A DERIVATIVE TRACKING MODEL FOR WIND POWER FORECAST ERROR

BY RENZO CABALLERO[1], AHMED KEBAIER[2,*], MARCO SCAVINO[3,†] AND RAÚL TEMPONE[4,‡]

[1]*CEMSE Division, King Abdullah University of Science and Technology, Saudi Arabia, Renzo.CaballeroRosas@kaust.edu.sa*

[2]*Université Sorbonne Paris Nord, LAGA, CNRS, UMR 7539, F-93430, Villetaneuse, France, *kebaier@math.univ-paris13.fr*

[3]*Instituto de Estadística (IESTA), Universidad de la República, Montevideo, Uruguay, †mscavino@iesta.edu.uy*

[4]*Chair of Mathematics for Uncertainty Quantification, RWTH Aachen University, Germany, ‡tempone@uq.rwth-aachen.de*

Reliable wind power generation forecasting is crucial for applications such as the allocation of energy reserves, optimization for electricity price, and operation scheduling of conventional power plants. We propose a data-driven model based on parametric Stochastic Differential Equations (SDEs) to capture the real asymmetric dynamics of wind power forecast errors. Our SDE framework features time-derivative tracking of the forecast, time-varying mean-reversion parameter, and an improved state-dependent diffusion term. The methodology we developed allows the simulation of future wind power production paths and to obtain sharp empirical confidence bands. All the procedures are agnostic of the forecasting technology, and they enable comparisons between different forecast providers. We apply the model to historical Uruguayan wind power production data and forecasts between April and December 2019.

1. Introduction. Despite the Covid-19 pandemic causing "the biggest shock to the global energy system in more than seven decades" the recent report released on April 30, 2020, by the International Energy Agency (https://www.iea.org/reports/global-energy-review-2020) stresses that projections of energy demand will fall 6% in 2020 – seven times the decline after the 2008 global financial crisis – It is emphasized that "renewables are set to be the only energy source that will grow in 2020, with their share of global electricity generation projected to jump thanks to their priority access to grids and low operating cost" (IEA (2020)).

Although the growth of renewables in electricity generation in 2020 is set to be lower than in previous years, solar photo-voltaic (PV) and wind are on track to help lift renewable electricity generation by 5% in 2020, aided by higher output from hydropower.

Wind and solar energy are expanding renewable generation capacity, experiencing record growth in the last years.
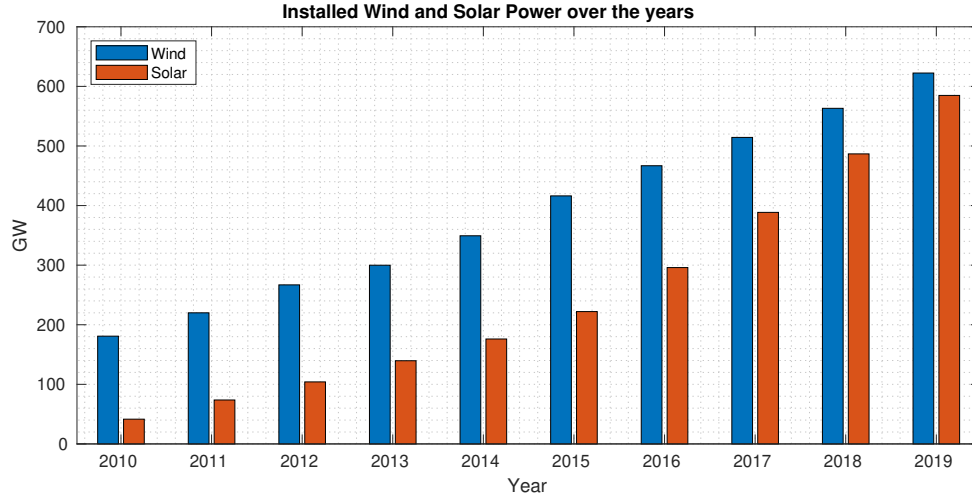
FIG 1. *Worldwide installed wind and solar power 2010–2019 (IRENA, 2020). We recall the importance of accurate forecasts to use green energies optimally.*

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.

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 forecast ensembles. However, there is a lack of simulating and producing data-driven stochastic forecasts based on 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 (2014)). This is due to many factors that 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 (Schicker et al. (2017)). 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 is also asymmetric due to different costs for up and down power regulations, which are determined by the electricity market (Tsitsiklis and Xu (2015)).

We propose to model wind power forecast 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 can evaluate wind power forecasts according to their real-world performance, and we can compare different forecasting technologies. Most notably, we can 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.

Some interesting works have been devoted to probabilistic forecasting related to renewable energies based on stochastic differential equations, among them (Møller, Zugno and Madsen (2016)) on wind power forecast, (Iversen et al. (2014)) and (Badosa et al. (2018)) on forecasts of solar irradiance. 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.

The rest of the paper is organized as follows. In Section 2, we introduce the main characteristics of a real data set encompassing the normalized wind power production in Uruguay between April and December 2019, 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. The application of the Lamperti transform with unknown parameters in Section 4 leads to model the forecast error through a stochastic differential equation with a unit diffusion coefficient. In Section 5, we write down the expressions for the likelihood functions of the forecast error in its original space and 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 numerical estimation procedures 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. Section 7 concludes the paper. The proofs of the existence, strong uniqueness, and boundedness of the SDE solutions 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 (2019), p.23), Uruguay was 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 (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 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 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 between April and December 2019, that we adequately normalized with respect to the present $1474\,\mathrm{MW}$ maximum installed wind power capacity. Each day, wind power production recordings are available every ten minutes. In this work, we 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 hours 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.
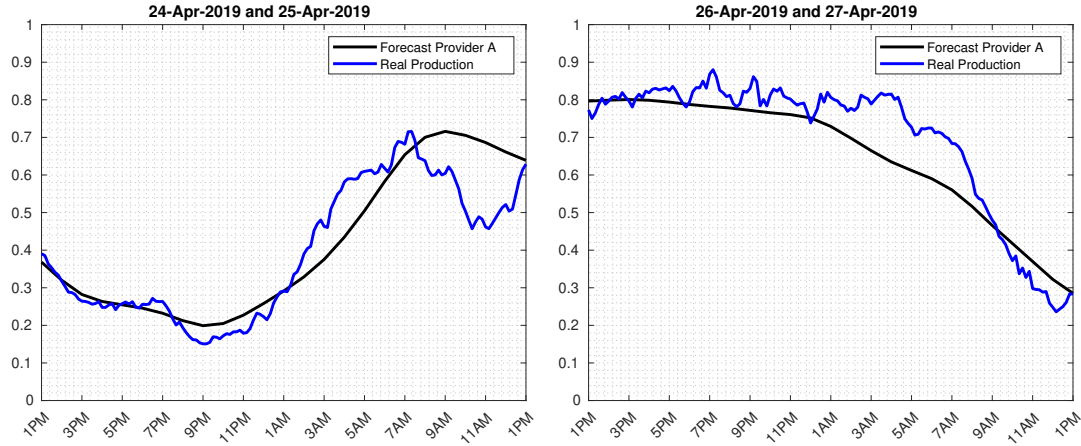

