligned}
$$

Initial guess for (θ_0, α)

Least square minimization

- As equation (16) is convex in c, then

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

Set $\theta_0^* = c^*$.

Initial guess for (θ_0, α)

Quadratic variation

We approximate

- the quadratic variation of the Itô's process V is $[V]_t = \int_0^t b(V_s; \boldsymbol{\theta}, p_s)^2 \mathrm{d}s$ where $b(V_s; \boldsymbol{\theta}, p_s) = \sqrt{2\alpha\theta_0 \left(V_s + p_s\right)\left(1 - V_s - p_s\right)}$ with
- the discrete process quadratic variation : $\sum_{0<t_{j,i}\leq t} \left(V_{t_{j,i+1}} - V_{t_{j,i}}\right)^2$.

Initial guess for the diffusion variability coefficient $\theta_0\alpha$:

$$\theta_0^*\alpha^* \approx \frac{\sum_{j=1}^{M}\sum_{i=0}^{N-1}\left(v_{j,i+1} - v_{j,i}\right)^2}{2\Delta\sum_{j=1}^{M}\sum_{i=0}^{N-1}\left(v_{j,i+1} + p_{j,i+1}\right)\left(1 - v_{j,i+1} - p_{j,i+1}\right)}$$

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

Model specification with the additional parameter δ

To ensure that $E(X_t) = p_t$ at all times, we need $E[V_0] = 0$. For most days, the forecast error at time $t_{j,0} = 0$ is not zero.

1. Assume that there is a time in the past $t_{j,-\delta} < t_{j,0}$, such that the forecast error is zero, $V_{j,-\delta} = 0$.
2. Extrapolate backward linearly the truncated prediction function to get its value at time $t_{j,-\delta}$, $p_{j,-\delta}$, and set $v_{t_{j,-\delta}} = 0$.

   Given the parameters (θ_0, α), find δ by *maximizing the likelihood of initial transitions*:

$$\arg\max_{\delta} \tilde{\mathcal{L}}_{\delta}\left(\boldsymbol{\theta}, \delta; v^{M,1}\right) = \arg\max_{\delta} \prod_{j=1}^{M} \rho_0\left(v_{j,0} | v_{j,-\delta}; \boldsymbol{\theta}, \delta\right), \qquad (17)$$

where $\tilde{\mathcal{L}}_{\delta}$ is the approximated $\delta-$likelihood.
Now assume that the initial transition density has a Beta distribution and apply the moment matching technique.

Model specification with the additional parameter δ

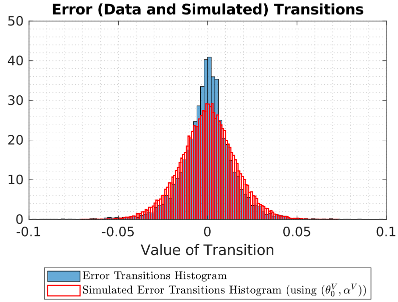
The approximated complete likelihood $\tilde{\mathcal{L}}_c$, which estimates the vector $(\theta_0, \alpha, \delta)$, is given by

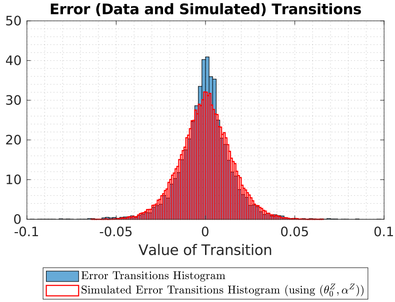
$$\tilde{\mathcal{L}}_c\left(\boldsymbol{\theta}, \delta; v^{M,N+1}\right) = \tilde{\mathcal{L}}\left(\boldsymbol{\theta}; v^{M,N+1}\right)\tilde{\mathcal{L}}_\delta\left(\boldsymbol{\theta}, \delta; v^{M,1}\right), \tag{18}$$

where $\tilde{\mathcal{L}}\left(\boldsymbol{\theta}; v^{M,N+1}\right)$ is the non-log version of (13). As we can provide initial guesses for $\boldsymbol{\theta}$ and δ, we have a starting point for the numerical optimization of the approximated complete likelihood (18).

