Wind Power Prediction Stochastic Model

Khaoula Ben Chaabane

Supervised by

Pr. Raul Tempone
Dr. Nadhir Ben Rached

Vis-à-vis

Dr. Yousra Gati

4 June 2020

Khaoula Ben Chaabane
4 June 2020

Acknowledgements

I would like to express my deepest appreciation towards all the people who have contributed to the achievement of this internship.

I would like to express my deepest gratitude to Prof. Raul Tempone for taking a chance and accepting me to work in this group.

I would also like to express my sincere thanks to Dr. Nadhir Ben Rached, who has recommended me and helped me during this internship.

Finally, I would like to express my deepest thanks to Renzo Caballero Rosas for his help and patience during this project.

List of Figures

List of Tables

Contents

Abstract

Having access to a reliable wind power generation forecast is a necessity for several industry applications: the optimal control of electricity costs, creating an optimal energy mix and the good management of ressources overall.
Through this work, we propose a stochastic prediction approach based on parametric stochastic differential equations modelling the forecast error. We use an approximation of the maximum likelihood method to infer the model parameters. We applied our model to Uruguayan wind power production as determined by historical data and corresponding numerical forecasts from three official providers for the period ranging from April 24 to December 31, 2019.

1 Introduction

With the increase of global warming and its harm on the environment as well as the inevitable shortage of fossil fuels, renewable energy sources like wind and solar, represent a realistic alternative for a clean energy transition and a necessity to be added in the energy mix. However, for an effective management of these ressources, it is crucial to have a reliable power generation forecast. These predictions are necessary for different situations:

- Optimizing the wind farms and solar plant's operations and energy trading.
- Controlling electricity costs in the market.
- Allocating the portion of each power source in the energy mix.

For this work, we focused on wind power generation forecasting. Several methods have been developed for this purpose and can be divided into four categories: physical models, statistical models, artificial intelligence approaches and hybrid methods. All these methods lead to a deterministic result.

For our approach, we will focus on stochastic forecast models. We propose to model wind power forecasts errors using parametric stochastic differential equations (SDEs). Their resultant stochastic process describes the time evolution of wind power forecast errors from which we will be able to derive a prediction for the wind power generation. As inputs of our model, we will be using historical power production as well as an available deterministic forecast provided by official sources.

We will be working on data from Uruguay as the country has been a global leader in the renewable energy transition. The International Energy Agency (IEA) announced in 2019 that the country is in fourth place globally, producing 36% of its electricity from wind and solar energy.

In the following section, we will describe the available datasets and the processing procedure as well as present some statistical analysis of the data.

In Section 3, we will present basic notions related to stochastic differential equations. We will then propose and explain our parametric model and proceed to parameter estimation in sections 4, 5 and 6. Finally, we will expose and discuss our results before concluding.

2 Wind Data Processing

2.1 Description of the data

We were provided data by three different official electricity providers from Uruguay. We will denote them by **Provider A, Provider B and Provider C**.
The data corresponds to the real production and the forecast data in (MW) for the year 2019 of the wind power in Uruguay.
All three providers have the same data structure for both the real production and the forecast.
The first step in processing the data was eliminating the days showing curtailment or errors in measurements. In the energy industry, curtailment means the act of reducing or restricting energy delivery from a generator to the electrical grid. It can be happen for different reasons (Surplus in production, system emergency, maintenance). However, curtailment distort the data and gives of false representation of the real production. Thus the need to eliminate days showing instances of curtailment.
From a total of 365, only 147 were useable.

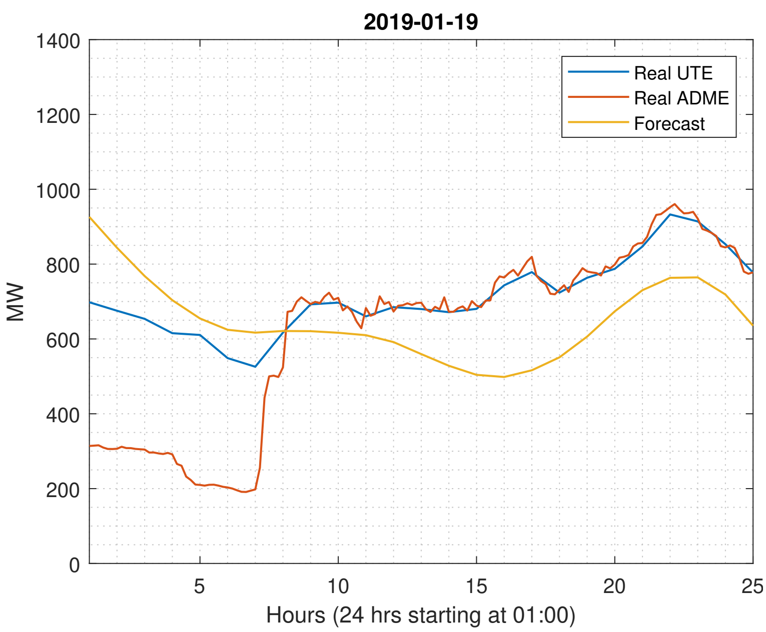


Figure 1: Example of a day with curtailment

As we can see from figure (1), this day shows signs of curtailment at the beginning. The data "Real_ADME" represents the raw real production and "Real_UTE" is the real production with a tentative correction of the curtailment.

Each dataset starts from 24/04/2019 and ends on 31/12/2019, making it a total of 147 days from every provider. For everyday, the measurements start at 00:00 and end at 24:00 on a 24 hours time scale. We have a measurement every 10 minutes, which means a total of 145 values each day.

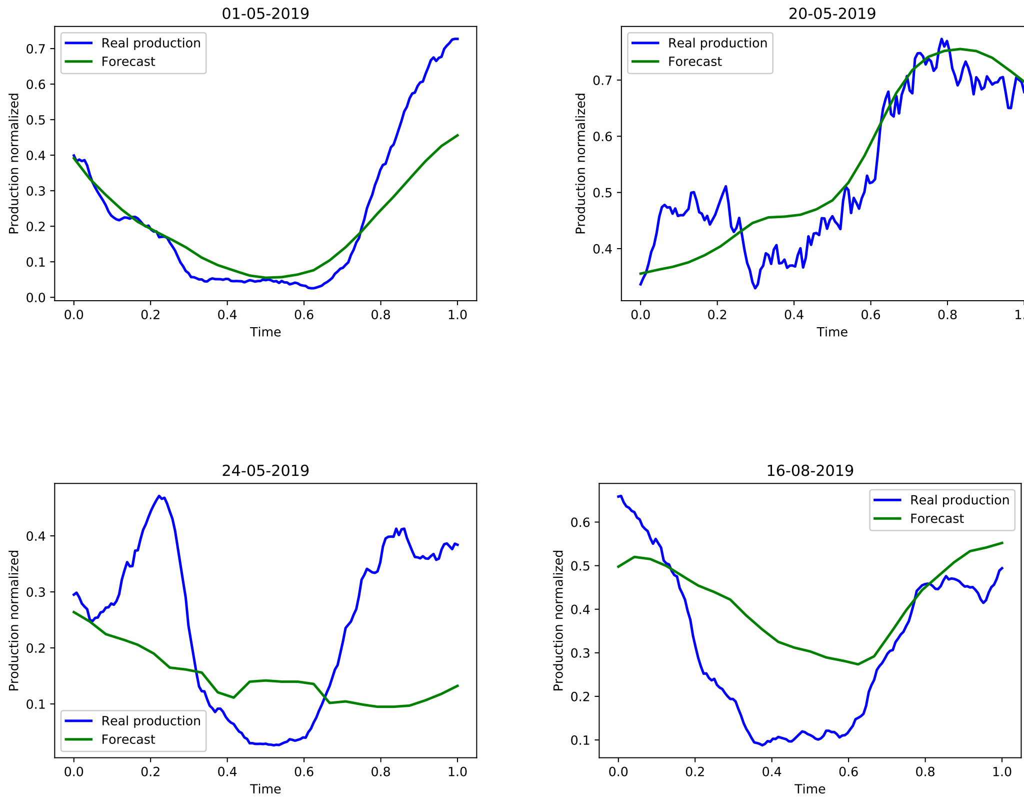


Figure 2: Real production vs forecast

These plots show the real production and the forecast normalized for some of the days. For the normalization we divided by the maximum production 1 476.5 MW. We can see that for some days the forecast corresponds to the real production as shown in figures (2) and (3), however, for some other days as show in figures (5) and (6), the transition error is quite important.

2.2 Data analysis

The variable of interest for our model is the transition error$V$ defined as the difference between the real production and the forecast: $V = X - p$ where X is the real normalized production and p is the normalized forecast.
In this section we will focus on the analysis of V.

2.2.1 Mean Square Error

To guarantee an homogeneous year, we study the Mean Absolute Error (MAE) for the forecast for each day and each hour.
Let V_{ij} $j \in \{0..146\}$ the number of the day and $i \in \{1...145\}$ the time, be the error. The MAE is defined as follows for both cases:

- **Daily MAE:** To see the error throughout the year. Using the formula:

$$\hat{V_j} = \frac{1}{145} \sum_{i=1}^{145} |V_{ij}| \quad \forall j \in \{0...146\}$$

- **Hourly MAE:** To see the error throughout the day. Using the formula:

$$\bar{V_i} = \frac{1}{147} \sum_{j=0}^{146} |V_{ij}| \quad \forall i \in \{1...145\}$$

We can see from figures (3), (4) and (5) that the error is important in the beginning of the day and as we advance in time and near the late afternoon, the forecast becomes less accurate.