FIG 2. *Two 24-hour segments with the normalized wind power real production in Uruguay (blue line) recorded every ten minutes, and the hourly wind power production forecasted by provider A (black line).*

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 in 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 $]0.6, 1]$.
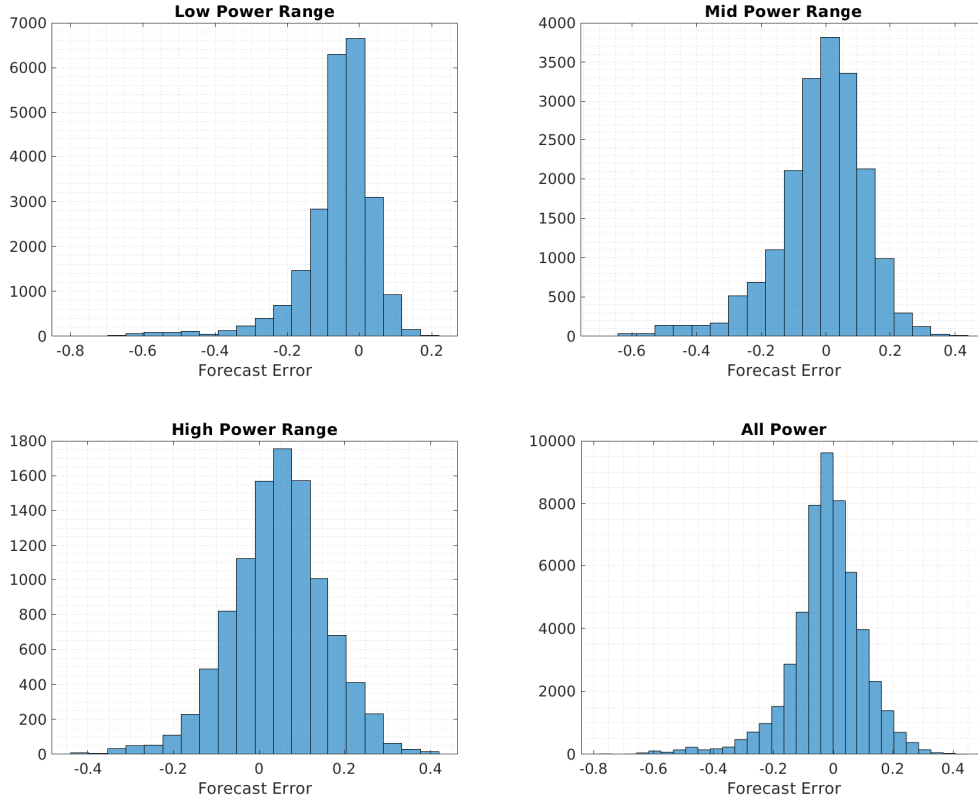
FIG 3. *Wind production forecast error histograms during the period April-December 2019: low-power (upper-left plot), mid-power (upper-right plot), high-power (lower-left plot), and the global range of power (lower-right plot).*

We may observe that all the histograms in Figure 3 exhibit skewed patterns, to a different extent, as well as extreme observations. The presence of these features can be partly explained. The data analysis highlighted that, during several 24-hour segments, the system operators decided to reduce or even cease the wind power production. Indeed, as recalled in (IRENA (2018), p.8), "Uruguay experiences high curtailment levels because generation exceeds demand." Despite the large country's interconnection capacity with Argentina and Brazil, there is no active cross-border market; the energy is traded via ad hoc short-term agreements. (IRENA (2018), p.3) "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 the inclusion of 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 shown below in Figure 4, can appreciate skewness reduction, except for low power forecast error histogram.
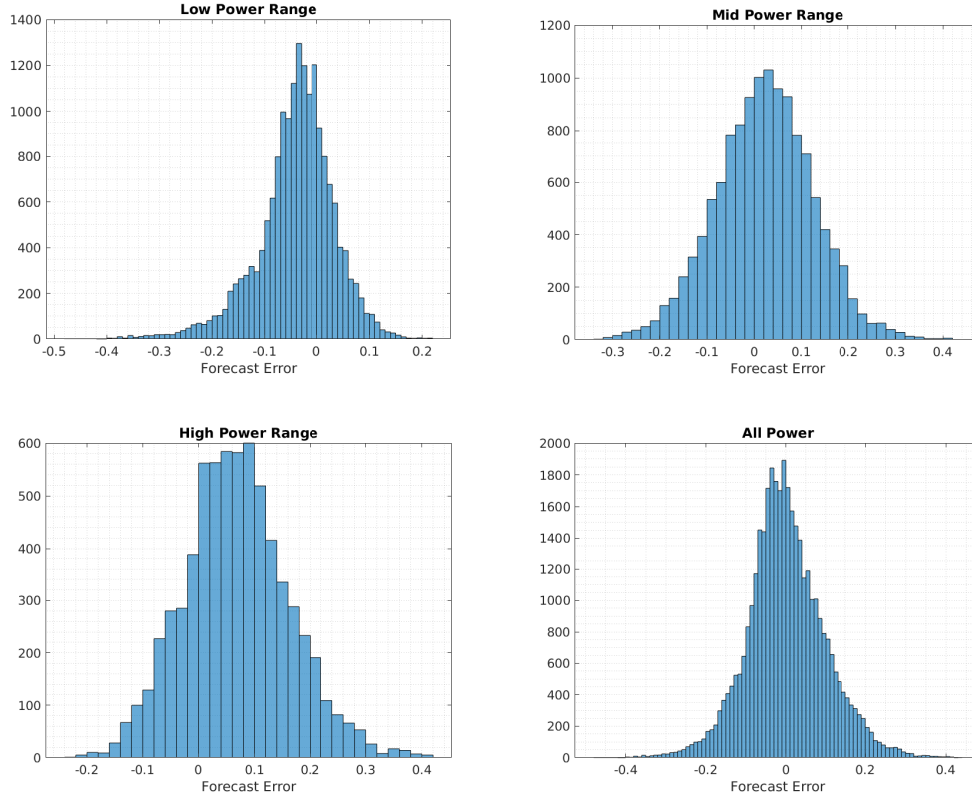
FIG 4. *Wind production forecast error histograms during the period April-December 2019 after removing 24-hour segments with artificial wind curtailment: low-power (upper-left plot), mid-power (upper-right plot), high-power (lower-left plot), and the global range of power (lower-right plot).*

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 later constitute a reference for the visual assessment of the global fit of the proposed models.

