Probabilistic Forecast of Wind Power Generation by Stochastic Differential Equation Models

Thesis by

Soumaya Elkantassi

In Partial Fulfillment of the Requirements

For the Degree of

Masters of Science

King Abdullah University of Science and Technology

Thuwal, Kingdom of Saudi Arabia

April, 2017

EXAMINATION COMMITTEE PAGE

The thesis of Soumaya Elkantassi is approved by the examination committee

Committee Chairperson: Raul Tempone

Committee Co-Chair: Evangelia Kalligiannaki

Committee Members: Raphaël Huser, Marco Scavino

ABSTRACT

Probabilistic Forecast of Wind Power Generation by Stochastic Differential Equation Models

Soumaya Elkantassi

Reliable forecasting of wind power generation is crucial to optimal control of costs in generation of electricity with respect to the electricity demand. Here, we propose and analyze stochastic wind power forecast models described by parametrized stochastic differential equations, which introduce appropriate fluctuations in numerical forecast outputs. We use an approximate maximum likelihood method to infer the model parameters taking into account the time correlated sets of data. Furthermore, we study the validity and sensitivity of the parameters for each model. We applied our models to Uruguayan wind power production as determined by historical data and corresponding numerical forecasts for the period of March 1 to May 31, 2016.

Key words: Indirect inference, wind power, probabilistic forecasting, model selection, sensitivity.

ACKNOWLEDGEMENTS

I am grateful to Allah for the good health and the well-being that were indispensable for achieving this Masters thesis.

I take this opportunity to express my sincere gratitude to my supervisor Prof. Raul Tempone and my co-supervisor Dr. Evangelia Kalligiannaki for accepting me as a member of their research group for the last two years. I am really grateful for their assistance and guidance throughout my Masters degree.

I owe my gratitude to all my professors at KAUST and in particular the committee members Prof. Marco Scavino and Prof. Raphaël Huser.

I want to express my heartfelt appreciation to my family for their continuous encouragement and their deep support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

Symbol	Meaning
$x_{\max}$	Maximum capacity 793.7 MWatts
$\tilde{p}_{ji}$	Numerical forecast day j at time i
$\tilde{d}_{ji}$	Real production day j at time i
p_{ji}	Normalized numerical forecast day j at time i
d_{ji}	Normalized real production day j at time i
$\tilde{p}(t)$	Numerical forecast, $\tilde{p}(t) \in [0, \mathrm{x}_{\max}]$
$p(t)$	Normalized numerical forecast, $p(t) = \frac{\tilde{p}(t)}{\mathrm{x}_{\max}} \in [0, 1]$
$\tilde{X}(t)$	Stochastic forecast, $\tilde{X}(t) \in [0, \mathrm{x}_{\max}]$
$X(t)$	Normalized stochastic forecast, Model 1
$Y(t)$	Normalized stochastic forecast, Model 2
$Z(t)$	Lamperti transform of $Y(t)$
$\mu_X(t)$	Mean of the process $X(t)$
$v_X(t)$	Variance of the process $X(t)$
$v_X(t, s)$	Co-variance of the process $X(t)$
$\mu_Y(t)$	Mean of the process $Y(t)$
$v_Y(t)$	Variance of the process $Y(t)$
$v_Y(t, s)$	Co-variance of the process $Y(t)$

Chapter 1

Introduction

Over the past 10 years, Uruguay has seen a spectacular transformation in production of wind-based power. In 2007, Uruguay did not produce energy from wind; by 2016, the installed capacity for wind power generation was 1210 MWatts satisfying more than 12% of the total electricity demand of the country. Moreover, renewable energy production in Uruguay provides 94.5% of the country's electricity and 55% of the country's total energy mix[1]. To incorporate the increased amount of generated power into the supply system, reliable forecasting models are crucial.

Many statistical forecast methods have been developed, for example persistence models [1], Kalman filters [2], and Autoregressive and Moving Average Model (ARMA) [3]. The main feature that distinguishes the forecast methods, and the corresponding time-horizon they are valid, is the purpose of its use (operation scheduling, electrical grid management, maintenance). A nice review of different wind power forecasting approaches is given in [4]. Stochastic forecast models are adequate for relative long-term applications because of their dependence on Numerical Weather Prediction (NWP) models. They consider the uncertainty of the deterministic forecast by using historical power production data. Probabilistic forecast models can also be used for stochastic optimization problems for optimizing the wind farms operation and energy trading. Taking into account the production costs of each energy source (fossil fuel, thermal hydropower, biomass, wind) and the storage capacity, such a model makes

[1]The statistics are sourced from the Global Wind Energy Council and the World Resources Institute.

it possible for decision makers to control electricity costs in the market.

Recently, a probabilistic forecast model based on stochastic differential equations (SDEs) [5] was introduced. Our goal is to extend the previous work to include a parametric form for the drift coefficient that controls the quality of the forecast, to explore the need for Lamperti transformation in [5] by considering inferences directly from state-dependent diffusion, and to find the best model fit by pursuing variability analysis and examining sensitivity of the models parameters.

In the following section, we describe the available datasets and the preprocessing procedure.

In Section 2 we formulate the SDEs and propose two models. Then, benchmark examples and numerical results for each model are given in Chapters 4 and 5 and we conclude by discussing our results.

1.1 Description of the data

The stochastic wind power forecasting model is built on two datasets. The first dataset consists of 72 hours of numerical predictions of wind power production, $\tilde{p}_i,\ i = 1,\ldots,72$. The second dataset includes the real (historical) wind power produced during the same time, $\tilde{d}_i,\ i = 1,\ldots,72$, (one observation every 1 hour) in Uruguay. The available data sets range from March 1 to May 31 2016, thus the available observations are $(\tilde{p}_{ji}, \tilde{d}_{ji}),\ i = 1,\ldots,72,\ j = 1,\ldots,81$, after discarding corrupted and missing observations. The value of the data represents the aggregate power (in MWatts) in the following wind farms: ARTI, CAPE, FACE, CPPP, FLO1, JPTE, KENT, LRLM, MIN1, MWIN, RSUR and TDMA. The nominal (maximum) capacity of this set of farms is $x_{\max} = 793.7$ MWatts. We normalize each of the data sets with the maximum installed capacity (793.7 MW) to obtain values between $[0, 1]$ for modelling purposes. We will thus refer to the normalized data sets $(p_{ji}, d_{ji}),\ i = 1,\ldots,72,\ j = 1,\ldots,81$ in the sequel.