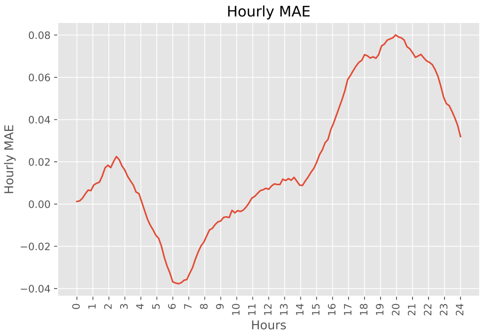


Figure 3: Hourly MAE for provider A

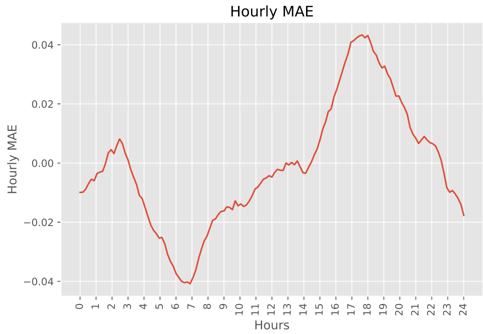


Figure 4: Hourly MAE for provider B

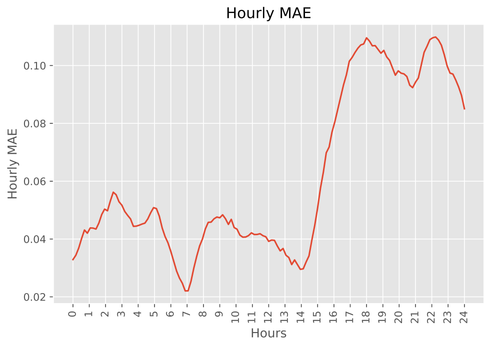


Figure 5: Hourly MAE for provider C

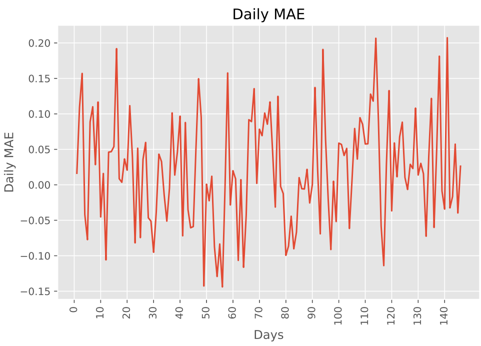


Figure 6: Daily MAE for provider A

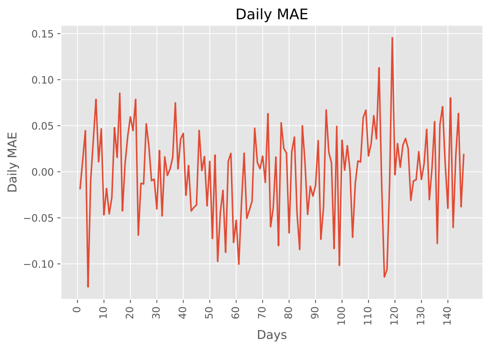


Figure 7: Daily MAE for provider B

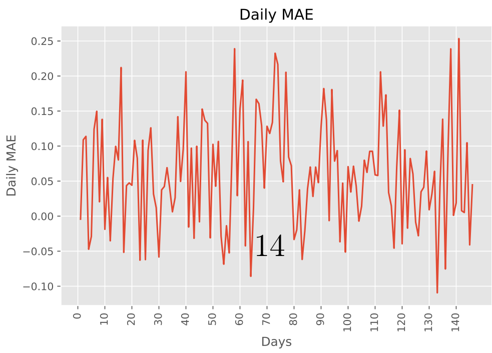


Figure 8: Daily MAE for provider C

From figures (6), (7) and (8) we can see that there are no big variations along the year for the three datasets.

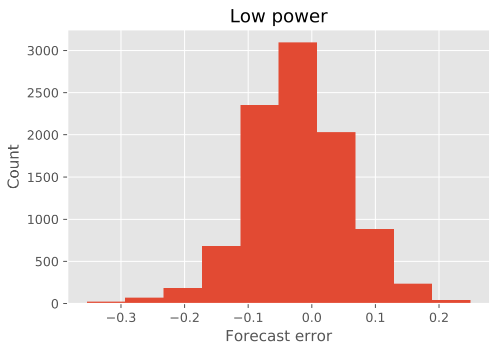
2.3 Power range

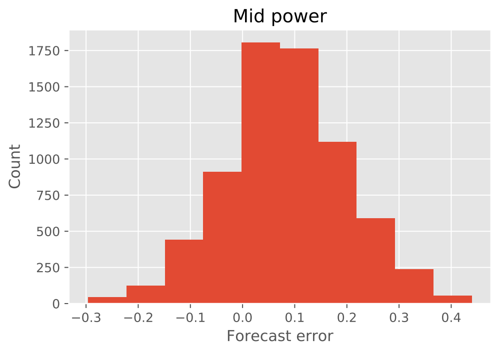
In this section, we will focus on analysing the transition error based on the distribution of the real normalized production into different power ranges.
We define three power range:

- **Low power range:** when the real production is in the interval [0,0.3[.
- **Mid power range:** when the real production is in the interval [0.3,0.6[.
- **High power range:** when the real production is in the interval [0.6,1[.

We plotted the different power ranges for every provider.

- Provider A:





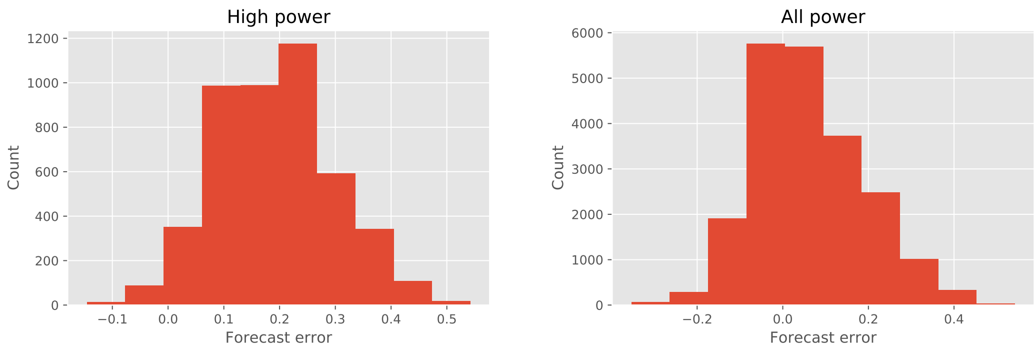


Figure 9: Power range for provider A

- Provider B:

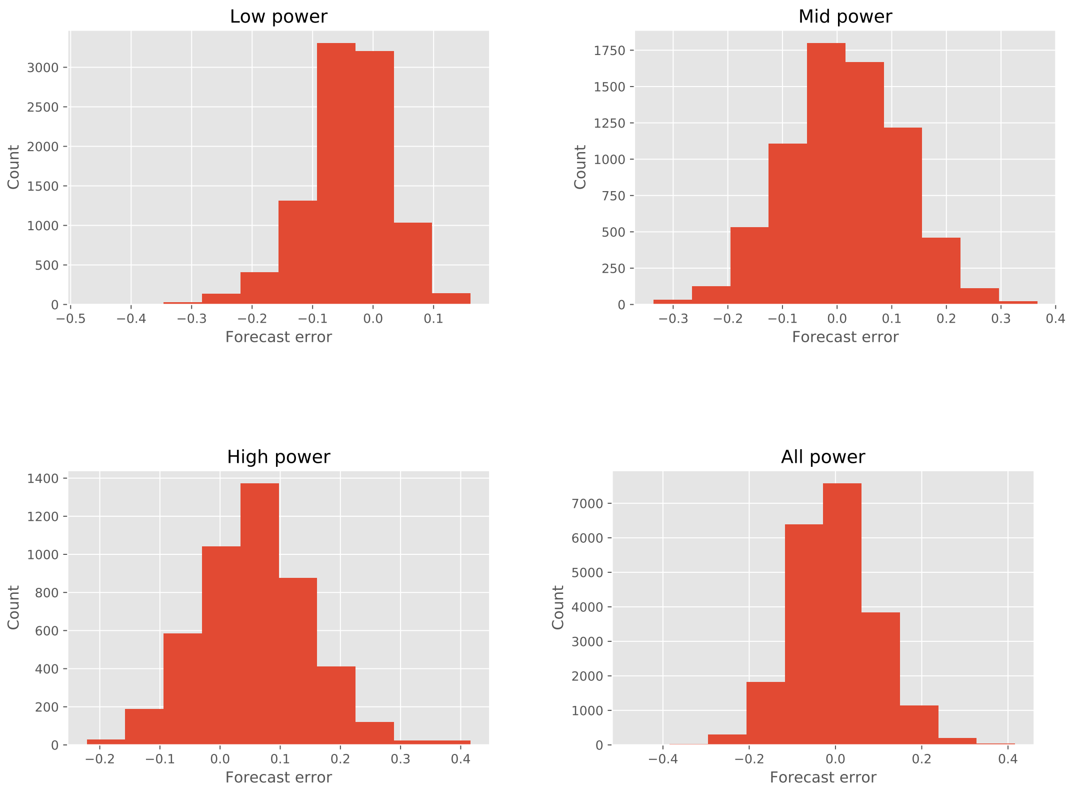


Figure 10: Power range for provider B

- Provider C:

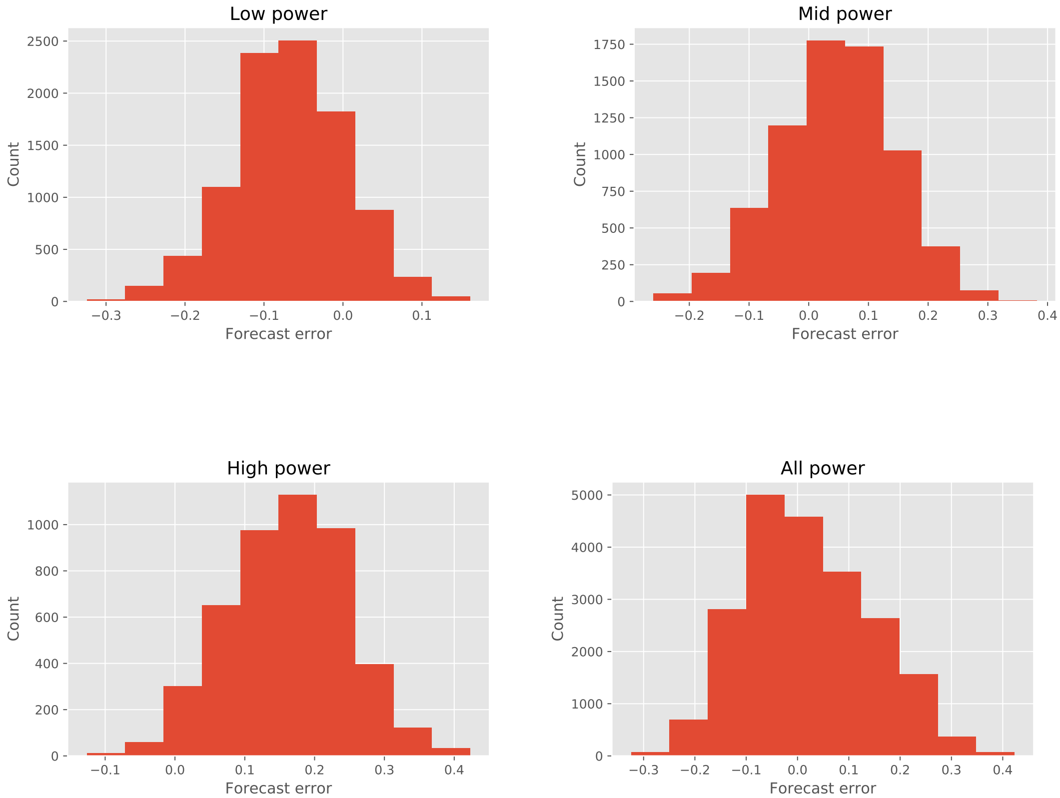


Figure 11: Power range for provider A

From these figures, we can see the repartition of the data in the different power ranges. For the three providers, the majority of the data belongs to the low power range.

2.3.1 Error over forecast

In this section we plotted the transition error over the forecast for the three different datasets.

We divided our data into a training set that we will use to determine the model's parameters and a testing set used to test our model. To do this, we separated the

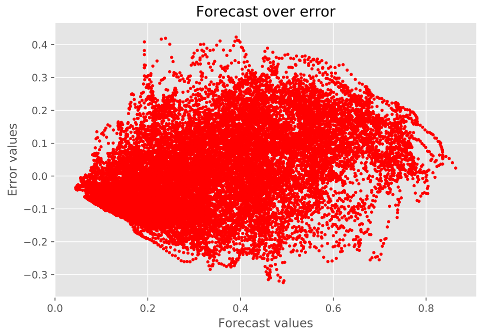


Figure 12: Error over forecast provider A

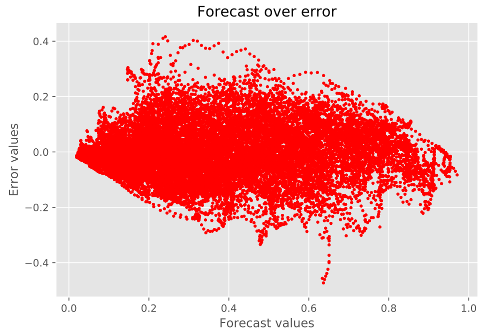


Figure 13: Error over forecast for provider B

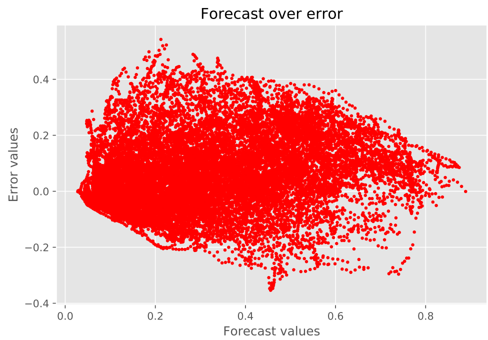


Figure 14: Error over forecast for provider C

complete data for each provider in a way to never have two consecutive days in the same set.

3 Basic notions and literature review

3.1 Introduction

Before getting into the details of our model, we should introduce the basic mathematical concepts necessary for our work. In the first section, we will introduce Stochastic Differential Equations and the theorems used for our model. The second section will be dedicated to explaining statistical inference and more particularly the Maximum Likelihood approach and its approximation.

3.2 Stochastic Differential Equation (SDE)

Lets consider a stochastic process, $X(t)_{t\in[0,T]}$ taking value in $\mathcal{A}$, that satisfies the following Itô SDE:

$$\begin{cases} dX(t) = a(X(t), t; \boldsymbol{\theta})dt + b(X(t), t; \boldsymbol{\theta})dW \\ X(0) = x_0 \end{cases} \tag{1}$$

Where W is a Brownian motion also known as a Wiener process. This random process adds uncertainty to the equation and differentiate the SDE from an ODE. $a(X(t), t; \boldsymbol{\theta})$ is known as the drift term and $b(X(t), t; \boldsymbol{\theta})$ as the diffusion term.

Equation (1) can also be written in an integral form as follows:

$$X(t) = x_0 + \int_0^t a(X(u), u; \boldsymbol{\theta})du + \int_0^t b(X(u), u; \boldsymbol{\theta})dW(u)$$

When $X(t)$ is a Markovian process, its transition probability density is:

$$p(t, y|s, x; \boldsymbol{\theta}) = P(X(t) = y|X(s) = x; \boldsymbol{\theta}), t \in [0, T]$$

This probability density satisfies the following Fokker-Planck equation:

$$\begin{cases} \frac{\partial p(y,t|x,s;\theta)}{\partial t} = -\frac{\partial}{\partial y}(a(y,t)p(y,t|x,s;\boldsymbol{\theta})) + \frac{1}{2}\frac{\partial^2}{\partial^2 y}(b(y,t;\boldsymbol{\theta})p(y,t|x,s;\boldsymbol{\theta})) \\ p(0,y) = g_0(y) \end{cases} \tag{2}$$

such that $s, t \in [0, T], x, y \in \mathcal{A}$, where $\mathcal{A}$ is the space domain, $g_0(.)$ is the deterministic function that defines the initial condition for the Fokker-Planck equation. When the Markov process starts at $X(0) = x_0$ (deterministic), the initial condition is $p(0, y) = \delta(x - y)$.

3.3 Maximum Likelihood Inference for discretely observed diffusion processes

Given a discrete sample set $\mathbf{x}_N = \{X(t_i)\}_{i=1}^N$ associated to the process $X(t)_{t\in[0,T]}$ solution of the SDE (1), we can infer the model parameters $\boldsymbol{\theta}$ by maximizing the (exact) likelihood function

$$L(\boldsymbol{\theta}; \mathbf{x}_N) = p(\mathbf{x}_N; \boldsymbol{\theta})$$

where $p(X(t_1), \ldots, X(t_N); \boldsymbol{\theta})$ denotes the finite-dimensional joint density of the sample $X(t_1), \ldots, X(t_N)$. Since $X(t)$ is a Markov process we can rewrite the likelihood function

$$L(\boldsymbol{\theta}; \mathbf{x}_N)) = p_1(x_0) \prod_{i=1}^{N} p_i(x_i, t_i | x_{i-1}, t_{i-1}; \boldsymbol{\theta})$$

in terms of the transition probabilities p_i.
The data sets for this project are of the form

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

where M represents the number of paths and (N+1) the time discretization points, equidistant with a given interval of length Δ. $X_{t_j}^{N+1} = \{X_{t,+i\Delta}, i = 0, \ldots, N\}$, we define $X_{t,+i\Delta} \equiv X_{j,i}\ \forall j \in \{1, \ldots, M\}$.