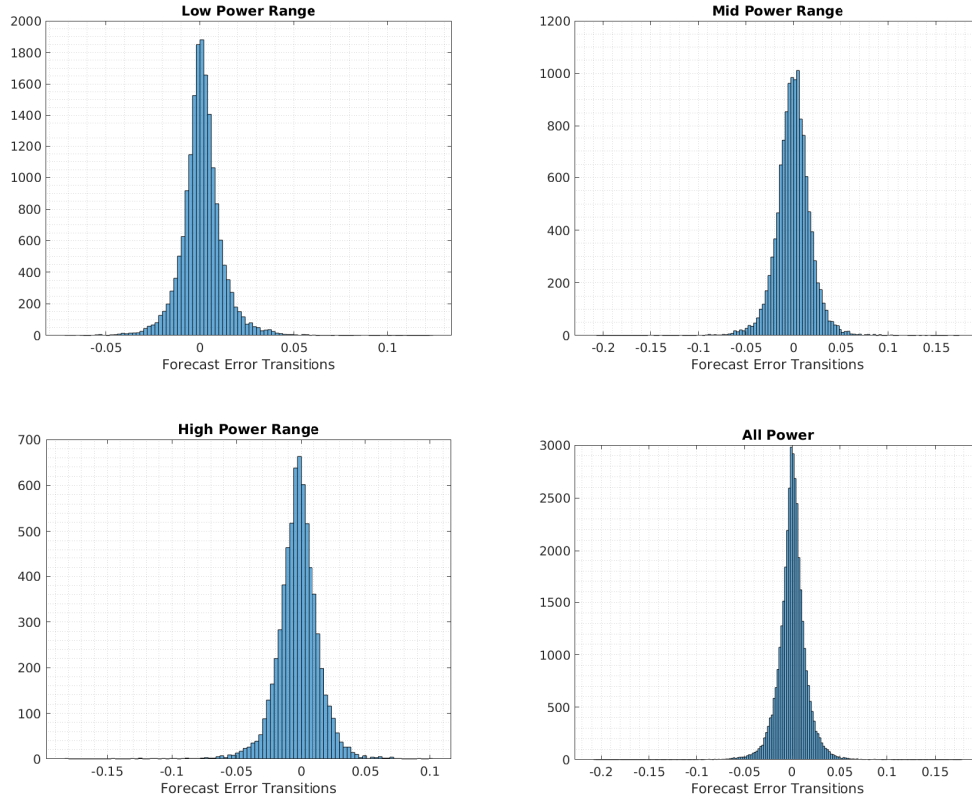FIG 5. *Forecast error transition histograms during the period April-December 2019 without wind power production curtailment: low-power (upper-left plot), mid-power (upper-right plot), high-power (lower-left plot), and the global range of power (lower-right plot).*

The histograms in Figure 5 feature a non-Gaussianity trait and provide initial input for the model-building stage, which is described in the next section. 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):

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

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 a deterministic forecast for the normalized wind power generation, which is provided by Administrator of Electric Market.

Our goal is to achieve a specification of the model (1) to follow the available normalized wind power forecasts closely 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:

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

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_t, 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 leads to

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

Taking expectation on (3) we obtain

$$(4) \qquad \mathbb{E}[X_t] = e^{-\int_0^t \theta_s \, ds} \left(\mathbb{E}[X_0] + p_t \, e^{\int_0^t \theta_s \, ds} - p_0 \right) = p_t.$$

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_t from its mean,
- it tracks the time derivative $\dot{p}_t$.

Remark: Observe that a mean-reverting model without derivative tracking shows a delayed path behavior. For instance, consider the diffusion model (1) with $a(X_t; p_t, \boldsymbol{\theta}) = -\theta_0(X_t - p_t), \theta_0 > 0$. In this case, given $\mathbb{E}[X_0] = p_0$, the diffusion has mean $\mathbb{E}[X_t] = p_t - e^{-\theta_0 t} \int_0^t \dot{p}_s e^{\theta_0 s} \, ds$. The next Figure (6) illustrates how different behave the estimated confidence bands for two diffusion models with and without derivative tracking, fitting the same daily segment.
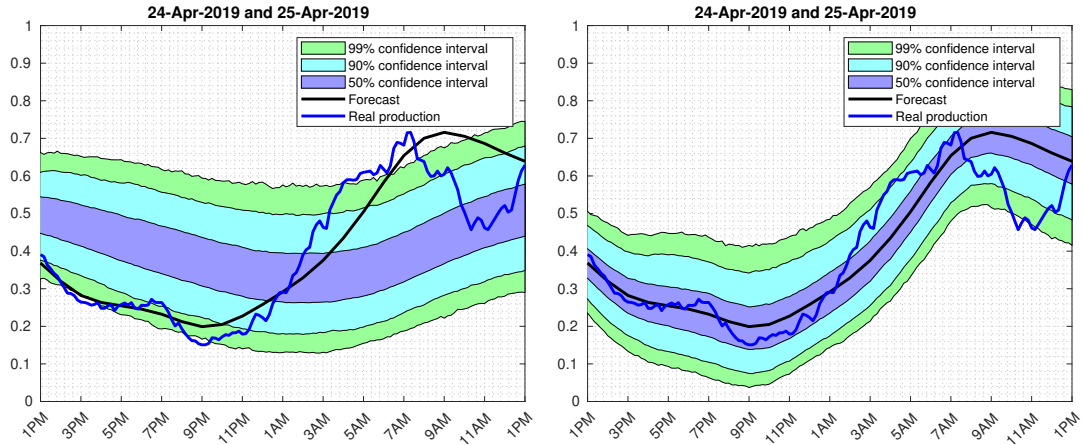

FIG 6. *Pointwise confidence bands fitted, for the same daily segment, through diffusion models without derivative tracking (plot on the left) and with derivative tracking (plot on the right).*

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_t 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_t 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:

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

where $\alpha > 0$ is an 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 (Forman and Sorensen, 2008, 440) a Pearson diffusion is a stationary solution to a stochastic differential equation of the form

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

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 (Forman and Sorensen, 2008, 440-443) 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, (Leonenko and Phillips, 2012, 2860-2861)).

It is worth mentioning that Jacobi diffusions have been successfully applied in several disciplines, among them finance (see (Valéry and Gouriéroux (2011)) and references therein) and neuroscience (D'Onofrio, Tamborrino and Lansky (2018)).

However, a distinctive feature in our proposed model

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

is that the drift term contains the time-varying parameter θ_t, rendering the solution X_t of (7) to a non-stationary and time-inhomogeneous process. To ensure that the process X_t 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:

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

The proof of this theoretical statement is presented in the Appendix.