Partition the 147 segments of normalized wind power production, each 24-hours long. 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.

- Optimal parameters in the V-space: $(\theta_0^V, \alpha^V) = (1.93, 0.050)$
- Optimal parameters in the Z-space: $(\theta_0^Z, \alpha^Z) = (1.87, 0.043)$





Model comparison and assessment of the forecast providers.

- Model 1: (Elkantassi et al., 2017, p.383): This model does not feature derivative tracking:

$$\begin{cases} dX_t &= -\theta_0(X_t - p_t)dt + \sqrt{2\alpha\theta_0 X_t(1-X_t)}dW_t, \quad t \in [0, T] \\ X_0 &= x_0 \in [0,1], \end{cases} \tag{19}$$

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

- Model 2: This model features derivative tracking and time-varying mean-reversion parameter, $\theta_t = \max\left(\theta_0, \frac{\alpha\theta_0 + |\dot{p}_t|}{\min(p_t, 1-p_t)}\right)$,

$$\begin{cases} dX_t &= \left(\dot{p}_t - \theta_t(X_t - p_t)\right)dt + \sqrt{2\alpha\theta_0 X_t(1-X_t)}dW_t, \quad t \in [0, T] \\ X_0 &= x_0 \in [0,1], \end{cases} \tag{20}$$

  with $\theta_0 > 0$, $\alpha > 0$ and θ_t satisfying condition (6).

Application: Uruguay wind and forecast dataset

Table 1: Model comparison.

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

The optimal estimates of the parameters of Model 2, for the three forecast providers, with Beta surrogates for the transition density:

Table 2: Optimal parameters for the three different forecast providers using Model 2 with Beta proxies.

Forecast Provider	Parameters (θ_0, α)	Product $\theta_0\alpha$
Provider A	$(1.93, 0.050)$	**0.097**
Provider B	$(1.42, 0.069)$	0.098
Provider C	$(1.38, 0.078)$	0.108

Application: Uruguay wind and forecast dataset

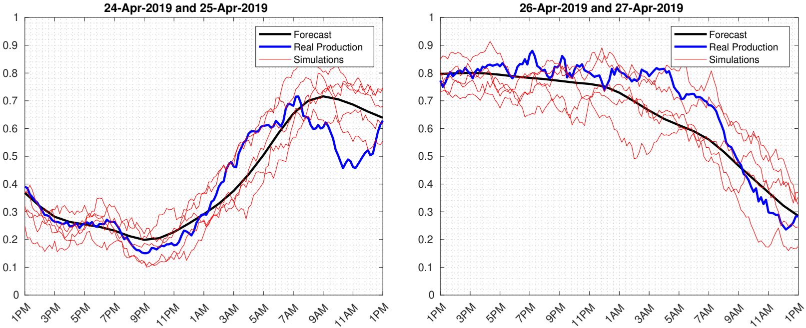


Figure 8: Two days with five simulated wind power production paths.

Given optimal estimates of the parameters of the complete likelihood for Model 2, obtain empirical pointwise confidence bands for wind power production (5000 simulations per day).

Application: Uruguay wind and forecast dataset

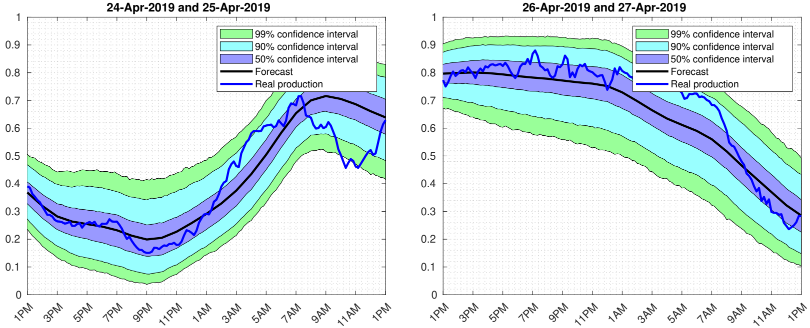


Figure 9: Empirical pointwise confidence bands for the wind power production using the approximate MLEs for Model 2.