The family of stochastic processes $X_{j,i}$ satisfies different SDEs which depend on the same parameter set $\boldsymbol{\theta}$.
Let $\rho\left(X_{j,i}|X_{j,i-1};\theta\right)$ be the conditional probability density of $X_{j,i}$ given $X_{j,i-1}$
For this case, assuming that $\{X_{j,i}\}_{j=1}^{M}$ are independent, the exact likelihood function is the following:

$$L\left(\boldsymbol{\theta};\mathbf{x}_{MN}\right)=\prod_{j=1}^{M}\prod_{i=1}^{N}\rho\left(X_{j,i}|X_{j,i-1};\boldsymbol{\theta}\right)$$

and the log-likelihood is given by:

$$\mathcal{L}\left(\boldsymbol{\theta};\mathbf{x}_{MN}\right)=\sum_{j=1}^{M}\sum_{i=1}^{N}log\left(\rho\left(X_{j,i}|X_{j,i-1};\boldsymbol{\theta}\right)\right) \tag{3}$$

Thus, given the data $X^{M,N+1}$, the maximum likelihood estimates of the parameters $\boldsymbol{\theta}$ are obtained by maximizing the function (3).

However, the transition probabilities $\rho\left(X_{j,i}|X_{j,i-1};\boldsymbol{\theta}\right)$ may not be easy to be determined in closed-form. One approach is to solve numerically the corresponding Fokker-Planck equations, but this can be not always possible and is computationally costly. Thus, we use some scheme approximation to approximate the transition density and compute the likelihood function.

3.4 Transition density approximation and moment matching

To approximate the transition density of X we will suppose that it follows a proxy distribution whose density is known and we will seek to find its parameters. To do so, we will assume the equality of the first and second moments of X deduced from the SDE (1) with the ones of the proxy: This is what we call moment matching.

3.4.1 Two-moment equations

We denote by $\mu(t) = E[X(t)]$ and $\sigma(t) = E\left[(X(t) - \mu(t))^2\right]$ the first and second moments of $X(t)$ respectively, $t \in [0, T]$.

The first moment equation comes directly from the application of the expectation to the equation (1), which directly gives us that:

$$\frac{d\mu(t)}{dt} = E[a(X(t))]$$

If the drift $b(X(t))$, is linear then

$$\frac{d\mu(t)}{dt} = a(E[X(t)]) = a(\mu(t)).$$

The second moment equation is obtained by applying Itô's formula (see appendix) to the process $Y(t)$, defined as $Y(t) = g(X(t)) = X(t)^2$. Then, $g_t = 0; g'(X(t)) = 2X(t); g''(X(t)) = 2$, and

$$\begin{aligned} dY(t) &= 2X(t)dX(t) + \frac{1}{2}2b^2(X(t))dt \\ &= 2X(t)\left(a(X(t))dt + b(X(t))dW_t\right) + \frac{1}{2}2b^2(X(t))dt \\ &= 2\left(X(t)a(X(t)) + \frac{1}{2}b^2(X(t))\right)dt + 2X(t)b(X(t))dW_t \end{aligned}$$

Applying the expectation on both sides of equation, we obtain:

$$dE[Y(t)] = dE[X(t)^2] = 2E\left[X(t)a(X(t)) + \frac{1}{2}b^2(Xt))\right]dt$$

Subtracting the derivative of the squared mean gives:

$$\begin{aligned} d\sigma^2(t) &= dE\left[Y(t)\right] - d\left(\mu^2(t)\right) \\ &= dE\left[X(t)^2\right] - 2\mu(t)d\mu(t) \\ &= 2E\left[X(t)a(X(t)) + \frac{1}{2}b^2(X(t))\right]dt - \mu(t)b(\mu(t))dt \end{aligned}$$

Thus, the approximate equation of the variance is given by

$$d\sigma^2(t) = 2E\left[X(t)a(X(t)) + \frac{1}{2}b^2(X(t))\right]dt - \mu(t)b(\mu(t))dt$$

3.5 Moment matching

Let $Z(t)$ be the selected proxy random variable to approximate $X(t)$ $\forall t \in [0,T]$ with a known probability density function f_Z. We approximate $X(t) = a + (a - b)Z(t)$ with support $[a,b]$ with a probability density function f_X. We have:

$$\begin{cases} \mu(t) = a + (b-a)E[Z(t)] \\ \sigma^2(t) = (b-a)^2 V[Z(t)] \end{cases}$$

and
$f_X(X) = f_Z(g^{-1}(X)) \left|\frac{\mathrm{d}}{\mathrm{d}X} g^{-1}(X)\right|$ with $g(x) = a + (b-a)x$ and $g^{-1}(x) = \frac{x-a}{b-a}$.

4 The proposed model

4.1 Construction of the model

In this section we introduce a stochastic process, $X(t) \in [0,1]$ that represents the normalized wind power as well as the normalized deterministic forecast, $p(t) \in [0,1]$, provided by an official source, for $t \in [0,T]$. We consider that the stochastic process, $X(t)$, is the solution of the following parametrized Ito stochastic differential equation:

$$\begin{cases} dX_t = a\left(X_t; p_t, \dot{p}_t, \theta\right) dt + b\left(X_t; p_t, \dot{p}_t, \theta\right) dW_t, \quad t \in [0, T] \\ X_0 = x_0 \in [0, 1] \end{cases} \tag{4}$$

where:

- $a\left(0, \dot{p}_t, \theta\right) : [0, 1] \to R$ denotes a drift function,
- $b\left(\cdot; p_t, \dot{p}_t, \theta\right) : [0, 1] \to R_+$ a diffusion function,
- θ is a vector of unknown parameters
- $\left(p_t\right)_{\mathrm{t}\in[0,T]}$ is a time-dependent deterministic function [0,1]-valued and $\left(\dot{p}_t\right)_{t\in[0,T]}$ is its time derivative
- $\{W_t, t \in [0, T]\}$ is a standard real-valued Wiener process.

4.1.1 Drift term

Let $\left(p_t\right)_{t\in[0,T]}$ be the available prediction function for the normalized wind power, which is an input to this model. As we expect the stochastic process, X(t), to align with the numerical forecast p(t), we choose the drift to be a linear polynomial of the state with a derivative tracking:

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

where $\left(\theta_t\right)_{t\in[0,T]}$ is a positive deterministic function.
With this drift, we can check that, $E(X_t) = p_t$ given $E(X_0) = p_0$.
The application of Itô's formula (see appendix) to the process $g\left(X_t, t\right) = X_t e^{\int_0^t \theta_s, d_s}$, gives us:

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

whose integral form is

$$e^{\int_0^t \theta_t dt} X_t = X_0 + \int_0^t \left(\dot{p}_s + \theta_s p_s\right) e^{\int_0^s \theta_u du} ds + \int_0^t b\left(X_s; p_s, \dot{p}_s, \theta\right) e^{\int_0^t \theta_u du} dW_s$$

Taking expectation we obtain

$$
\begin{aligned}
E\left(X_t\right) &= e^{-\int_0^t \theta_t dt}\left[p0 + \int_0^t \left(\dot{p}_s + \theta_s p_s\right) e^{\int_0^s \theta_u du} ds\right] \\
&= e^{-\int_0^t \theta_t dt}\left[p_0 + \int_0^t \dot{p}_s e^{\int_0^s \theta_u du} ds + \left[p_s e^{\int_0^t \theta_u du}\right]_0^t - \int_0^t \dot{p}_s e^{\int_0^s \theta_u du} ds\right] \\
&= e^{-\int_0^t \theta_t dt}\left[p_0 + e^{\int_0^t \theta_t dt} p_t - p_0\right] \\
&= p_t
\end{aligned}
$$

Therefore, the process X, modeled as solution to the SDE (9) using this drift, reverts back to its mean p_t with a time-varying speed θ_t that is proportional to its deviation from the mean. This drift term also allows it tracks the time derivative $\dot{p}_t$.

4.1.2 Diffusion term

Let $\boldsymbol{\theta} = (\theta_0, \alpha)$, we chose a state-dependent diffusion term that avoids the process exiting outside the range [0,1]:

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

where $\alpha > 0$ is a unknown parameter that controls the path variability. This diffusion term belongs to the Pearson diffusion family and, in particular, it defines a Jacobi type diffusion.

Therefore, the stochastic process, $X(t) \in [0, 1]$ that represents the normalized wind power is solution the following SDE:

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

A property of this model is that the drift term contains the time-varying parameter θ_t, rendering the solution X to (10) a non-stationary and time-inhomogeneous process. To ensure that the process X_t be the unique solution to (10) for all

$t \in [0, T]$ with state space [0,1], θ_t must verify the following property:

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

The proof of this theoretical statement is presented in the appendix.
From this condition we can see that the parameter θ_t becomes unbounded when $p_t = 0$ or $p_t = 1$. To avoid this, we will consider the following truncated prediction function:

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

The steps to find the parameter ϵ will be explained later on.
From now on, we will keep the notation p_t to denote the truncated prediction function p_t^{ϵ} unless specified otherwise.