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

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

that satisfies $p_t^\epsilon \in [\epsilon, 1 - \epsilon]$ for any $0 < \epsilon < \frac{1}{2}$ and $t \in [0, T]$, 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:

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

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 the use of Itô's formula on a well-chosen transformation of the process V, John Lamperti (Lamperti (1964)) first showed that the transformed process is again a diffusion process that is solution to a SDE with unit coefficient for the diffusion term. The vast literature nowadays refers to this result as the so-called Lamperti transform (see, for example, (Iacus, 2008, 40–41); Møller and Madsen (2010); (Panik, 2017, 199–203); (Särkkä and Solin, 2019, 98–100)), 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 Lamperti's change of variable as part of a more complex approximation procedure (for example, in the case of one-dimensional diffusions, the local linearization method in Shoji and Ozaki (1998), or the expansion method in Aït-Sahalia (2002), later extended to time-inhomogeneous SDEs in Egorov, Li and Xu (2003)).

We consider the following Lamperti transform with unknown parameters

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

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

$$(11) \qquad \begin{aligned} \mathrm{d}Z_t = \Bigg[&\frac{\dot{p}_t}{\sqrt{2\alpha\theta_0 (V_t + p_t)(1 - V_t - p_t)}} \\ &+ \frac{-\theta_t V_t}{\sqrt{2\alpha\theta_0 (V_t + p_t)(1 - V_t - p_t)}} - \frac{1}{4} \frac{\sqrt{2\alpha\theta_0} \left(1 - 2(V_t + p_t)\right)}{\sqrt{(V_t + p_t)(1 - V_t - p_t)}} \Bigg] \mathrm{d}t + \mathrm{d}W_t. \end{aligned}$$

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_t satisfies the SDE

$$
\mathrm{d}Z_t = \left[\frac{\dot{p}_t - \theta_t\left(1 - p_t - \sin^2\left(-\sqrt{\frac{\alpha\theta_0}{2}}Z_t\right)\right)}{\sqrt{2\alpha\theta_0}\cos\left(-\sqrt{\frac{\alpha\theta_0}{2}}Z_t\right)\sin\left(-\sqrt{\frac{\alpha\theta_0}{2}}Z_t\right)} \right.
$$

(12)

$$
\left. -\frac{1}{4}\frac{\sqrt{2\alpha\theta_0}\left(1 - 2\cos^2\left(-\sqrt{\frac{\alpha\theta_0}{2}}Z_t\right)\right)}{\cos\left(-\sqrt{\frac{\alpha\theta_0}{2}}Z_t\right)\sin\left(-\sqrt{\frac{\alpha\theta_0}{2}}Z_t\right)} \right]\mathrm{d}t + \mathrm{d}W_t
$$

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

To summarize the effect of the Lamperti transform visually, we can see in the next Figure (7) how the forecast error transition histograms (without curtailment) modify in comparison with Figure (5).

In this case, the Lamperti transform has been applied using the optimal estimates of the parameters in the SDE model (9), obtained applying our numerical procedure detailed later in Subsection 5.4.
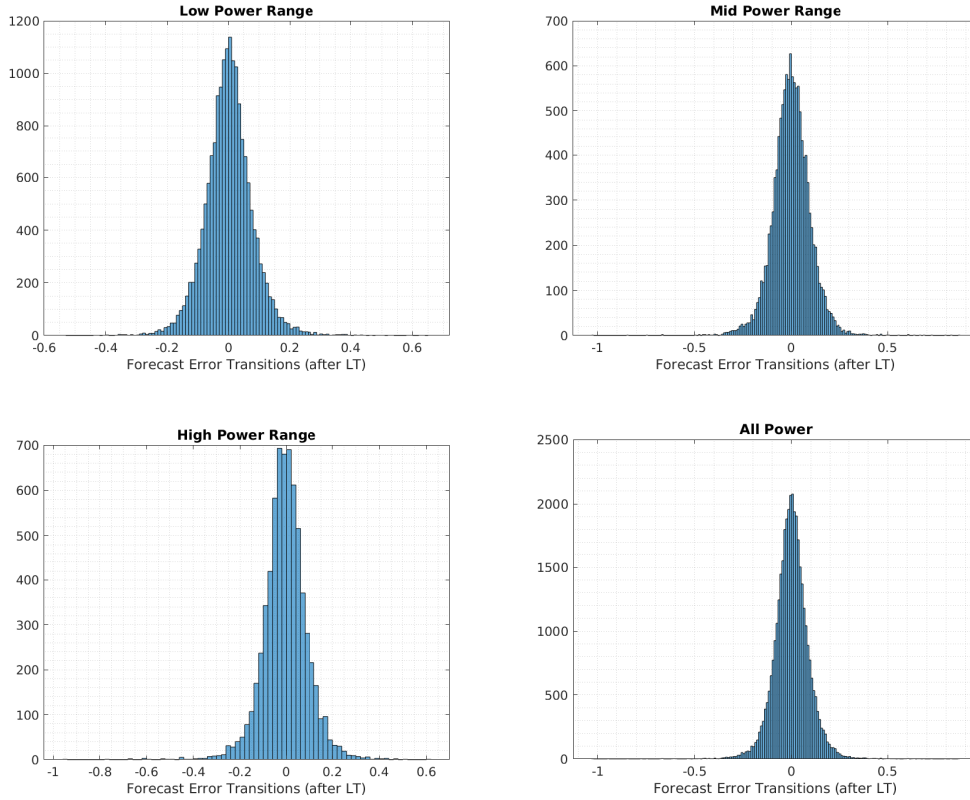

FIG 7. *Lamperti transformed forecast error transition histograms between April and December 2019 without wind power production curtailment: low-power (upper-left plot), mid-power (upper-right plot), high-power (lower-left plot), and the global range of power (lower-right plot).*

The shape of the forecast error transition histograms after applying the Lamperti transform has similarities with the Gaussian distribution, motivating toward the use of Gaussian-like approximations of the unknown density transition functions of the process Z_t.

Remark: Moreover, this obtained Gaussian distribution supports the validity of the choice of our model diffusion coefficient given by (6).

5. Likelihood functions of the forecast error 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 $V = \{V_t, t \in [0, T]\}$, each one starting at a different time t_j with $j = 1, \ldots, M$, is sampled at $N + 1$ equispaced discrete points with given length interval Δ. 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 $V_{t_j^{N+1}} = \left\{ V_{t_j + i\Delta}, i = 0, \ldots, N \right\}$.

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, and the likelihood function of the sample $V^{M,N+1}$ can be written as the following product of transition densities:

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

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

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 together or 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 ((Iacus, 2008, 36); (Särkkä and Solin, 2019, 61-68)):

$$\frac{\partial f}{\partial t} \rho(v, t|v_{j,i-1}, t_{j,i-1}; \boldsymbol{\theta}) = -\frac{\partial}{\partial v}(-\theta_t v\, \rho(v, t|v_{j,i-1}, t_{j,i-1}; \boldsymbol{\theta}))$$

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

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 $V_{j,i-1}$.