Summary and conclusions

- A methodology is developed to assess the short-term forecast of the normalized wind power, which is agnostic of the wind power forecasting technology.
- We built a phenomenological stochastic differential equation model for the normalized wind power production forecast error, with time-varying mean-reversion parameter and time-derivative tracking of the forecast in the linear drift coefficient, and state-dependent and time non-homogenous diffusion coefficient.
- The Lamperti transform with unknown parameters provides a version of the proposed model with a unit diffusion coefficient.
- We used approximate likelihood-based methods for models' calibration.
- The incorporation of an early transition with an additional parameter accounts for the forecast's uncertainty at the beginning of each future period.

Summary and conclusions

- We obtained a robust procedure for synthetic data generation that, using the available forecast input, embraces future wind power production paths through empirical pointwise bands with prescribed confidence.
- Application to the wind power production and three forecast providers dataset in Uruguay between April and December 2019.
- An objective tool is available for forecast assessment and comparison through model selection.
- This work contributes toward the efficient management of renewable energies.

Main references

Alfonsi, A. (2015). *Affine Diffusions and Related Processes: Simulation, Theory and Applications* (Vol. 6). Springer.

Badosa, J., Gobet, E., Grangereau, M., & Kim, D. (2018). Day-ahead probabilistic forecast of solar irradiance: A stochastic differential equation approach (P. Drobinski, M. Mougeot, D. Picard, R. Plougonven, & P. Tankov, Eds.). In P. Drobinski, M. Mougeot, D. Picard, R. Plougonven, & P. Tankov (Eds.), *Renewable energy: Forecasting and risk management*, Cham, Springer International Publishing.

Caballero, R., Kebaier, A., Scavino, M., & Tempone, R. (2021). Quantifying uncertainty with a derivative tracking SDE model and application to wind power forecast data. *Stat. Comput.*, *31*(5), Paper No. 64, 19. https://doi.org/10.1007/s11222-021-10040-8

Carlsson, J., Moon, K.-S., Szepessy, A., Tempone, R., & Zouraris, G. (2010). Stochastic differential equations: Models and numerics. *Lecture notes.*

Main references

Chang, W.-Y. (2014). A Literature Review of Wind Forecasting Methods. *Journal of Power and Energy Engineering*, *2*(4), 161–168. https://doi.org/10.4236/jpee.2014.24023

Egorov, A. V., Li, H., & Xu, Y. (2003). Maximum likelihood estimation of time-inhomogeneous diffusions. *Journal of Econometrics*, *114*, 107–139.

Elkantassi, S., Kalligiannaki, E., & Tempone, R. (2017). Inference and Sensitivity in Stochastic Wind Power Forecast Models. In M. Papadrakakis, V. Papadopoulos, & G. Stefanou (Eds.), *2nd ECCOMAS Thematic Conference on Uncertainty Quantification in Computational Sciences and Engineering* (pp. 381–393). Rhodes Island, Greece, Eccomas Proceedia UNCECOMP 2017. https://doi.org/10.7712/120217.5377.16899

Main references

Forman, J. L., & Sørensen, M. (2008). The Pearson diffusions: A class of statistically tractable diffusion processes. *Scandinavian Journal of Statistics*, *35*(3), 438–465.

Giebel, G., Brownsword, R., Kariniotakis, G., Denhard, M., & Draxl, C. (2011). The state-of-the-art in short-term prediction of wind power: A literature overview. *ANEMOS. plus*.