4.2 Forecast error model

Having specified a model for the normalized production X, we can define a model for the forecast error $V = (V(t))_{t\in[0,T]}$, $V_t = X_t - p_t$ $\forall t \in [0, T]$ with a change of variable. We find:

$$\begin{cases} \mathrm{d}V_t = -\theta_t V_t \mathrm{d}t + \sqrt{2\alpha\theta_0 \left(V_t + p_t\right)\left(1 - V_t - p_t\right)}\mathrm{d}W_t, & t \in [0, T] \\ V_0 \quad = \mathrm{v}_0 \in [-1 + p_0, 1 - p_0] \end{cases} \tag{6}$$

Our goal will be to estimate the parameter $\boldsymbol{\theta} = (\theta_0, \alpha)$ using the maximum likelihood approach on the likelihood functions related to the equation (6) to then use it to find the normalized production with the equation (5).

5 Likelihood functions and parameter estimations

As discussed in section (3.3), we sample each of our M path continuous-time Itô process $V = (V_t)_{t\in[0,T]}$ (each path representing a day of the year) at $N + 1$

equidistant discrete points with a given length interval Δ.

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

denotes this random sample with $V_{t_j}^{N+1} = \left\{V_{t_j+i\Delta}, i = 0, \ldots, N\right\}, \forall j \in \{1, \ldots, M\}$ and $V_{t_j+i\Delta} \equiv V_{j,i}$.
Let $\rho\left(V_{j,i}|V_{j,i-1}; \theta\right)$ be the conditional probability density of $V_{j,i}$ given $V_{j,i-1}$ where $\theta = (\theta_0, \alpha)$ are the unknown model parameters.

The Itô process V defined by the SDE (6) is Markovian, then the likelihood function of the sample $V^{M,N+1}$ can be written as follows:

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

where $t_{j,i} \equiv t_j + i\Delta$ for any $j \in \{1, \ldots, M\}$ and $i \in \{0, \ldots, N\}$.
As discussed in section (3.4), we will approximate the transition density with a proxy distribution. we need to compute the first two moments.

5.1 First moment equation

For the first moment, since the drift function is linear, we have that for every day $j \in \{1, \ldots, M\}$

$$\begin{cases} \mathrm{d}E(V(t)) = \mathrm{d}\mu(t) = -\mu(t)\theta_t \mathrm{d}s \\ \mu\left(t_{i-1}\right) = V_{j,i-1} \end{cases}$$

$\forall t \in [t_{j,i}, t_{j,i+1}]$ $\forall i \in \{0, \ldots, N\}$.

If $\theta\left(t_{j,i}\right) = \theta\left(t_{j,i+1}\right) = \theta$ then the exact solution is given by :

$$\mu\left(t_{j,i+1}\right) = \mu\left(t_{j,i}\right)\exp\left(-\theta\left(t_{j,i+1} - t_{j,i}\right)\right).$$

otherwise, we approximate the ODE using Forward-Euler (see appendix).

5.2 Second moment equation

For the second moment, as seen in section (3.4) we have that $\forall t \in [t_{j,i}, t_{j,i+1}]$ $\forall j \in \{1, \ldots, M\}$ and $\forall i \in \{0, \ldots, N\}$:

$$dE[V^2(t)] = 2E\left[-\theta_t V^2(t) + 2\alpha\theta_0 (V(t) + p_t)(1 - V(t) - p_t)\right] dt$$

Therefore:

$$\begin{cases} \mathrm{d}E[V^2(t)] = -2E[V^2(t)]\,(\theta(t) + \alpha\theta_0) + 2\alpha\theta_0\mu(t)(1 - 2p(t)) + 2\alpha\theta_0 p(t)(1 - p(t))\mathrm{d}t \\ E[V^2(t_{j,i})] = V^2_{t_{j,i}} \end{cases}$$

This ODE is solved using forward-Euler (see appendix). We finally have that the second moment is $\sigma^2(t) = E(V^2(t)) - \mu^2(t)$ $\forall t \in [t_{j,i}, t_{j,i+1}]$ $\forall j \in \{1, \ldots, M\}$ and $\forall i \in \{0, \ldots, N\}$.

5.3 Moment matching

A Beta distribution $\beta(\xi_1, \xi_2)$ is the most suitable for a proxy because for the family of diffusion term in our SDE (11) (Pearson diffusion), it has been proven to be the best approximation.

We define $\forall t \in [t_{j,i}, t_{j,i+1}]$ $\forall j \in \{1, \ldots, M\}$ and $\forall i \in \{0, \ldots, N\}$ $Z(t) \sim \beta(\xi_1, \xi_2)$ and we have that $V(t) = a + (a - b)Z(t)$ with support $[a, b] = [-1, 1]$. Therefore:

$$\begin{cases} \mu(t) = a + (b - a)E[Z(t)] = a + (b - a)\frac{\xi_1}{\xi_1 + \xi_2} \\ \sigma^2(t) = (b - a)^2 V[Z(t)] = \frac{(b-a)^2\xi_1\xi_2}{(\xi_1+\xi_2)^2(\xi_1+\xi_2+1)} \end{cases}$$

We finally find that:

$$\xi_1 = -\frac{(1+\mu)\left(\mu^2+\sigma^2-1\right)}{2\sigma^2}$$
$$\xi_2 = \frac{(\mu-1)\left(\mu^2+\sigma^2-1\right)}{2\sigma^2}$$

5.4 Log-likelihood approximation

Now that we have determined the proxy's density parameters, we can compute the approximated density probability and the log-likelihood.

5.4.1 Log-density approximation

Having made the approximation $\forall t \in [t_{j,i}, t_{j,i+1}]$ $\forall j \in \{1,\ldots,M\}$ and $\forall i \in \{0,\ldots,N\}$ $V(t) = a + (a-b)Z(t)$.
Then,

$$f_V(v) = \frac{1}{|(b-a)|}\frac{1}{B(\xi_1,\xi_2)}\left(\frac{v-a}{b-a}\right)^{\xi_1-1}\left(1-\frac{v-a}{b-a}\right)^{\xi_2-1},$$

Finally, the log-density of the stochastic random variable V(t) is given by:

$$\log\left(f_V(v)\right) = \log\left(\frac{1}{\beta(\xi_1,\xi_2)}\right) + (\xi_1-1)\log\left(\frac{v-a}{b-a}\right) + (\xi_2-1)\log\left(\frac{b-v}{b-a}\right)$$

$[a,b] = [-1,1]$,

$$\log\left(f_V(v)\right) = \log\left(\frac{1}{\beta(\xi_1,\xi_2)}\right) + (\xi_1-1)\log\left(\frac{v+1}{2}\right) + (\xi_2-1)\log\left(\frac{1-v}{2}\right)$$

5.4.2 Log-likelihood

The log-likelihood as defined in the section (3.3) is given by:

$$\begin{aligned}
\mathcal{L}\left(V^{M,N+1},\boldsymbol{\theta}\right) &= \sum_{j=1}^{M}\sum_{i=1}^{N+1}\log\left[f_V\left(V_{j,i}|V_{j,i-1}\right)\right]\\
&= \sum_{j=1}^{M}\sum_{i=1}^{N}\log\left(\frac{1}{\beta(\xi_1,\xi_2)}\left(\frac{V_{j,i}+1}{2}\right)^{\xi_1-1}\left(\frac{1-V_{j,i}}{2}\right)^{\xi_2-1}\right)
\end{aligned}$$

$$\mathcal{L}\left(V^{M,N+1},\boldsymbol{\theta}\right) = \sum_{j=1}^{M}\sum_{i=0}^{N}\log\left(\frac{1}{\beta(\xi_1,\xi_2)}\left(\frac{V_{j,i}+1}{2}\right)^{\xi_1-1}\left(\frac{1-V_{j,i}}{2}\right)^{\xi_2-1}\right) \tag{7}$$

where and M is the number of paths and (N+1) the number of transitions per path.

6 Initial parameters and ε determination

In order to have a smooth optimization process of our log-likelihood function, we want the initial parameters (θ_0, α) to be close to the optimal solution. Therefore, we will estimate these initial parameters using Least-Square Minimization and Quadratic Variation. This initial estimation will also allow us to compute the parameter ε the we need to truncate the forecast p. We will also introduce another parameter δ that we will use to correct our data.