However, closed-form solutions to initial-boundary value problems for time-inhomogeneous diffusions can be obtained only in a few cases (see, for example, (Egorov, Li and Xu, 2003, Section 3.1)). In our case, solving numerically (14) for the transition densities of the process V at every transition step is computationally expensive. Several numerical techniques have been devised to obtain estimates for the unknown parameters of continuous-time SDE models with discrete observations (see, for example, Preston and Wood (2012) for likelihood-based

inference techniques, Sørensen (2012) for an estimating function approach). As explained in the next subsection, we have considered approximate likelihood methods, similar in spirit to (Särkkä and Solin, 2019, Section 11.4).

5.2. *Approximate likelihood in the $V-$space.* Gaussian approximations to the transition densities of nonlinear time-inhomogeneous SDEs are available through different algorithms (Särkkä and Solin, 2019, Chapter 9). However, as Figure 5 may suggest at 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 surrogate density moments.

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

For $m \geq 2$, using Itô's lemma we derive
(15)
$$\frac{\mathrm{d}\mathbb{E}[V_t^m]}{\mathrm{d}t} = -m\theta_t \mathbb{E}[V_t^m] + m(m-1)\alpha\theta_0 \mathbb{E}\left[-V_t^m + (1-2p_t)V_t^{m-1} + p_t(1-p_t)V_t^{m-2}\right].$$

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

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

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 parameters, ξ_1, ξ_2. Recall that the choice of the beta proxy distribution is a natural choice as it is the invariant distribution of the Jacobi type processes.

For any $t \in [t_{j,i-1}, t_{j,i}[$, we 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
(17)
$$\xi_1(t) = -\frac{(\mu_t + 1 - \epsilon)(\mu_t^2 + \sigma_t^2 - (1-\epsilon)^2)}{2(1-\epsilon)\sigma_t^2}, \quad \xi_2(t) = \frac{(\mu_t - 1 + \epsilon)(\mu_t^2 + \sigma_t^2 - (1-\epsilon)^2)}{2(1-\epsilon)\sigma_t^2},$$

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 $v^{M,N+1}$ can be expressed as

(18) $\tilde{\ell}\left(\boldsymbol{\theta}; v^{M,N+1}\right)$

$$= \sum_{j=1}^M \sum_{i=1}^N \log \left\{ \frac{1}{2(1-\epsilon)} \frac{1}{B(\xi_1(t_{j,i}^-), \xi_2(t_{j,i}^-))} \left(\frac{v_{j,i} + 1 - \epsilon}{2(1-\epsilon)}\right)^{\xi_1(t_{j,i}^-)-1} \left(\frac{1 - \epsilon - v_{j,i}}{2(1-\epsilon)}\right)^{\xi_2(t_{j,i}^-)-1} \right\},$$

where the shape parameters $\xi_1(t_{j,i}^-)$ and $\xi_2(t_{j,i}^-)$, according to (17), depend on the limit quantities $\mu(t_{j,i}^-; \boldsymbol{\theta})$ and $\sigma^2(t_{j,i}^-; \boldsymbol{\theta})$ as $t \uparrow t_{j,i}$ that are computed solving numerically the initial-value problem (16). $B(\xi_1, \xi_2)$ denotes the Beta distribution with parameters ξ_1 and ξ_2.

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

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

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

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

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

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

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 (19).

5.4. *Algorithm for the approximate maximum likelihood estimations.* In this subsection, we aim to infer the model's parameters using optimization techniques. We start by finding an initial guess close enough to the optimal value, and from that point, start the optimization.

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 (θ_0^*, α^*).

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

$$
\begin{cases} \mathrm{d}\mathbb{E}[V](t) & = -\theta_t\mathbb{E}[V](t)\,\mathrm{d}t \\ \mathbb{E}[V](t_{j,i-1}) & = v_{j,i-1}, \end{cases}
$$

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

$$
\hat{c} = \arg\min_{c \geq 0}\left[\sum_{j=1}^{M}\sum_{i=1}^{N}\left(v_{j,i} - \mathbb{E}[V](t_{j,i}^-)\right)^2\right]
$$

(21)

$$
\approx \arg\min_{c \geq 0}\left[\sum_{j=1}^{M}\sum_{i=1}^{N}\left(v_{j,i} - v_{j,i-1}(1 - c\Delta)\right)^2\right].
$$

As Equation (21) is convex in c, it is enough to verify the first order optimality conditions. It follows that

$$(22) \qquad \hat{c} \approx \frac{\sum_{j=1}^{M} \sum_{i=1}^{N} v_{j,i-1}(v_{j,i-1} - v_{j,i})}{\Delta \cdot \sum_{j=1}^{M} \sum_{i=1}^{N} (v_{j,i-1})^2}.$$

We approximate θ_0 by Equation (22) setting $\theta_0^* = \hat{c}$.

• Quadratic variation: We approximate the quadratic variation of the Itô's process V

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

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

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

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

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 $h(v_{j,i}, t_{j,i}; \boldsymbol{\theta})$ the Lamperti transform of the observation $v_{j,i}$. As we can see in Section 4, the transformed observations $z^{M,N+1}$ depend on the vector $\boldsymbol{\theta}$.

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 $\boldsymbol{\theta}^\star$ such that

$$(24) \qquad \boldsymbol{\theta}^\star = \arg \max_{\boldsymbol{\theta}} \tilde{\ell}_Z \left(\boldsymbol{\theta}; \{h(v_{j,i}, t_{j,i}; \boldsymbol{\theta}^\star)\}_{j=1,i=0}^{M,N} \right).$$

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

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

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

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

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

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

16

and remark that $\boldsymbol{\theta}^\star$ is a fixed point if and only if $\Psi(\boldsymbol{\theta}^\star) = 0$. 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 $t_{j,-\delta}$, $p_{j,-\delta}$, 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 vector of parameters $\boldsymbol{\theta}$, we estimate δ solving the following problem

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

where $\tilde{\mathcal{L}}_\delta$ is the approximated $\delta-$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 $p_{j,-\delta}$ at each $j \in \{1, 2, \ldots, M\}$.

As remarked, we extend the statistical model (13) to include the extra parameter δ. The approximated complete likelihood $\tilde{\mathcal{L}}_c$, which estimates the vector $(\theta_0, \alpha, \delta)$, is given by

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

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 (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-hours 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. Additional cross-correlation tests were performed to ensure this assumption.