To illustrate more the data used, we present in Fig (1.1) the numerical forecast and the real wind power production for a time horizon of $T = 72$ hours for three successive days. The figure depicts the redundancy in the data, that is for example the recorded real production for the 25th up to 48th hours of March 3, is identical to the 1st up to 24th hour of March 4. We point out the redundancy of the data here, as it must be considered in the inference approach, see section 2.3.

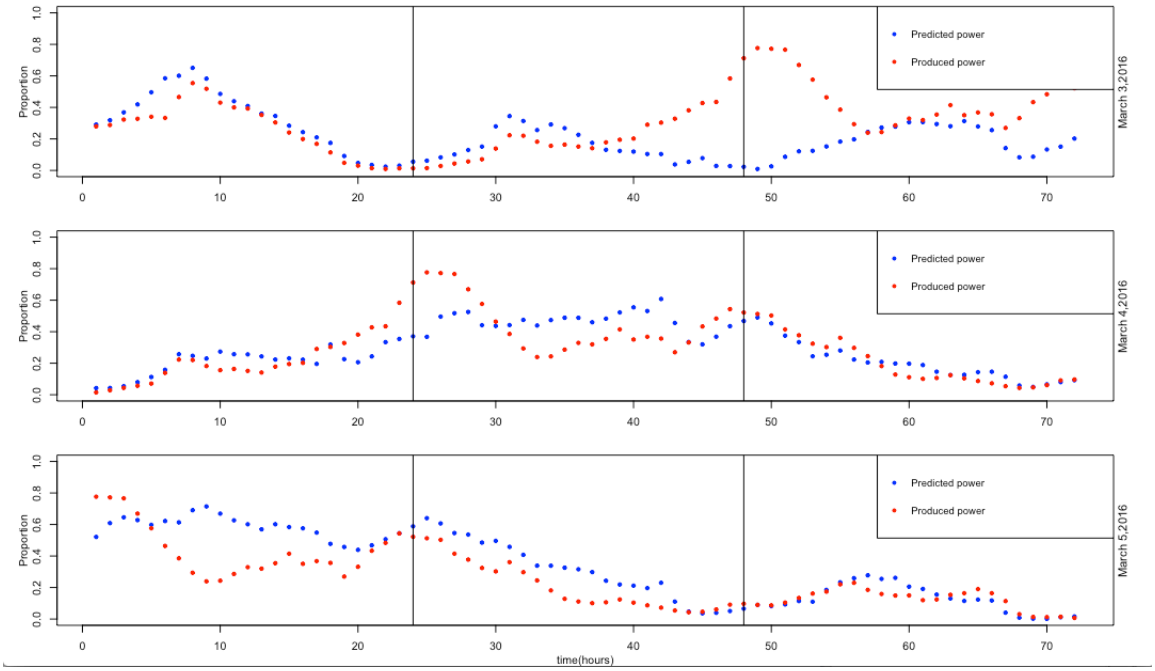

Figure 1.1: Numerical forecast and real production comparison

Chapter 2

Stochastic differential equations and inference

2.1 Introduction

In this chapter, we present the theoretical formulation of our problem. We define a continuous time stochastic process governed by a parametrized SDE. Given a set of observations our goal is to estimate the optimal parameter set for the SDE model. The parameter estimation is a crucial step in all applications. In our work, we consider the maximum likelihood approach for the estimation of the parameters, which are ideally obtained by maximizing the exact likelihood function. However, the exact likelihood function is usually unknown or very costly to calculate numerically. Thus, we use an approximate likelihood for the parameters inference. Here, we consider a Gaussian approximation for the likelihood and derive the equations for the first two moments that define it.

2.2 Stochastic differential equations

Consider a stochastic process, $X(t), t \geq 0$ taking values in $\mathcal{X}$ that satisfies the following Itô diffusion SDE

$$\begin{cases} dX(t) = b(X(t), t; \boldsymbol{\theta})dt + \sigma(X(t), t; \boldsymbol{\theta})dW(t), \quad t > 0\,, \\ X(0) = x_0\,, \end{cases} \tag{2.1}$$

The stochastic differential equation (2.1) can also be written in integral form:

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

When $X(t)$ is a Markov process, its transition probability density is $p(t, y|s, x; \boldsymbol{\theta}) = P(X(t) = y|X(s) = x; \boldsymbol{\theta}), t \leq s$ and satisfies the Fokker-Planck equation

$$\begin{cases} \frac{\partial p(y,t|x,s;\boldsymbol{\theta})}{\partial t} = & -\frac{\partial}{\partial y}\big(b(y,t)p(y,t|x,s;\boldsymbol{\theta})\big) + \frac{1}{2}\frac{\partial^2}{\partial^2 y}\big(\sigma(y,t;\boldsymbol{\theta})p(y,t|x,s;\boldsymbol{\theta})\big) \\ p(0,y) = & f_0(y) \end{cases}$$

such that $s, t \in [0\ T]$, $x, y \in \mathcal{X}$, where $\mathcal{X}$ is the space domain, $f_0(.)$ 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 $p(0, y) = \delta(x - y)$.

2.3 Maximum Likelihood Inference for discretely observed diffusion processes

Given a discrete sample set $\mathbf{d}_N = \{d_i = X(t_i)\}_{i=1}^N$ associated to the process $X(t)$, solution of the SDE (2.1), we can infer the model parameters $\boldsymbol{\theta}$ by maximizing the (exact) likelihood function

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

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

$$L(\boldsymbol{\theta}; \mathbf{d}_N)) = p_1(d_0) \prod_{i=1}^{N} p_i(d_i, t_i | d_{i-1}, t_{i-1}; \boldsymbol{\theta}) \,,$$

in terms of the transition probabilities p_i. In the real application of the current study, the data sets are of the form

$$\mathbf{d}_{JN} = \left\{ \mathbf{d}^{(j)} = \left\{ d_{ji} = X^{(j)}(t_i) \right\}_{i=1}^{N} \right\}_{j=1}^{J} \,,$$

associated to a family of stochastic processes $X^{(j)}(t)$ satisfying different SDE's but which depend on the same parameter set $\boldsymbol{\theta}$. For this case, assuming that $\{\mathbf{d}^{(j)}\}_{j=1}^{J}$ are independent, the exact likelihood function is

$$L(\boldsymbol{\theta}; \mathbf{d}_{JN}) = \prod_{j=1}^{J} p^{(j)}(\mathbf{d}^{(\mathbf{j})}; \boldsymbol{\theta}) \,. \tag{2.2}$$

Thus, given the data $\mathbf{d}_{JN}$, the maximum likelihood estimates of the parameters $\boldsymbol{\theta} \in \Theta$ are obtained by maximizing the function (2.2). The transition probabilities $p^{(j)}(x, t|y, s; \boldsymbol{\theta})$ 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 prohibitively costly in many applications. Thus, to obtain parameter estimates we use some scheme approximation. These approximation schemes may approximate the transition density directly or the path of the process, [6].