6.1 Least-square minimization

In order to evaluate the initial parameters of our model we apply the least square method on the forecast error V. We consider for every day j, $j \in \{0, \ldots, M\}$, the transition $\Delta V_{j,i} = V_{j,i+1} - V_{j,i}$ with $\Delta = t_{j,i+1} - t_{j,i}$, $i \in \{1, \ldots, M\}$. $(V_{j,i+1}|V_{j,i})$ is a random variable which conditional mean can be approximated by the solution of the following system :

$$\begin{cases} \mathrm{d}E[V]_j = -\theta_t E[V]_j \mathrm{d}t \\ E\left[V_j\right](t_{j,i}) = V_{j,i} \end{cases}$$

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

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

We take the first order approximation of $e^{-c\Delta t}$ w.r.t. c

$$e^{-c\Delta} = 1 - c\Delta + O\left((c\Delta)^2\right)$$

and introduce it in the previous equation. We get

$$c^* \approx \arg\min_{c\geq 0}\left[\sum_{j=1}^{M}\sum_{i=0}^{N-1}\left(V_{j,i+1}-V_{j,i}(1-c\Delta)\right)^2\right] \tag{8}$$

As equation 8 is convex in c, finding c^* is equivalent to solving:

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

Therefore,

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

Then, c^* satisfies the following:

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

6.2 Quadratic variation

We approximate the SDE by its Euler-Maruyama scheme (see appendix). In particular, we approximate the Itô quadratic variation with the discrete one:

- Itô process quadratic variation : $[V]_t = \int_0^t b(V_s;\theta,p_s)^2\mathrm{d}s$
  where $b(V_s;\theta,p_s) = \sqrt{2\alpha\theta_0\left(V_s+p_t\right)\left(1-V_s-p_s\right)}$
- Discrete process quadratic variation : $[V]_t = \sum\limits_{0<s\leq t}\left(\Delta V_s\right)^2$

Then, considering Δ the time between two consecutive measurements, we approximate:

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

6.3 Determination of ε

If we fix ε, we define the forecast error $\forall j \in \{1...M\}$, $\forall i \in \{1...N\}$ as $V_i = X_i - p_i^{\varepsilon}$. If we also fix θ_0 and α, we can define the set of indexes:

$$\mathbf{K} = \left\{ k \in \{1, \ldots, N\} : \text{ the LSM estimation will estimate } \frac{\theta_0 \alpha}{\varepsilon} \right\}$$

$$\mathbf{I} = \{ i \in \{1, \ldots, N\} : \text{ the LSM estimation will estimate } \theta_0 \}$$

In LSM part, we assumed that $\theta_t = c \in R^+$, and we defined previously (section...) that:

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

$\forall t \in [0, T]$.

From the definition of θ_t, we have can see that for $\varepsilon << 1$, and $p_t = \varepsilon$ or $p_t = 1 - \varepsilon$, $\theta_t \approx \frac{\theta_0 \alpha}{\varepsilon}$ holds.
Then, for ε small enough, K can be approximated by the following:

$$\mathbf{K} \approx \{ k \in \{1, \ldots, N\} : p_k^{\varepsilon} \in \{\varepsilon, 1 - \varepsilon\} \}$$

Besides, we have that $\theta_t^{\varepsilon} = \theta_0$ if $p_t^{\varepsilon} \approx \frac{1}{2}$. Then, we can approximate $\mathbf{I}$ by

$$\mathbf{I} \approx \{ i \in \{1, \ldots, N\} : p_i \in (\gamma, 1 - \gamma) \}, \quad \gamma = \frac{1}{2}$$

Given an approximated value of θ_0 and $\theta_0\alpha$, if we can estimate $\frac{\theta_0 \alpha}{\varepsilon}$, then we can estimate ε.
Let $m := \frac{\theta_0 \alpha}{\varepsilon}$, the goal is to find values for ε that satisfy $\varepsilon << 1$. For that we start by randomly choosing a small initial value for ε (that we will call ε_0), and iterating we aim to converge to a local minimum. We proceed with the following steps:

1. We sample ε_0 from $U[0.01, 0.1]$ and initialize ε with ε_0
2. We create use the set $\mathbf{K}$ to approximate m.

- If $m < \theta_0$, then the assumption $\theta_t = c \in R^+$ is not correct and we reduce the value of ε, by multiplying it with 0.999.
- If $m \geq \theta_0$, then ε takes the value of $\frac{\theta_0^* \alpha^*}{k}$(we allow a maximum relative change of 1%).

We repeat this step 100 times.

3. We repeat steps 1 and 2, 50 times.

6.4 Time-parameter δ

In order to have the property $E(X_t) = p_t$ correct at all times, our SDE (1) needs to have a starting time t_0 where $V_{t_0} = 0$ i.e ($X_{t_0} = p_{t_0}$).

However, we noticed that with the data provided, the initial error $V(0) \neq 0$ for all the paths. To unsure the validity of this property and therefore correct results for our model, we suppose that there existed a time $t_\delta < 0$, such as $t_0 - t_\delta = \delta$, in the past where the condition $X_{t_\delta} = p_{t_\delta}$ is verified. (This can be justified as we assume that the forecast providers had at some point perfect information on the data).

Knowing the optimal parameters (θ_0, α), we will linearly interpolate the forecast p to the time t_δ, and for the transition $V_{t_0}^j | V_{t_\delta}^j$ $\forall j \in \{1, \ldots, M\}$ we will solve the following optimization problem:

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

We will approximate again the density with a β distribution and approximate its parameters using the moment matching method.

7 Results

7.1 Initial parameters and ε

The initial guess for the parameters θ_0 and α using LSM as described in sections 6.5.1 and 6.5.3 gives us the following values :

$$(\theta_0, \alpha) = (1.25, 0.08)$$

We plotted the parameter ε over the number of iterations as we can see on the following graph:

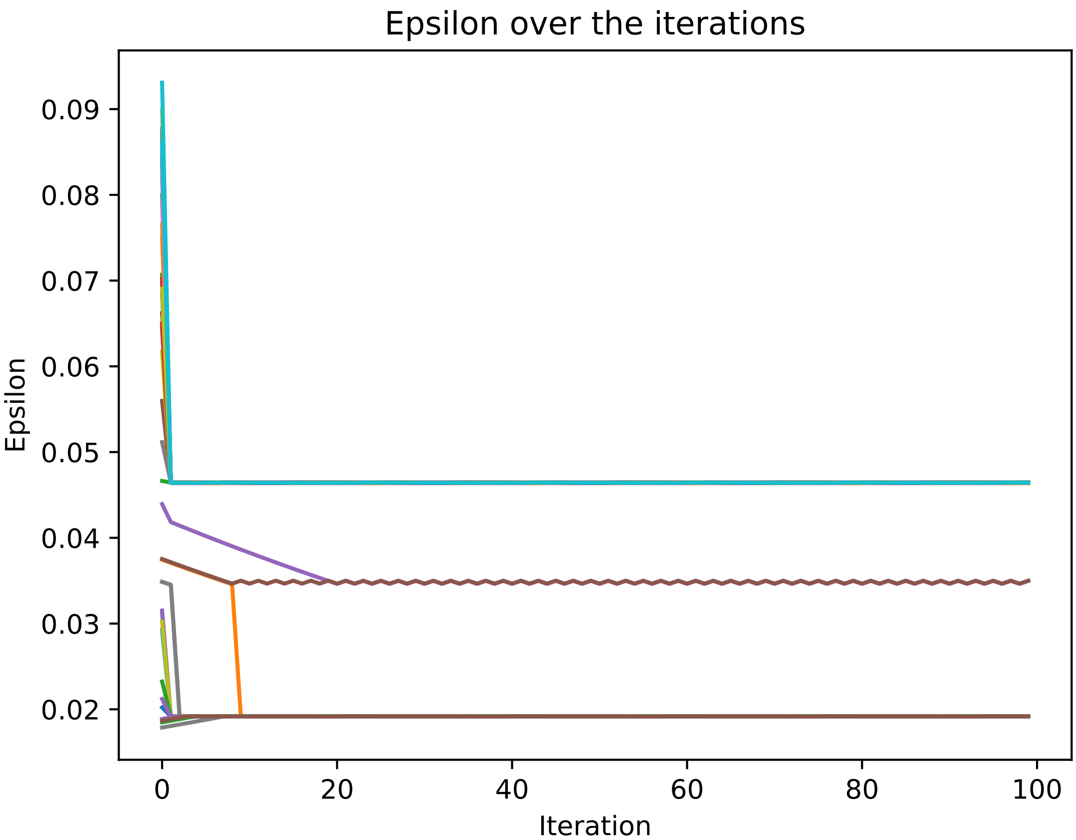


Figure 15: ε over the iterations

As we can see from figure 1, we have 3 possible values for ε, taking the minimum $\varepsilon = 0.018$ is a good approximation.

7.2 Likelihood optimization

We minimized the negative log-likelihood (7) using $(\theta_0, \alpha) = (1.25, 0.08)$ as initial parameters and found the following results for the different data providers:

Data providers	θ_0	α	$\theta_0\alpha$	δ
Provider A	1.357	0.0809	0.108	0.0821
Provider B	1.175	0.0856	0.100	0.0827
Provider C	1.196	0.0846	0.101	0.0827

Table 1: Optimal parameters for the different providers

We can see a slight difference in the optimal parameters between the 3 providers. However the product $\theta_0\alpha$ is almost the same for each one.

7.3 Path simulations

After computing the optimal parameters $(\theta_0, \alpha, \delta)$ we can use them in the equation (1) to simulate the real production that we find using our model.

The following plots represent different simulations of the real normalized production and the provided forecast for different days of the year for each provider. We plotted five paths per day.