All the following results involving a single provider refer to provider A. Furthermore, all calibrations involve the training sets and all simulations, the test sets. Following the instruction for the initial guesses from Subsection 5.4 and assuming that

$$\theta_t = \max\left(\theta_0, \frac{\alpha\theta_0 + |\dot{p}_t|}{\min(p_t, 1 - p_t)}\right),$$

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 V-space and Z-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. Moreover, we realized an additional inference utilizing the test sets to guarantee the robustness of our numerical methods.
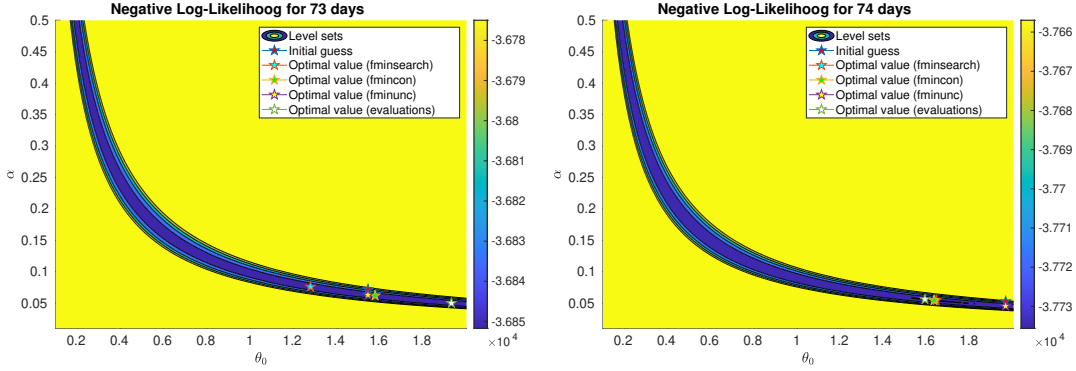
FIG 8. *Negative log-likelihood's level sets for the training sets (plot on the left), and for the test sets (plot on the right). All optimal values are located over the curve $\theta_0\alpha = 0.097$ and $\theta_0\alpha = 0.089$ for the training and test sets, respectively.*

On Figure (8), we can see the level sets for the negative log-likelihood for both training and test sets. The numerical values of each relevant point can be seen in Table (1). We set the optimal parameters in the V-space $(\theta_0^V, \alpha^V) = (1.93, 0.050)$, as it is where the negative log-likelihood for the training sets reaches its minimum value.

	Training sets			Test sets		
	θ_0	α	$\theta_0\alpha$	θ_0	α	$\theta_0\alpha$
Initial guess	1.54	0.072	0.111	1.96	0.053	0.104
fminsearch	1.14	0.076	0.097	1.64	0.054	0.089
fmincon	1.58	0.062	0.097	1.63	0.055	0.089
fminunc	1.54	0.063	0.097	1.96	0.045	0.089
Evaluations	1.93	0.050	0.097	1.59	0.056	0.089

TABLE 1
Coordinates of the initial guess point and of the optimal points represented in Figure 8.

We observe that all the local (possibly global) minimizers are located over the curves $\theta_0\alpha = 0.097$ and $\theta_0\alpha = 0.089$ for the training and test sets, respectively. This effect shows that the optimization is more sensitive to the diffusion than to the drift.

In the Z-space, we obtain the optimal parameters $(\theta_0^Z, \alpha^Z) = (1.87, 0.043)$ as minimizers for the relative error function (25) using *fminsearch* from MATLAB R2019b and the training sets.

To verify and compare these 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).

6.2. *Model comparison and assessment of the forecast providers.* We compare two candidate models to find the best-fit that maximizes the retained information, the Model 1, introduced in (Elkantassi, Kalligiannaki and Tempone (2017), p.383), and our proposed model (7), from hereafter called Model 2.
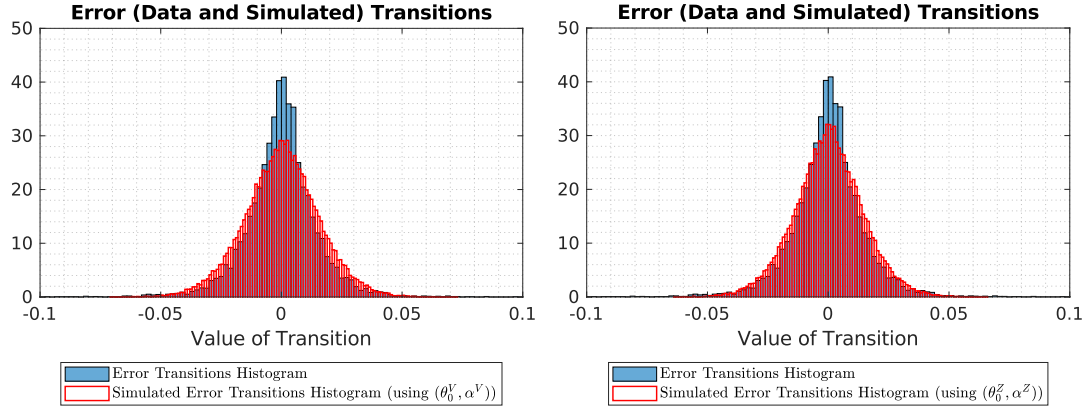
FIG 9. *Density histograms for error transitions. Using provider A, we overlap the real transitions from the test set with the simulated ones from the $V-$space SDE. On the left, simulations use (θ_0^V, α^V). On the right, simulations use (θ_0^Z, α^Z).*

- Model 1: This model does not feature derivative tracking:

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

with $\theta_0 > 0$, $\alpha > 0$.
- Model 2: This model features derivative tracking and time-varying mean-reversion parameter:

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

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

To show the better performance of Model 2, we have computed the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) for the two considered models, and any combination of the three different forecast providers with three approximate likelihood methods, the one based on the Beta surrogate density in the V-space (Subsection 5.2), the one based on the Gaussian surrogate density in the Z-space (Subsection 5.3), and the Shoji-Ozaki local linearization method (Shoji and Ozaki (1998)). Table (2) summarizes these results, also reporting the estimate of the variability diffusion coefficient $\alpha\theta_0$. It is worth observing that the best fitting is achieved with Model 2 and adopting Beta distributions as proxies of the transition densities.

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

Model comparison based on Akaike and Bayesian information criteria.