2.3.1 Likelihood and Gaussian process approximation

In this section, we consider a Gaussian approximation to obtain an approximation to the likelihood function (2.2). Let $G^{(j)}(t)$ be the Gaussian approximation of processes $X^{(j)}(t),\ j = 1, \ldots, J$ satisfying equation (2.1), and the corresponding likelihood function associated to (2.2). We denote by $\mu^{(j)}(t) = \mathbb{E}[X^{(j)}(t)]$, $v^{(j)}(t) =$

$\mathbb{E}\left[\left(X^{(j)}(t)-\mu^{(j)}(t)\right)^2\right]$, $v^{(j)}(t,s)=\mathbb{E}\left[\left(X^{(j)}(t)-\mu^{(j)}(t)\right)\left(X^{(j)}(s)-\mu^{(j)}(s)\right)\right]$ the mean, the variance and the covariance function respectively, see [7]. Within this approximation we consider the following observation equation to incorporate the measurement error

$$\mathbf{D}^{(j)}=\mathbf{G}^{(j)}+\epsilon^{(j)}\,. \tag{2.3}$$

Here $\epsilon^{(j)}$ denotes the measurement error and we assume that it follows a Gaussian distribution $\mathcal{N}(0,\mathbf{\Sigma}^{\epsilon,(j)})$ with zero mean and variance-covariance matrix $\mathbf{\Sigma}^{\epsilon,(j)}$. The corresponding likelihood function is

$$L(\boldsymbol{\theta};\mathbf{d}_{JN})=\prod_{j=1}^{J}(2\pi)^{-N/2}|\mathbf{\Sigma}^{(j)}(\boldsymbol{\theta})|^{-1/2}\exp\left[(\mathbf{d}^{(j)}-\mu^{(j)}(\boldsymbol{\theta}))^T[\mathbf{\Sigma}^{(j)}(\boldsymbol{\theta})]^{-1}(\mathbf{d}^{(j)}-\mu^{(j)}(\boldsymbol{\theta}))\right]\,, \tag{2.4}$$

where $\boldsymbol{\mu}^{(j)}(\boldsymbol{\theta})=(\mu(t_1),\ldots,\mu(t_N))$, $\mathbf{\Sigma}^{(j)}(\boldsymbol{\theta})=\mathbf{V}^{(j)}(\boldsymbol{\theta})+\mathbf{\Sigma}^{\epsilon,(j)}$ and $\mathbf{V}^{(j)}(\boldsymbol{\theta})$ is the variance-covariance matrix with elements $[\mathbf{V}^{(j)}]_{kl}=v(t_k,t_l)$, $k,l=1,\ldots N$. $|\mathbf{\Sigma}|$ denotes the determinant of the matrix $\mathbf{\Sigma}$.

2.3.2 Two-moment equations

In this section we derive the system of differential equations for the mean $\mu(t)$, the variance $v(t)$, and the covariance functions $v(t,s)$ for the process $X(t)$ satisfying the diffusion equation (2.1). In the sequel, for ease of notation we will remove the explicit dependence on $\boldsymbol{\theta}$.

The first moment equation is the direct result of application of the expectation to the equation (2.1), resulting

$$\frac{d\mu(t)}{dt}=\mathbb{E}\big[b(X(t))\big]\,.$$

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

$$\frac{d\mu(t)}{dt} = b(\mathbb{E}\big[X(t)\big]) = b(\mu(t))\,.$$

Otherwise, we expand the drift around the mean using a Taylor series expansion

$$\frac{d\mu(t)}{dt} = \mathbb{E}\big[b(\mu(t)) + b'(\mu(t))(X(t)-\mu(t)) + \frac{1}{2}b''(\mu(t))(X(t)-\mu(t))^2 + \mathcal{O}(X(t)^3)\big]$$

to obtain a closed-form equation of the approximate mean

$$\begin{cases} \frac{d\mu(t)}{dt} = b(\mu(t)) + \frac{1}{2}b''(\mu(t))v(t)\,, \\ \mu(0) = x_0\,. \end{cases} \tag{2.5}$$

The second moment equation is obtained by applying Itô formula (see appendix $A.2.1$) to the process $Y(t)$, defined as $Y(t) = g(X(t)) = X(t)^2$. Then, $g_t = 0;\ g'(X(t)) = 2X;\ g''(X(t)) = 2$, and

$$\begin{aligned} dY(t) &= 2X(t)dX(t) + \frac{1}{2}2\sigma^2(X(t))dt \\ &= 2X(t)\Big(b(X(t))dt + \sigma(X(t))dW_t\Big) + \frac{1}{2}2\sigma^2(X(t))dt \\ &= 2\Big(X(t)b(X(t)) + \frac{1}{2}\sigma^2(X(t))\Big)dt + 2X(t)\sigma(X(t))dW_t\,. \end{aligned} \tag{2.6}$$