Iacus, S. M. (2008). *Simulation and Inference for Stochastic Differential Equations: With R Examples*. New York, Springer.

IRENA. (2019). *Innovation landscape for a renewable-powered future: Solutions to integrate variable renewables*. Abu Dhabi.

IRENA. (2020). *Renewable Energy Statistics 2020 The International Renewable Energy Agency*. Abu Dhabi.

Jang, H. S., Bae, K. Y., Park, H.-S., & Sung, D. K. (2016). Solar power prediction based on satellite images and support vector machine. *IEEE Transactions on Sustainable Energy*, *7*(3), 1255–1263.

Main references

Karatzas, I., & Shreve, S. E. (1998). Brownian motion. In *Brownian motion and stochastic calculus* (pp. 47–127). New York, NY, Springer New York. https://doi.org/10.1007/978-1-4612-0949-2_2

Kuo, H. (2006). *Introduction to stochastic integration. universitext*, Springer New York.

Lamperti, J. (1964). A simple construction of certain diffusion processes. *J. Math. Kyoto Univ.*, *4*(1), 161–170. https://doi.org/10.1215/kjm/1250524711

Møller, J. K., Zugno, M., & Madsen, H. (2016). Probabilistic Forecasts of Wind Power Generation by Stochastic Differential Equation Models. *Journal of Forecasting*, *35*(3), 189–205.

Rana, M., Koprinska, I., & Agelidis, V. G. (2016). Univariate and multivariate methods for very short-term solar photovoltaic power forecasting. *Energy Conversion and Management*, *121*, 380–390.

REN21. (2019). *Renewables 2019 Global Status Report*. Paris.

Main references

Särkkä, S., & Solin, A. (2019). *Applied Stochastic Differential Equations*. Cambridge University Press.

Wu, Y.-K., Chen, C.-R., & Abdul Rahman, H. (2014). A novel hybrid model for short-term forecasting in pv power generation. *International Journal of Photoenergy*, *2014*.

Zhou, Z., Botterud, A., Wang, J., Bessa, R., Keko, H., Sumaili, J., & Miranda, V. (2013). Application of probabilistic wind power forecasting in electricity markets. *Wind Energy*, *16*(3), 321–338. https://doi.org/10.1002/we.1496

Thank you very much for your attention!

Behavior of δ as function of $\boldsymbol{\theta}$

- 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, less time is needed 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, the mean reversion becomes larger and reduces the variance for the initial transition density. Then, more time is needed for the initial transition density to cover the initial error observations.

Behavior of δ as function of $\boldsymbol{\theta}$

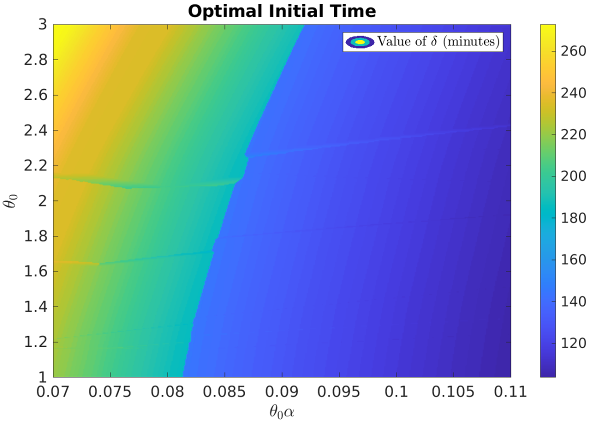


Figure 10: Initial value for δ as a function of the elements of the parameter vector $\boldsymbol{\theta}$.