- Provider A:

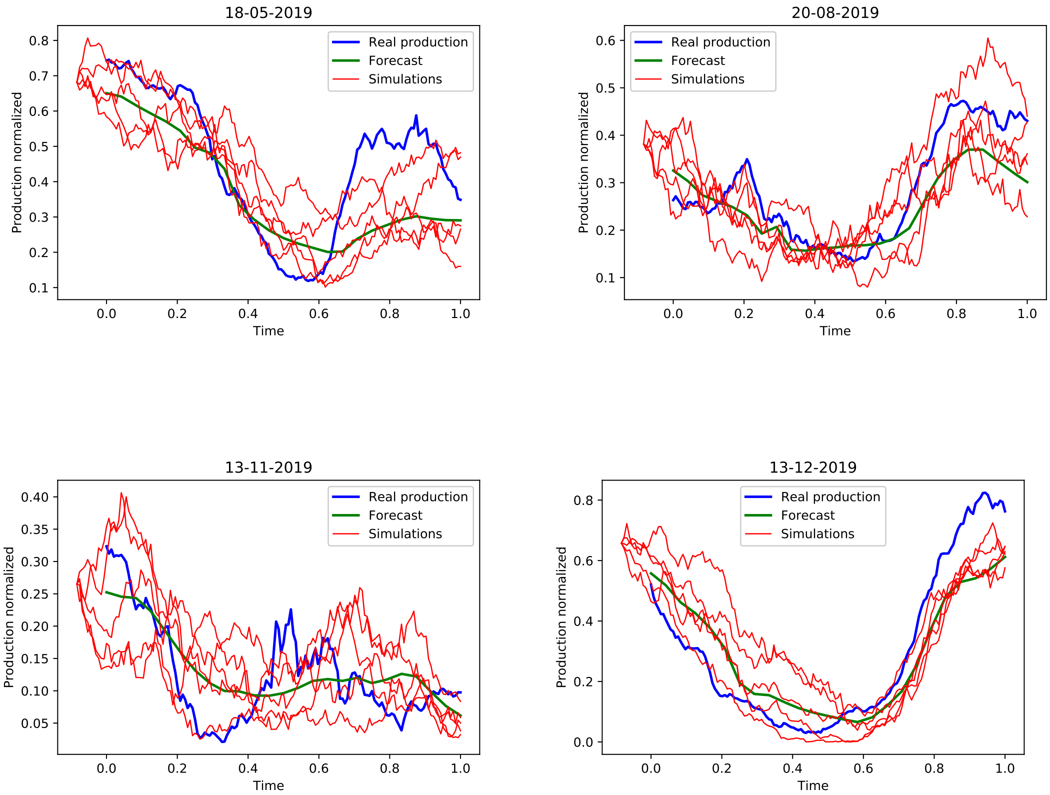
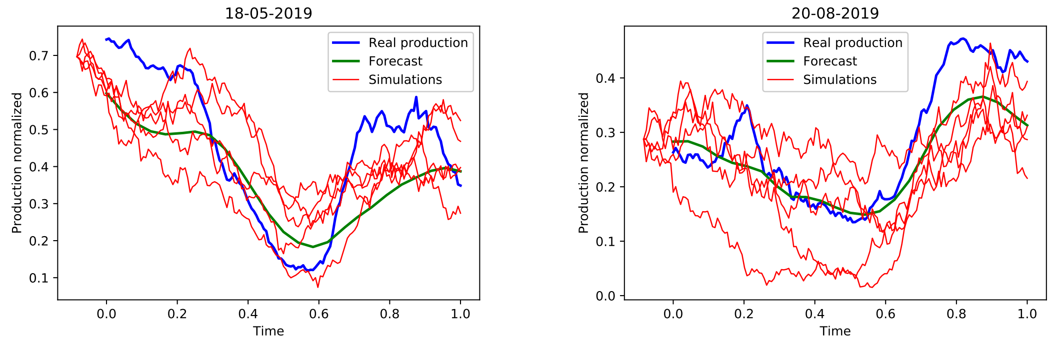


Figure 16: Path simulation for provider A

- Provider B:



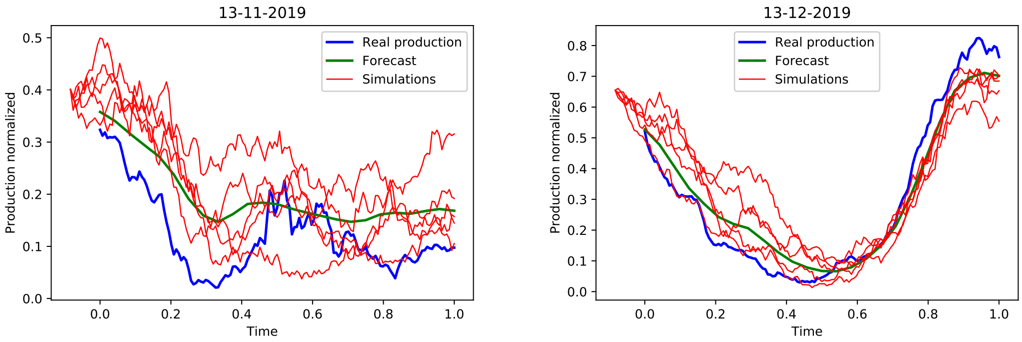


Figure 17: Path simulation for provider B

- Provider C:

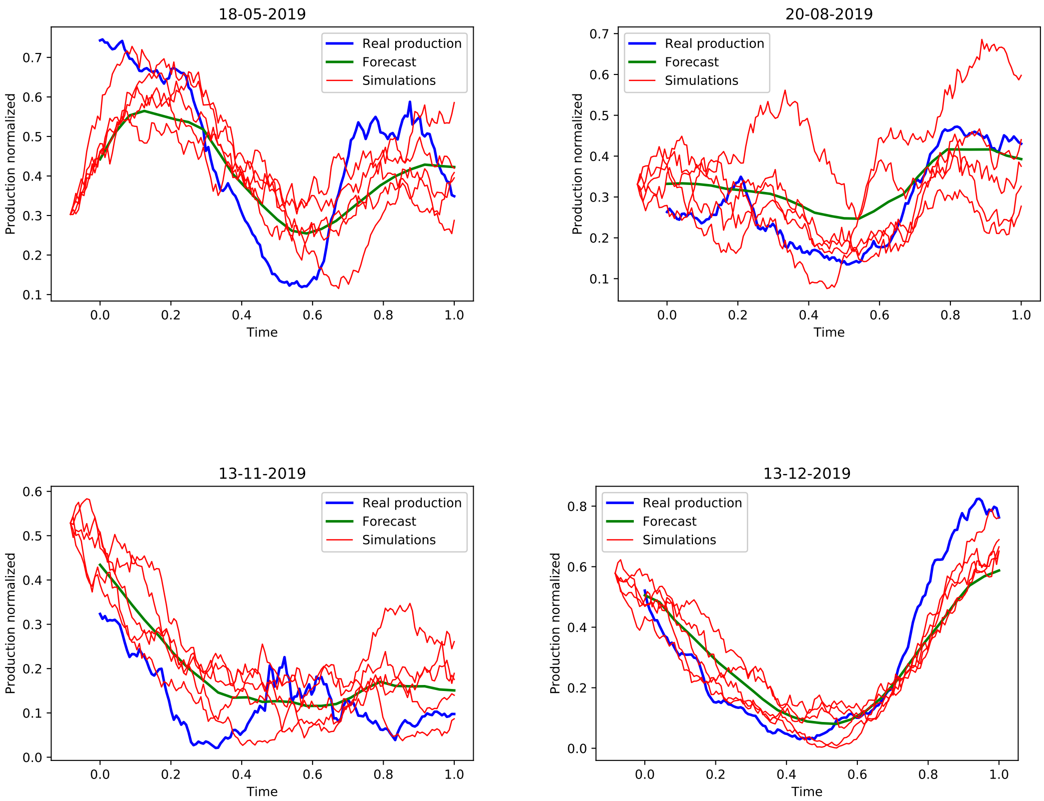


Figure 18: Path simulation for provider C

We can see from these plots that the simulations with our new model are closer to the real production than the forecast.

7.4 Confidence bands

In this section we computed three different confidence interval bands: 99%, 90% and 50% from our simulations (we created a 100 simulations per day) using quantiles.