Applying the expectation on both sides of equation (2.6) and using Taylor series expansion of b and σ^2 around the mean, we obtain

$$\begin{aligned} d\mathbb{E}\big[Y(t)\big] &= 2\mathbb{E}\Big[X(t)b(X(t)) + \frac{1}{2}\sigma^2(X(t))\Big]dt \\ &= 2\mathbb{E}\Big[X(t)\big(b(\mu(t)) + b'(\mu(t))(X(t)-\mu(t)\big) + \mathcal{O}(X^3(t))\Big]dt \\ &+ 2\mathbb{E}\big[\frac{1}{2}\sigma^2(\mu(t)) + \frac{1}{2}\sigma^2(x)'|_{x=\mu(t)}\big(X(t)-\mu(t)\big) + \frac{1}{4}\sigma^2(x)''|_{x=\mu(t)}\big(X(t)-\mu(t)\big)^2 + \mathcal{O}(X^3(t))\big]dt \\ &\approx 2\Big(\mu(t)b(\mu(t)) + b'(\mu(t))v(t) + \frac{1}{2}\sigma^2(\mu(t)) + \frac{1}{4}(\sigma^2)''(\mu(t))v(t)\Big)dt\,. \end{aligned} \tag{2.7}$$

Subtracting the derivative of the squared mean from (2.7) gives

$$\begin{aligned} dv(t) &= d\mathbb{E}\big[X^2(t)\big] - d(\mu^2(t)) \\ &= d\mathbb{E}\big[Y(t)\big] - 2\mu(t)d\mu(t) \\ &\approx 2\Big(\mu(t)b(\mu(t)) + \frac{1}{2}\sigma^2(\mu(t)) + b'(\mu(t))v(t) + \frac{1}{4}(\sigma^2)''(\mu(t))v(t)\Big)dt - 2\mu(t)b(\mu(t))dt \\ &\approx 2\Big(\frac{1}{2}\sigma^2(\mu(t)) + b'(\mu(t))v(t) + \frac{1}{4}(\sigma^2)''(\mu(t))v(t)\Big)dt \end{aligned}$$

Thus, the approximate equation of the variance is given by

$$\begin{cases} \frac{dv(t)}{dt} = \Big(2b'(\mu(t)) + \frac{1}{2}(\sigma^2)''(\mu(t))\Big)v(t) + \sigma^2(\mu(t)) \\ v(0) = 0 \end{cases} \tag{2.8}$$

The covariance function between $X(t)$ and $X(s)$ such that $0 \leq s \leq t \leq T$, is defined by

$$v(t,s) = \mathbb{E}\Big[(X(t) - \mu(t))(X(s) - \mu(s))\Big] = \mathbb{E}\Big[X(t)X(s)\Big] - \mu(t)\mu(s)$$

For a given $s \in [0,T]$, $v(t,s)$ can be considered as a function of time $\forall t \in [s,T]$, then

$$\begin{aligned} \frac{dv(t,s)}{dt} &= \frac{d\mathbb{E}\Big[X(t)X(s)\Big]}{dt} - \mu(s)\frac{d\mu(t)}{dt} \\ &= \frac{d\mathbb{E}\Big[X(t)X(s)\Big]}{dt} - \mu(s)b(\mu(t)) \end{aligned}$$

Applying Dynkin's formula (appendix, A.2.3) to the first term such that $f(X(t)) =$

$X(t)X(s)$ and Taylor series expansion of b around the mean, we obtain

$$\begin{aligned}\frac{d\mathbb{E}\Big[X(t)X(s)\Big]}{dt} &= \mathbb{E}\Big[b(X(t))\frac{df(X(t))}{dX(t)} + \frac{1}{2}\sigma^2(X(t))\frac{d^2f(X(t))}{dX^2(t)}\Big] \\ &= \mathbb{E}\Big[b(X(t))X(s)\Big] \\ &\approx \mu(s)b(\mu(t)) + b'(\mu(t))v(t,s)\,.\end{aligned}$$

Thus, we conclude that the approximate covariance equation is

$$\begin{cases} dv(t,s) = b'(\mu(t))v(t,s)dt \\ v(s,s) = v(s) \end{cases} \tag{2.9}$$

2.4 The proposed models

In this section we introduce a stochastic process, $X(t) \in [0,1]$ that serves as the stochastic analogue of the normalized deterministic forecast, $p(t) \in [0,1]$, for $t \in [0,T]$. We consider that the stochastic process, $X(t)$, is the solution of the following parametrized Ito diffusion SDE

$$\begin{cases} dX(t) = b(X(t),t;\boldsymbol{\theta})dt + \sigma(X(t),t;\boldsymbol{\theta})dW(t), \quad t>0\,, \\ X(0) = x_0\,, \end{cases}$$

and we assume that the drift, $b(X(t),t;\boldsymbol{\theta})$, and the diffusion coefficients are such that a solution exists and is unique, [8]. Here, $\boldsymbol{\theta}$ denotes the set of parameters for each model. $W(t)$ is a one-dimensional Brownian motion. 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

$$b(x,t;\boldsymbol{\theta}) = -\theta(t)(x-p(t))\,, \tag{2.10}$$

where $\theta(t)$ represents the rate by which the variable reverts towards $p(t)$ in time.

We distinguish two major models (Model 1 and Model 2) based on the choice of the diffusion coefficient. This distinction is mainly for an inference purpose and not for a modeling reason. We would like to explore the need of the transformation to state independent diffusion coefficient.

2.4.1 Model 1

We want the diffusion at the boundaries to vanish (i.e. if $X(t) \in \{0, 1\}$, then $\sigma(0) = \sigma(1) = 0$) such that the stochastic forecast does not exceed the total capacity. Thus, we choose the diffusion to be a second-order polynomial with respect to the state of the process of the form

$$\sigma(X(t), t; \boldsymbol{\theta}) = \sqrt{2\theta(t)\alpha X(t)(1 - X(t))}.$$

Summarizing, the first proposed model is

$$\begin{cases} dX(t) = -\theta(t)\Big(X(t) - p(t)\Big)dt + \sqrt{2\alpha\theta(t)X(t)\Big(1 - X(t)\Big)}dW_t, \quad t > 0\,, \\ X(0) = x_0\,, \end{cases} \tag{Model 1}$$

and

$$\boldsymbol{\theta} = (\theta(t), \alpha), \quad \alpha > 0,\ \theta(t) > 0, \quad \text{for all } t > 0.$$

The inference approach for the parameters is based on the two moment equations for the process, $X(t)$. For that, we denote $\mu_X(t) = \mathbb{E}[X(t)]$, $v_X(t) = \text{var}[X(t)] = \mathbb{E}[(X(t) - \mu_X(t))^2]$ and $v_X(t, s) = \text{cov}(X(t), X(s)) = \mathbb{E}[(X(t) - \mu_X(t))(X(s) - \mu_X(s))]$ as the mean, variance and covariance functions, respectively.