Theorems

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

$$\begin{cases} dX_t &= \left(\dot{p}_t - \theta_t(X_t - p_t)\right)dt + \sqrt{2\alpha\theta_0 X_t(1 - X_t)}dW_t, \quad t \in [0, T] \\ X_0 &= x_0 \in [0, 1], \end{cases} \tag{21}$$

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

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

The following proofs are based on standard arguments for stochastic processes that can be found e.g. in Alfonsi, 2015 and Karatzas and Shreve, 1998 that we adapted to the setting of our model (21).

Theorem

Assume that

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

Then, there is a unique strong solution to (21) *s.t. for all* $t \in [0,T]$, $X_t \in [0,1]$ *a.s.*

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

$$X_t = x_0 + \int_0^t \left(\dot{p}_s - \theta_s(X_s - p_s)\right) ds + \int_0^t \sqrt{2\alpha\theta_0 |X_s(1-X_s)|}\, dW_s, \quad 0 \le x_0 \le 1. \qquad (22)$$

According to Proposition 2.13, p.291 of Karatzas and Shreve, 1998, under assumption (A) there is a unique strong solution X_t to (22). Moreover, as the diffusion coefficient is of linear growth, we have for all $p > 0$

$$\mathbb{E}\left[\sup_{t\in[0,T]} |X_t|^p\right] < \infty. \tag{23}$$

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 bunch of useful properties:

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

$$|\psi_n(x)| \wedge |x\psi'_n(x)| \leq |x|, \;\; \psi'_n(x) \leq 1,$$

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

See the proof of Proposition 2.13, p. 291 of Karatzas and Shreve, 1998 for the construction of such function. Applying Itô's formula we get

$$\psi_n(X_t) = \psi_n(x_0) + \int_0^t \psi'_n(X_s)\big(\dot{p}_s + \theta_s p_s - \theta_s X_s\big)\,ds + \int_0^t \psi'_n(X_s)\sqrt{2\alpha\theta_0|X_s(1-X_s)|}\,dW_s + \alpha\theta_0 \int_0^t g_n(|X_s|)|X_s(1-X_s)|\,ds.$$

Now, thanks to (A), (23), and to the above properties of ψ_n and g_n, we get

$$\mathbb{E}\left[\psi_n(X_t)\right] \leq \psi_n(x_0) + \int_0^t \left(\dot{p}_s + \theta_s p_s - \theta_s \mathbb{E}[\psi'_n(X_s)X_s]\right)ds + \frac{2\alpha\theta_0}{n}\int_0^t \mathbb{E}\left[|1-X_s|\right]ds.$$

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

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

Besides, taking the expectation of (22), we get

$$\mathbb{E}\left[X_t\right] = x_0 + \int_0^t \left(\dot{p}_s + \theta_s p_s - \theta_s \mathbb{E}\left[X_s\right]\right) ds,$$

and thus we have

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

Then, Gronwall's lemma gives us $\mathbb{E}\left[|X_t|\right] = \mathbb{E}\left[X_t\right]$ 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

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

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

Theorem

Assume that assumptions of Theorem 3 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.*

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

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

so that

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

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

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

By way of contradiction let us assume that $\{\tau_0 < \infty\}$, then letting $t \to \tau_0$ we deduce that

$$\lim_{t\to\infty} 1_{\{\tau_0<\infty\}} M_{t\wedge\tau_0} = -1_{\{\tau_0<\infty\}}\infty \ \text{ a.s.}$$

This leads to a contradiction since we know that continuous martingales likewise the Brownian motion cannot converge almost surely to $+\infty$ or $-\infty$. It follows that $\tau_0 = \infty$ almost surely. Next, recalling that the process $(Y_t)_{t\geq 0}$ given by $Y_t = 1 - X_t$ is solution to

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

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

Remark: As the diffusion coefficient of X_t 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_t and X_t is bijective, so that we deduce the properties of existence and uniqueness of Z_t from those of X_t.

The application of Itô's formula is subjected to the condition (B) that avoids the process X_t hits the boundaries of the interval $]0,1[$, otherwise the Lamperti transform is not applicable.