The optimal estimates of the parameters of Model 2, for the three forecast providers, when using Beta surrogates for the transition density are presented next:

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

TABLE 3

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

6.3. *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 wind power production for each day of interest.
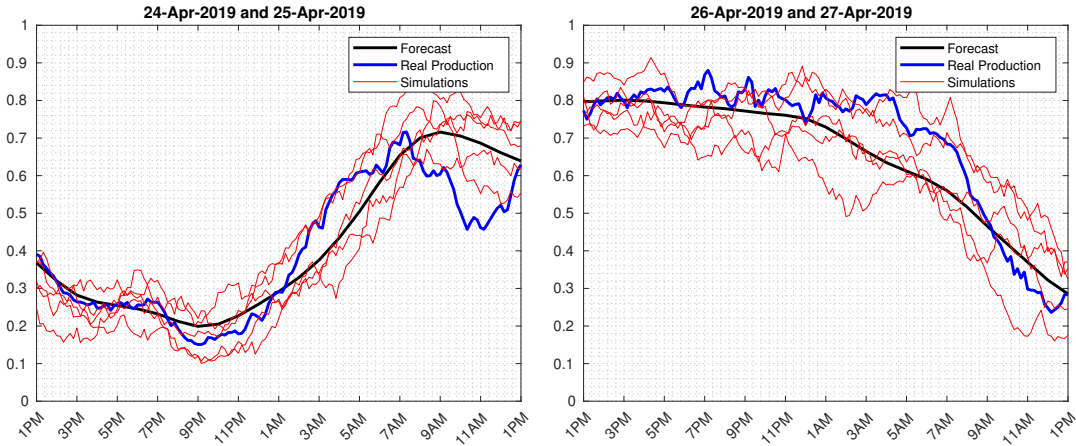

FIG 10. *Two arbitrary days with five simulated wind power production paths each.*

Once derived optimal estimates of the parameters of the complete likelihood for Model 2, we obtain empirical pointwise confidence bands for wind power production. Figure (11)

shows the empirical pointwise confidence bands for wind power production for each day of interest, assuming Model 2 specification, a given forecaster, and 5000 simulations per day.
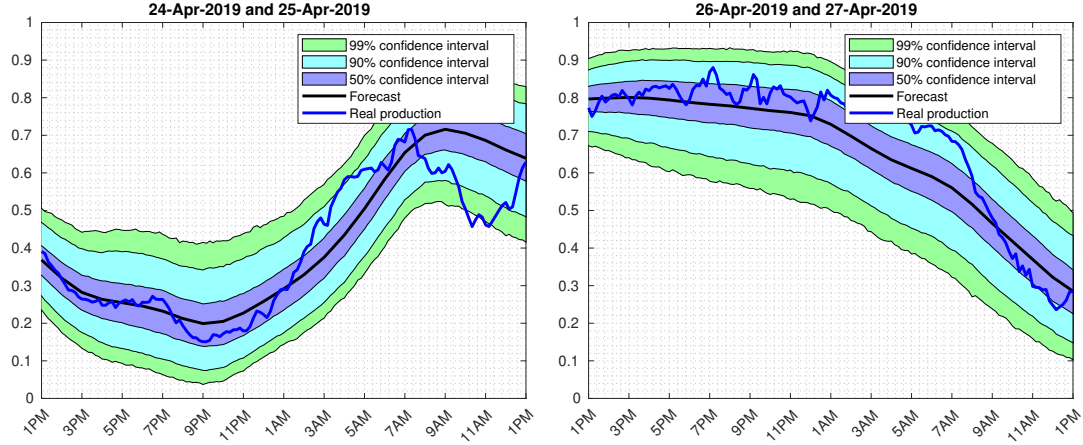

FIG 11. *Empirical pointwise confidence bands for the wind power production using the approximate MLEs for Model 2* $(\theta_0, \alpha, \delta) = (2.22, 0.044, 0.054)$. *Blue line: real production.*

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

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, 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.
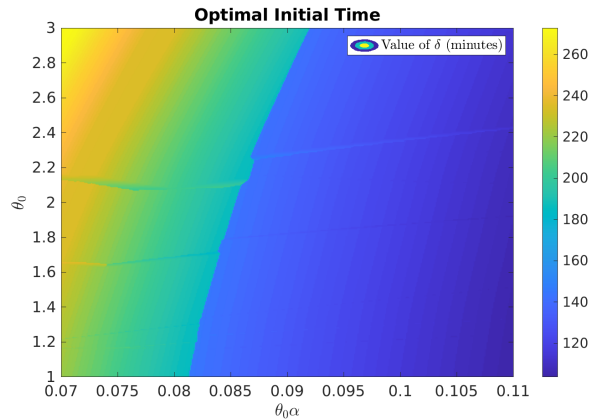

FIG 12. *Initial value for* δ *as a function of the elements of the parameter vector* $\boldsymbol{\theta}$.

7. Conclusions.   In this work, we propose a methodology to assess the short-term forecast of the normalized wind power, which is agnostic of the wind power forecasting technology.

To this purpose, we built a phenomenological stochastic differential equation model for the normalized wind power production forecast error, with time-varying mean-reversion parameter in the linear drift coefficient, and state-dependent and time non-homogenous diffusion coefficient. We also used the Lamperti transform with unknown parameters to provide a version of the proposed model with a unit diffusion coefficient, increasing its stability properties.

We used approximate likelihood-based methods for the models' calibration, and developed original algorithms to derive optimal estimates of the unknown parameters. The likelihood approach allowed for the extending of the SDE models in a very effective way, incorporating an early transition with an additional parameter that accounts for the forecast's uncertainty at the beginning of each future period. As a result, 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.

On the basis of historical data of wind power production and forecast from different sources, our method came up with an objective tool for forecast assessment and comparison by performing the model selection stage. The application of the modeling procedure, inference through numerical optimization, and model selection through information criteria, to the wind power production dataset in Uruguay between April and December 2019, with three different providers, shows the excellent performance of our proposed model, which preserves the asymmetry of wind power forecast errors and their correlation structure.

We conclude that our SDE model, featuring a time-derivative tracking of the forecast, a time-dependent mean-reversion parameter, and a state-dependent diffusion term that suitably adjusts to the problem under study, contributes toward the management of renewable energies efficiently. This methodology paves the way for stochastic optimal control methods enabling principled decision making under uncertainty in the presence of complex energy matrices.

APPENDIX

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

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

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 Alfonsi (2015) and Karatzas and Shreve (1998) that we adapted to the setting of our model (30).

THEOREM A.1.   *Assume that*

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

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 $t \in [0, T]$

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

According to Proposition 2.13, p.291 of Karatzas and Shreve (1998), under assumption (A) there is a unique strong solution X_t to (31). Moreover, as the diffusion coefficient is of linear growth, we have for all $p > 0$

$$(32) \qquad\qquad \mathbb{E}\left[\sup_{t \in [0,T]} |X_t|^p \right] < \infty.$$

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:

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

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

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

$$\psi_n(X_t) = \psi_n(x_0) + \int_0^t \psi'_n(X_s)(\dot{p}_s + \theta_s p_s - \theta_s X_s) \, \mathrm{d}s$$

$$+ \int_0^t \psi'_n(X_s)\sqrt{2\alpha\theta_0 |X_s(1 - X_s)|} \, \mathrm{d}W_s + \alpha\theta_0 \int_0^t g_n(|X_s|)|X_s(1 - X_s)| \, \mathrm{d}s.$$

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

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

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

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

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

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

and thus we have

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

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

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

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

THEOREM A.2.   *Assume that assumptions of Theorem A.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*

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

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

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

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

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

By way of contradiction let us assume that $\{\tau_0 < \infty\}$, then letting $t \to \tau_0$ we deduce that $\lim_{t \to \infty} \mathbf{1}_{\{\tau_0 < \infty\}} M_{t \wedge \tau_0} = -\mathbf{1}_{\{\tau_0 < \infty\}}\infty$ 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

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

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

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 in Section 4 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.