Following the same tools and steps as in Section 2.3.1 , the equation of the mean is given by

$$\begin{cases} d\mu_X(t) = -\theta(t)(\mu_X(t) - p(t))dt, \quad t > 0\,, \\ \mu_X(0) = \mu_0 = x_0\,, \end{cases}$$

with the solution

$$\mu_X(t) = e^{-\int_0^t \theta(s)ds} \left(\int_0^t \theta(s)p(s)e^{\int_0^s \theta(u)du} ds + \mu_0 \right). \tag{2.11}$$

The equation of the variance is

$$\begin{cases} dv_X(t) = -2\theta(t)\Big((1+\alpha)v_X(t) - \alpha\mu_X(t)(1-\mu_X(t))\Big)dt, \quad t > 0\,, \\ v_X(0) = 0\,, \end{cases}$$

from which we have

$$v_X(t) = e^{-2(1+\alpha)\int_0^t \theta(s)ds} \left(\int_0^t 2\theta(s)\alpha\mu_X(s)(1-\mu_X(s))e^{2(1+\alpha)\int_0^s \theta(s)du} ds \right). \tag{2.12}$$

The equation of the covariance between two times of $X(t)$ is given by

$$\begin{cases} \frac{dv_X(t,s)}{dt} = -\theta(t)v_X(t,s), \ \forall t > s\,, \\ v_X(t,s) = v_X(s), t = s\,, \end{cases}$$

from which we obtain

$$v_X(t,s) = v_X(s)e^{-\int_0^t \theta(u)du}, \ \forall t > s\,. \tag{2.13}$$

2.4.2 Model 2

In this section, we transform equation (Model 1) by means of the Lamperti transform such that the diffusion term of the transformed system is independent of the

state. This transformation is recommended in many references. In [6] it is stated that this type of transformation is worth using in both simulation and estimation because numerical schemes sampled from an SDE with additive noise are more stable and usually of higher strong order than those sampled from an SDE with multiplicative noise. Moreover, [9] and [5] suggested that the transformed process, $Z_X(t) = \Psi(X(t)) = \arcsin(2X(t) - 1)$, is better approximated by a Gaussian distribution.

Lamperti transformation

Let us denote by $Z(t) = \Psi(X(t))$ the transformed process that satisfies an SDE with state-independent diffusion, where $\Psi(X(t)) \in \mathbb{C}^2([0, \infty])$. Applying Itô formula to $Z(t)$, where $X(t)$ satisfies equation (Model 1), we get

$$\begin{aligned} dZ(t) &= \frac{\partial \Psi}{\partial t}dt + \frac{\partial \Psi}{\partial X}dX(t) + \frac{1}{2}\frac{\partial^2 \Psi}{\partial X^2}(dX(t))^2 \\ &= \Psi_t dt + \Psi'\big(b(X(t))dt + \sigma(X(t))dW_t\big) + \frac{1}{2}\Psi''\sigma^2(X(t))dt \\ &= \big(\Psi_t + \Psi'(b(X(t)) + \frac{1}{2}\Psi''\sigma^2(X(t))\big)dt + \Psi'(\sigma(X(t))dW_t \end{aligned}$$

Then by choosing :

$$\Psi(X(t)) = \int \frac{1}{\sigma(x)}dx\Big|_{x=X(t)}$$

$Z(t)$ has a state-independent diffusion and it is governed by the following SDE

$$dZ(t) = \big(\Psi_t(\Psi^{-1}(Z)) + \frac{b(\Psi^{-1}(Z))}{\sigma(\Psi^{-1}(Z))} - \frac{1}{2}\sigma_X(\Psi^{-1}(Z))\big)dt + dW_t\,.$$

The general use of the Lamperti transform is limited to univariate diffusion processes and a restricted class of multivariate diffusion processes [10]. To ensure the uniqueness of the solution $\mu_Z(t)$ for the mean equation of the transformed process $Z(t)$, the

following condition should be satisfied

$$\alpha \leq \min(p(t), 1 - p(t))\,.$$

One way to satisfy this condition is by choosing the parameter α

$$\alpha(t) = \alpha p(t)(1 - p(t))\,.$$

We denote the multiplicative constant with the same letter as in the function $\alpha(t)$, because in the sequel we will refer only to the constant α. This suggests the form of the second proposed model

$$\begin{cases} dY(t) = -\theta\Big(Y(t) - p(t)\Big)dt + \sqrt{2\theta\alpha p(t)\Big(1 - p(t)\Big)Y(t)\Big(1 - Y(t)\Big)}dW_t, \quad t > 0\,, \\ Y(0) = x_0\,, \end{cases} \tag{Model 2}$$

and

$$\boldsymbol{\theta} = (\theta(t), \alpha), \quad \alpha > 0,\ \theta(t) > 0, \quad \text{for all } t > 0.$$

Following the proof in the appendix A.3, the Lamperti transformation of the process $Y(t)$ defines the process $Z(t)$ that satisfies the following SDE:

$$dZ(t) = \tilde{b}(Z(t); \boldsymbol{\theta})dt + \tilde{\sigma}(t; \boldsymbol{\theta})dW_t\,, \tag{2.14}$$

where

$$\tilde{b}(x; \boldsymbol{\theta}) = \frac{-\theta(1 + \sin(x) - 2p(t)) + \frac{1}{2}\sin(x)\tilde{\sigma}^2(t; \boldsymbol{\theta})}{\cos(x)}\,,$$

and

$$\tilde{\sigma}(t; \boldsymbol{\theta}) = \sqrt{2\theta(t)\alpha p(t)(1 - p(t))}\,.$$

In this section we will derive the two moment equation of the process $Z(t)$ that satisfies

equation (2.14), then transform it back to $Y(t)$ in order to obtain an approximation of its mean and covariance functions.