For each provider we have the following plots:

- Provider A:

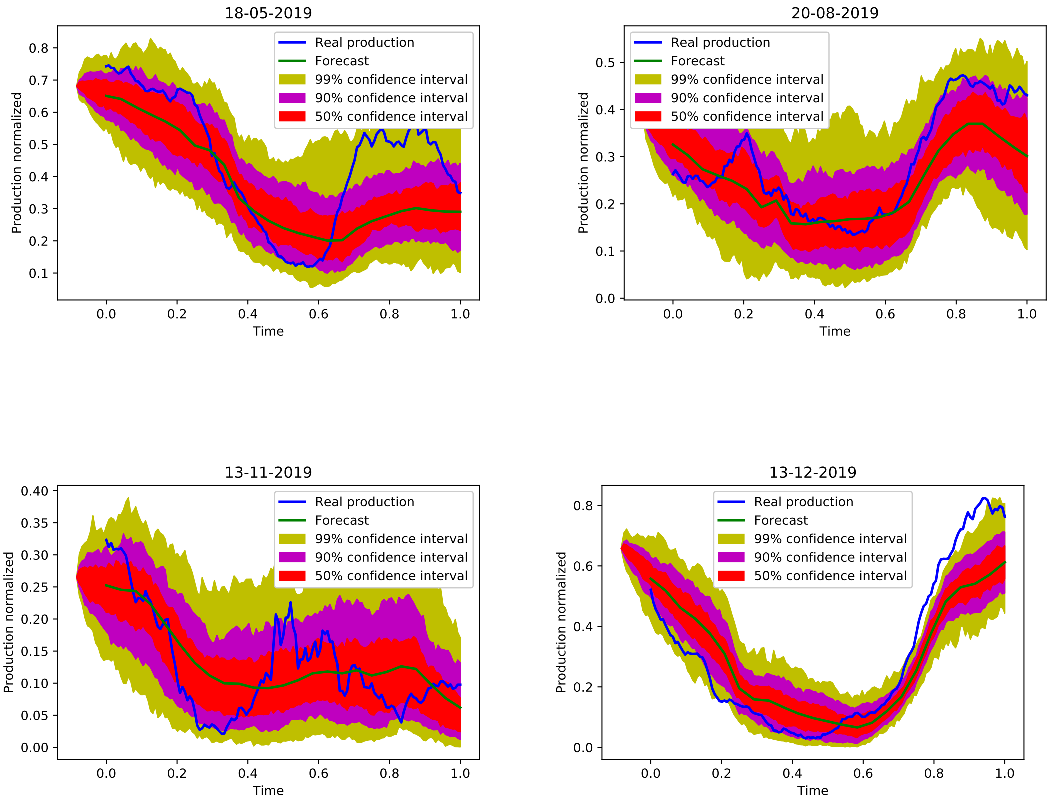


Figure 19: Confidence bands for provider A

- Provider B:

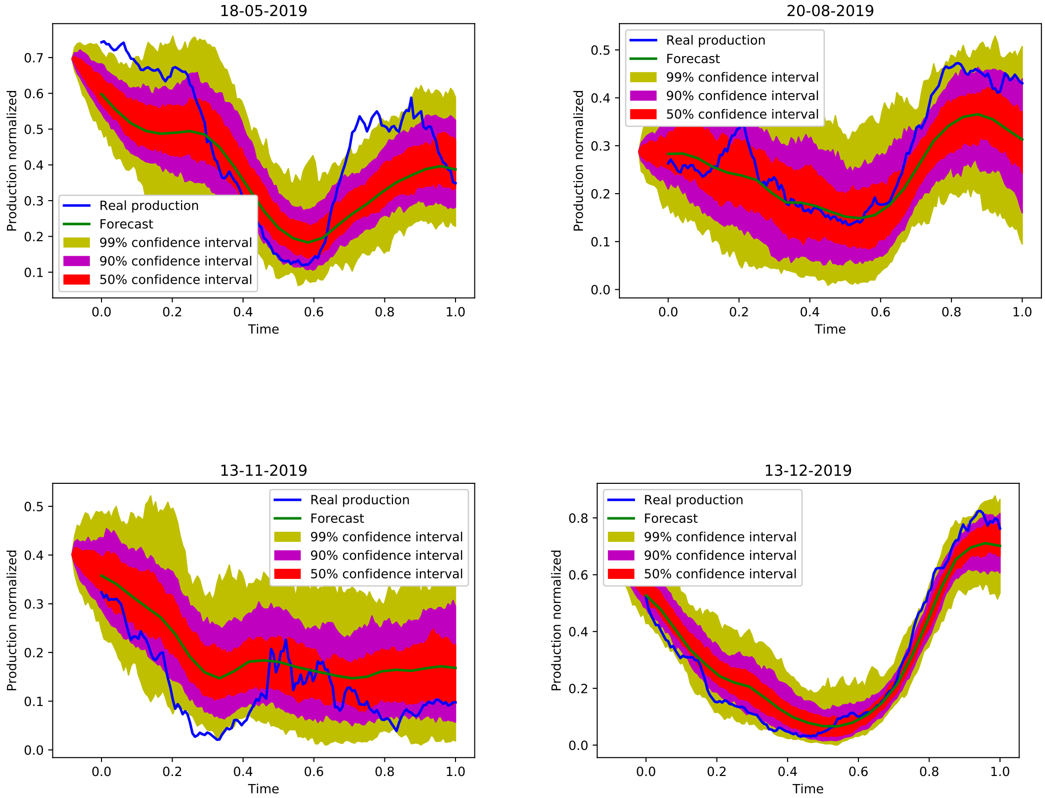


Figure 20: Confidence bands for provider B

- Provider C:

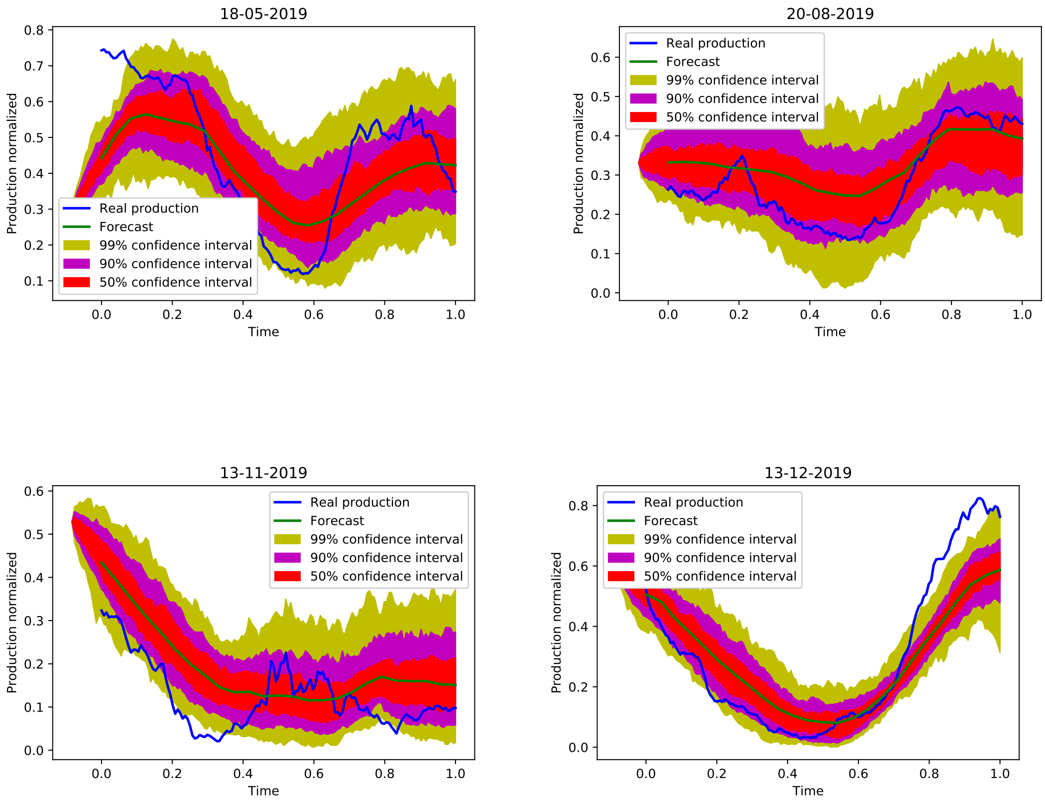


Figure 21: Confidence band for provider C

8 Conclusion

In this work, we proposed an approach to produce forecasts for wind power production using stochastic differential equations and parameter estimation. This method is independent of the forecasting technique used and serves as a complement to the prediction process.

This model features time-derivative tracking of the forecast and a time-dependent mean reversion parameter.

We focused on the data from Uruguay from three different official providers and we were able to simulate wind power production paths and compute confidence bands.

Finally, while this method was designed for wind power production forecast, a similar approach could be used for solar power prediction, taking into consideration the physical differences between the two.

9 Appendix

9.1 Itô's Formula

$X(t)$ is a stochastic process satisfying the following stochastic differential equation (SDE):

$$dX(t) = a(X(t))dt + b(X(t))dW(t)$$

where $W(t)$ is a Wiener process.
Let $g : [0, T] \times R \to R$ be a given bounded function in $C^2([0, T] \times R)$ Then $Y(t) \equiv g(t, X(t))$ satisfies the SDE:

$$dY(t) = \left(\partial_t g + \partial_x g a + \partial^2_{xx} g \frac{b^2}{2}\right)(t, X(t))dt + (\partial_x g b)(t, X(t))dW(t)$$

9.2 Euler-Maruyama method

Euler method, is a numerical approximation of the solution of a stochastic differential equation (SDE). Let us consider the following SDE

$$\begin{cases} dX(t) = a(X(t))t + b(X(t))dW(t) \\ X(0) = x_0 \end{cases}$$

Suppose that we wish to solve this SDE on some interval of time $[0, T]$. Then, the Euler approximation to the true solution X is the process $\bar{X}$ recursively defined on $\Delta t > 0 : 0 = t_0 < t_1 < \cdots < t_N = T$ and $\Delta t = T/N$, by

$$\bar{X}_{n+1} - \bar{X}_n = a\left(\bar{X}_n\right)\Delta t + b\left(\bar{X}_n\right)\Delta W_n$$

for $1 \leq n \leq N$ where $\Delta W_n = W_{t_{n+1}} - W_{t_n}$

9.3

10 References