SUPPLEMENTARY MATERIAL

Supplement: DATA AND SIMULATIONS

(doi: COMPLETED BY THE TYPESETTER; .pdf). Supplementary Material contains the entire dataset with normalized wind-power production, the simulation-based results, and the complete set of empirical pointwise confidence bands for wind power production – all corresponding to the forecast provider A.

REFERENCES

AÏT-SAHALIA, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica* **70** 223–262.

ALFONSI, A. (2015). *Affine Diffusions and Related Processes: Simulation, Theory and Applications*. *Bocconi & Springer Series* **6**. Springer.

BADOSA, J., GOBET, E., GRANGEREAU, M. and KIM, D. (2018). Day-Ahead Probabilistic Forecast of Solar Irradiance: A Stochastic Differential Equation Approach. In *Renewable Energy: Forecasting and Risk Management* (P. DROBINSKI, M. MOUGEOT, D. PICARD, R. PLOUGONVEN and P. TANKOV, eds.) 73–93. Springer International Publishing, Cham.

CHANG, W.-Y. (2014). A Literature Review of Wind Forecasting Methods. *Journal of Power and Energy Engineering* **2** 161–168.

D'ONOFRIO, G., TAMBORRINO, M. and LANSKY, P. (2018). The Jacobi diffusion process as a neuronal model. *Chaos* **28**.

EGOROV, A. V., LI, H. and XU, Y. (2003). Maximum likelihood estimation of time-inhomogeneous diffusions. *Journal of Econometrics* **114** 107–139.

ELKANTASSI, S., KALLIGIANNAKI, E. and TEMPONE, R. (2017). Inference and Sensitivity in Stochastic Wind Power Forecast Models. In *2nd ECCOMAS Thematic Conference on Uncertainty Quantification in Computational Sciences and Engineering* (M. Papadrakakis, V. Papadopoulos and G. Stefanou, eds.) 381–393. Eccomas Proceedia UNCECOMP 2017.

FORMAN, J. L. and SORENSEN, M. (2008). The Pearson Diffusions: A Class of Statistically Tractable Diffusion Processes. *Scandinavian Journal of Statistics* **35** 438–465.

GIEBEL, G., BROWNSWORD, R., KARINIOTAKIS, G. N., DENHARD, M. and DRAXL, C. (2011). The State of the Art in Short-Term Prediction of Wind Power. A Literature Overview Technical Report No. ANEMOS.plus and SafeWind projects.

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

IRENA (2018). *Uruguay Power System Flexibility assessment: IRENA FlexTool Case Study*.

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

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

IVERSEN, E. B., MORALES, J. M., MØLLER, J. K. and MADSEN, H. (2014). Probabilistic forecasts of solar irradiance using stochastic differential equations. *Environmetrics* **25** 152–164.

KARATZAS, I. and SHREVE, S. E. (1998). Brownian Motion. In *Brownian Motion and Stochastic Calculus*. second edition ed. *Graduate Texts in Mathematics* **113** 47–127. Springer.

LAMPERTI, J. (1964). A simple construction of certain diffusion processes. *J. Math. Kyoto Univ.* **4** 161–170.

LEONENKO, G. M. and PHILLIPS, T. N. (2012). High-order approximation of Pearson diffusions processes. *Journal of Computational and Applied Mathematics* **236** 2853–2868.

MØLLER, J. K. and MADSEN, H. (2010). From State Dependent Diffusion to Constant Diffusion in Stochastic Differential Equations by the Lamperti Transform Technical Report No. IMM-Technical Report-2010-16, Technical University of Denmark, DTU Informatics, Building 321, Kgs. Lyngby, Denmark.

MØLLER, J. K., ZUGNO, M. and MADSEN, H. (2016). Probabilistic Forecasts of Wind Power Generation by Stochastic Differential Equation Models. *Journal of Forecasting* **35** 189–205.

PANIK, M. J. (2017). *Stochastic differential equations : an introduction with applications in population dynamics modeling*. John Wiley & Sons, Inc., Hoboken, NJ.

PRESTON, S. P. and WOOD, A. T. A. (2012). Approximation of transition densities of stochastic differential equations by saddlepoint methods applied to small-time Ito–Taylor sample-path expansions. *Statistics and Computing* **22** 205–217.

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

SÄRKKÄ, S. and SOLIN, A. (2019). *Applied Stochastic Differential Equations*. Cambridge University Press.

SCHICKER, I., PAPAZEK, P., KANN, A. and WANG, Y. (2017). Short-range wind speed predictions for complex terrain using an interval-artificial neural network. *Energy Procedia* **125** 199–206.

SHOJI, I. and OZAKI, T. (1998). Estimation for nonlinear stochastic differential equations by a local linearization method. *Stochastic Analysis and Applications* **16** 733–752.

SØRENSEN, M. (2012). Estimating functions for diffusion-type processes. In *Statistical Methods for Stochastic Differential Equations. Monographs on Statistics and Applied Probability* **124** 1–107. Chapman & Hall/CRC.

SULTANA, W. R., SAHOO, S. K., SUKCHAI, S., YAMUNA, S. and VENKATESH, D. (2017). A review on state of art development of model predictive control for renewable energy applications. *Renewable and Sustainable Energy Reviews* **76** 391–406.

TSITSIKLIS, J. N. and XU, Y. (2015). Pricing of fluctuations in electricity markets. *European Journal of Operational Research* **246** 199–208.

VALÉRY, P. and GOURIÉROUX, C. (2011). A quasi-likelihood approach based on eigenfunctions for a bounded-valued Jacobi process (Working Paper). Available at https://www.researchgate.net/publication/251252253.

IEA Press release. https://www.iea.org/news/global-energy-demand-to-plunge-this-year-as-a-result-of-the-biggest-shock-since-the-second-world-war. Accessed: 2020-05-30.

ZHOU, Z., BOTTERUD, A., WANG, J., BESSA, R. J., KEKO, H., SUMAILI, J. and MIRANDA, V. (2013). Application of probabilistic wind power forecasting in electricity markets. *Wind Energy* **16** 321–338.