The approximate mean $\mu_Z(t) \approx \mathbb{E}[Z(t)]$ of $Z(t)$ is

$$\begin{cases} \frac{d\mu_Z(t)}{dt} = \tilde{b}(\mu_Z(t))\,, \\ \mu_Z(0) = E[Z_0] = z_0 = \arcsin(2x_0 - 1) \end{cases}$$

An exact solution of the mean can be written as

$$\mu_Z(t) = \arcsin\left\{ e^{-\int_0^t h(u)du} \Big(\int_0^t \theta(2p(s)-1) e^{\int_0^s h(u)du} ds + \sin(Z(0)) \Big) \right\} \tag{2.15}$$

$$\text{where} \quad h(t) = \theta[1 - \alpha p(t)(1-p(t))]$$

Using the inverse function of Ψ given by: $Y(t) = \Psi^{-1}(Z(t)) = \frac{1}{2}\Big(1 + \sin(Z(t))\Big)$, we obtain the following mean solution of the process $Y(t)$ after using Lamperti transform

$$\tilde{\mu}_Y(t) = \Psi^{-1}(\mu_Z(t)) = \frac{1}{2}(1 + \sin(\mu_Z(t)))\,. \tag{2.16}$$

The approximate variance $v_Z(t) \approx \mathbb{E}[\big(Z(t) - \mu_Z(t)\big)^2]$ is given by

$$\begin{cases} \frac{dv_Z(t)}{dt} = 2\tilde{b}'(\mu_Z(t))v(t) + \tilde{\sigma}^2(t)\,, \\ v_Z(0) = 0 \end{cases}$$

and an exact solution for this first-order ODE is given by

$$v_Z(t) = e^{2\int_0^t A(s;\boldsymbol{\theta})ds} \int_0^t \frac{1}{2}\tilde{\sigma}^2(s;\boldsymbol{\theta}) e^{-2\int_0^s A(u;\boldsymbol{\theta})du} ds,$$

where $A(t;\boldsymbol{\theta}) = \partial_x \tilde{b}(x;\boldsymbol{\theta}))|_{x=\mu_Z(t)}$.

The approximate covariance function $v_Z(t,s) \approx \mathbb{E}[\big(Z(t)-\mu_Z(t)\big)\big(Z(s)-\mu_Z(s)\big)]$ is

$$v_Z(t,s) = v_Z(s)e^{\int_s^t A(u;\boldsymbol{\theta})du}, \text{ for all } t > s\,.$$

Thus, the approximate variance and covariance functions for $Y(t)$ are

$$\tilde{v}_Y(t) = \frac{1}{4}\cos^2(\mu_z(t))v_Z(t)\,. \tag{2.17}$$

and

$$\tilde{v}_Y(t,s) = \frac{1}{4}\cos(\mu_z(t))v_Z(t,s)\cos(\mu_z(s))\,. \tag{2.18}$$

2.4.3 Bias introduced in Model 2

The equation of the mean of X can be written as

$$\begin{aligned}
\frac{d\tilde{\mu}_Y(t)}{dt} &= \frac{dg(\mu_Z(t))}{dt}, \quad (g = \Psi^{-1}) \\
&= g'(\mu_Z(t))\frac{d\mu_Z(t)}{dt} \\
&= \frac{1}{2}\cos(\mu_Z(t))\Big[\frac{-\theta(t)\big(\tilde{\mu}_Y(t) - p(t) - \alpha p(t)(1-p(t))(1-2\tilde{\mu}_Y(t))\big)}{\frac{1}{2}\cos(\mu_Z(t))}\Big] \\
&= -\theta(t)\Big(\tilde{\mu}_Y(t) - p(t) - \alpha p(t)(1-p(t))(1-2\tilde{\mu}_Y(t))\Big)
\end{aligned} \tag{2.19}$$

Thus, the mean of $Y(t)$ after the use of Lamperti transform satisfies

$$\frac{d\tilde{\mu}_Y(t)}{dt} = -\theta(t)\Big(\tilde{\mu}_Y(t) - p(t) - \alpha p(t)(1-p(t))(1-2\tilde{\mu}_Y(t))\Big)$$

However, for Model 2 the exact equation of the mean is given by

$$\frac{d\mu_Y(t)}{dt} = -\theta(t)\Big(\mu_Y(t) - p(t)\Big)$$

Then, the bias introduced in the mean is given by : $\alpha p(t)(1-p(t))(1-2\tilde{\mu}_Y(t))$. In order to correct this bias, we introduce a compensator and the resulting SDE is given by:

$$\begin{aligned} dY(t) = &-\theta(t)\Big(Y(t) - p(t) - cp(t)(1-p(t))(1-2Y(t))\Big)dt \\ &+ \sqrt{2\theta(t)\alpha p(t)\Big(1-p(t)\Big)Y(t)\Big(1-Y(t)\Big)}dW_t \end{aligned} \tag{2.20}$$

Ideally α and c should be chosen after a trade off between minimizing the bias in the mean and matching the drift term in equation (2.20). In case $c = 0$, we obtain equation (Model 2). The equation of the transformaed process, the equation of mean and the covariance in this case are detailed in appendix (A.3.2).

Chapter 3

Methods for model selection and sensitivity

In this chapter we present two likelihood-based statistics, AIC and BIC, that search the best model embedded in a wider class of models. Then, for the selected model, we are interested in studying the sensitivity of its parameters and assessing the effect of small perturbations. Thus, we introduce the Finite Differences FD method and the Fisher Information Matrix (FIM).

3.1 Model selection criteria

The Akaike information criterion (AIC) measures the goodness of fit of a given model, [6]. Suppose that we have a statistical model for some data $\mathbf{d}_N = \{d_i\}_{i=1}^N$ and let $\hat{L}$ be the likelihood function of the model evaluated at the optimal parameter values (i.e $\hat{L} = L(\hat{\theta}; \mathbf{d}_N)$) and k be the number of free parameters in the model. Then the AIC statistic is defined as

$$\text{AIC} = 2k - 2\ln(\hat{L}).$$

This criterion selects the model that has a good fit and avoids over-fitting by penalizing the inclusion further predictors.

The Bayesian information criterion (BIC) is another goodness of fit measure, [6]. However, the penalty term is larger in BIC than in AIC. The BIC is formally defined as

$$\text{BIC} = \ln(J)k - 2\ln(\hat{L}).$$

where, J is the number of observations, or equivalently, the sample size.

3.2 Sensitivity analysis

In this section we present two methods to assess the effect on the model output of a small change of a parameter. We discuss the sensitivity indices defined by the FD and FIM, [11, 12].

Let $F(X(t);\boldsymbol{\theta})$ be a quantity of interest with expectation

$$f(\boldsymbol{\theta}) = E[F(X(t);\boldsymbol{\theta})]$$

For example $F(X(t);\theta) = \frac{1}{T}\sum_{j=1}^{T} X(t_j;\boldsymbol{\theta}),$ could represent the average wind power production in $[0,\ T]$ and $F(X(t);\boldsymbol{\theta}) = X(t;\boldsymbol{\theta})$, is the normalized power production at time t.

We study the sensitivity of the proposed models with respect to the vector parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ through the derivative estimator ∇f.

Given a small increment δ, the Taylor series expansion of f around θ can be written as

$$f(\theta + \delta a) = f(\theta) + \delta a^T \nabla f(\theta) + \frac{\delta^2}{2} a^T H_f(\theta) a + \mathbb{O}(||\delta a||^3) \tag{3.1}$$

$$f(\theta - \delta a) = f(\theta) - \delta a^T \nabla f(\theta) + \frac{\delta^2}{2} a^T H_f(\theta) a + \mathbb{O}(||\delta a||^3) \tag{3.2}$$

where $a^T \nabla f(\theta)$ gives the directional derivative of f in the direction of a

$$a^T \nabla f(\theta) = \lim_{\delta \to 0} \frac{f(\theta + \delta a) - f(\theta)}{\delta}$$

and $H_f(\theta)$ denotes the Hessian of f evaluated at θ. Higher order terms can be written in terms of tensors but here we just consider second order expansion.

From 3.1, we have

$$\delta a^T \nabla f(\theta) = f(\theta + \delta a) - f(\theta) + \mathbb{O}(||\delta a||^2) \tag{3.3}$$

Approximating the first derivative by the first term, we obtain the following estimator S_{FD} known as the forward difference approximation of $\nabla f(\theta)$

$$a^T S_{\text{FD}} = f(\theta + \delta a) - f(\theta)$$

Similarly, using (3.2), we obtain the backward difference approximation of the first derivative denotes as S_{BD}

$$a^T S_{\text{BD}} = f(\theta) - f(\theta - \delta a)\,.$$

A second-order approximation of $\nabla f(\theta)$ called central-difference is given by

$$\delta a^T S_{\text{CFD}} = f(\theta + \delta a) - f(\theta - \delta a)\,.$$

Another sensitivity study is carried through the FIM defined as

$$\text{FIM}(\theta) = \text{E}\left[\left(\nabla \log p(X;\theta)\right)^2 \middle| \, \theta \right] \,.$$

The information theoretic inequality [12], based on the Cramer-Rao bound is given by

$$||\nabla f(\theta)|| \leq \sqrt{\text{var}(F)\text{tr}(\text{FIM}(\theta))}$$

where tr(FIM) is the trace function of the Fisher Information Matrix. The inequality above indicates that the diagonal elements of the FIM provides an upper bound for the sensitivity indices $\nabla f(\theta)$. Thus, we denote the sensitivity index with respect to

the FIM as

$$S_{\mathrm{FIM}} = \mathrm{FIM}(\theta)\,.$$

Chapter 4

Benchmark Applications

In this chapter, we present the application of the models described in chapter (2) to two benchmark examples. To test our models, we generate synthetic datasets from (Model 1) and (Model 2) using Euler Maruyama method (appendix A.2.2) with a predefined forecast function $p(t)$.

For the two models, we consider a parametrization with respect to time independent parameters. In order to make a fair comparison, we use the same parameters value, optimization method, initial conditions, number of synthetic paths.

4.1 Benchmark applications

Numerical implementation

Two methods can be used to solve the three differential equations of the mean, variance and co-variance. First, we can solve them using an ODE solver such as 'ode45', 'rk', 'euler', etc. Another way is to use numerical integration of their exact solutions. The numerical results presented in the sequel are derived with the second approach, where we use the trapezoidal quadrature with time step $\Delta t = 0.01$. For the minimization procedure we use the function 'fmincon' on MATLAB. We also implemented the first method using R with 'ode45' and 'optimR' packages but it was considerably more time consuming.

The predefined forecast function that we used are

1. The trigonometric function $p(t) = \sin^2(t/6)$, $j = 1, \ldots, J$, that is common for

all generated paths .

2. The functions, $p_j(t),\ i = j, \ldots, J$, fitted from the available numerical forecast data.

Next, we summarize the steps we follow for the benchmark examples independently of the model. First, we choose values for (θ, ϕ, α) and a time horizon, $T \in \{12, 24, 36, 48\}$. Second, we sample M paths from equation (Model 1) or paths from ((2.14)) and transform them to Model 2 using Ψ^{-1}. After sampling the paths using the predefined forecast function p(t), we add noise to the data, $e_i \sim N(0, \Sigma_i^e)$. Then, we compute the corresponding mean, variance and co-variance functions. Finally, we use the sampled synthetic data to maximize the likelihood function, compute the confidence interval of the optimal parameters and compare them to the true values.

4.1.1 Model 1

Trigonometric forecast

For the optimization problem, we include sequentially the following parameters in the likelihood function to test the ability of recovering each one of them.

- MLE0: maximum likelihood estimate of θ for fixed $(\alpha, \phi) = (0.3, 0.02)$.
- MLE: maximum likelihood estimates of (θ, α) for fixed $\phi = 0.02$.
- MLE1: maximum likelihood estimates of (θ, α, ϕ).

Based on the Figures (4.1, 4.2, 4.3), which present estimated values and confidence intervals with respect to the number of paths in the likelihood function, MLE1 is able to recover the true parameters $(0.2, 0.2, 0.05)$. Fig (4.4) confirms that MLE1 is converging with a rate of order $\mathcal{O}(n^{-\frac{1}{2}})$ in alignment with the theoretical convergence of the maximum likelihood estimation [13].

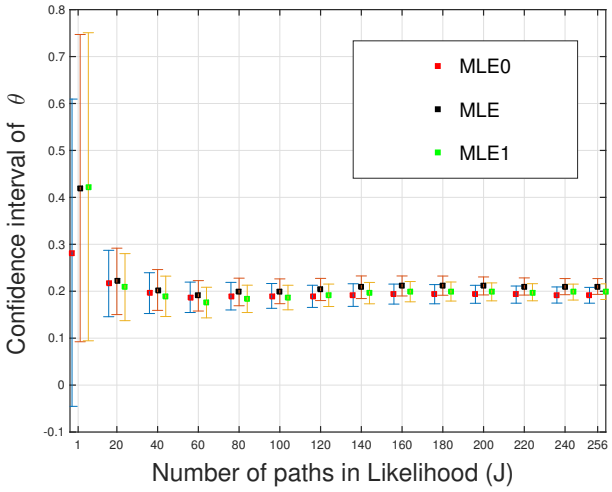

Figure 4.1: Convergence of θ for synthetic data with trigonometric forecast-Model 1

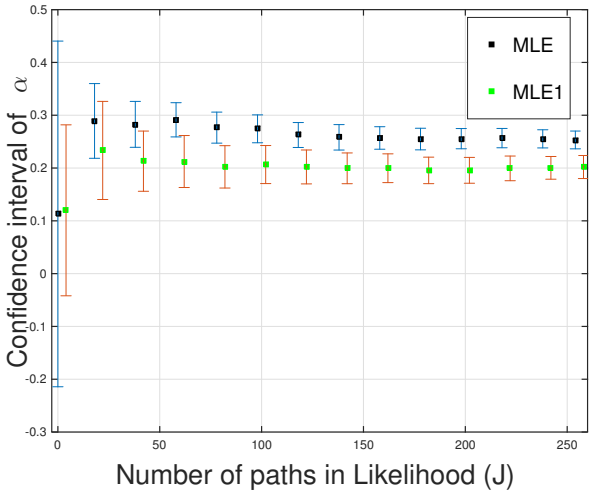

Figure 4.2: Convergence of α for synthetic data with trigonometric forecast-Model 1

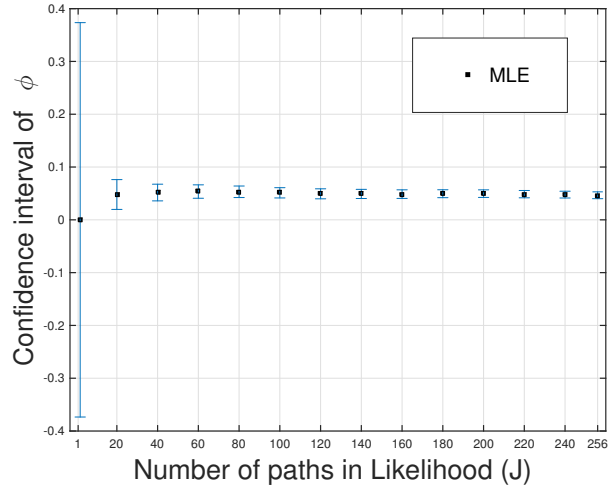

Figure 4.3: Convergence of ϕ for synthetic data with trigonometric forecast-Model 1

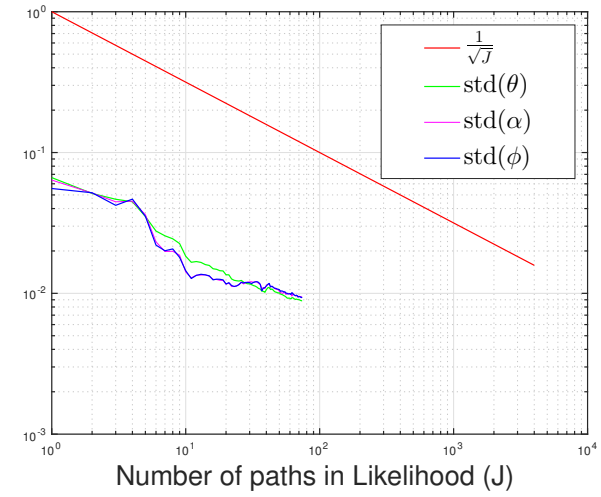

Figure 4.4: Rate of convergence of the maximum likelihood estimates

In Fig (4.5), we plot the sample mean, the exact mean, which is the solution of equation (2.11) with the true values (θ, α, ϕ) and the fitted mean with the maximum likelihood estimates $(\theta^*, \alpha^*, \phi^*)$. For both Figures , (4.5) and (4.6), we can see that the three curves have similar behavior which implies on one hand that the moment approximation provides good results. On the other hand, the estimated parameter values are very close to the true ones.

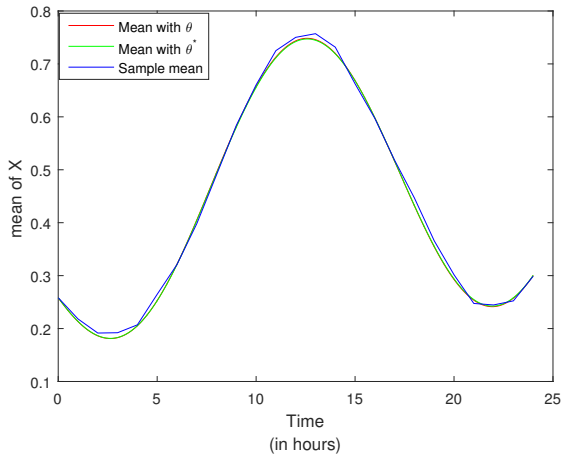

Figure 4.5: Mean vs sample mean with 2^8 synthetic paths.

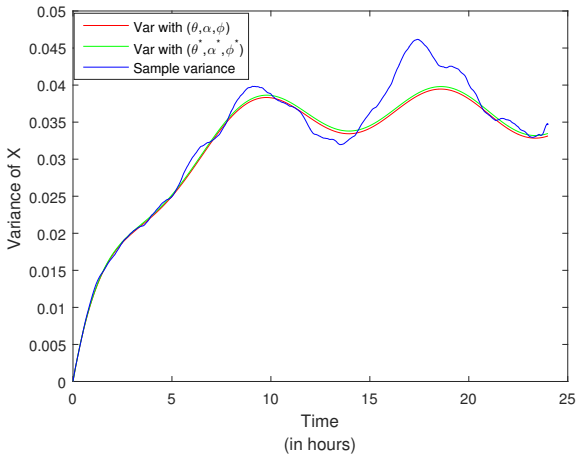

Figure 4.6: Variance vs sample variance with 2^8 synthetic paths.

In Fig (4.7), we present the confidence interval obtained with the Gaussian approximation {(2.11) ,(2.12) ,(2.13)} and two paths: one represents the real production (i.e generated using the true values of (θ, α, ϕ)) and another one represents the numerical prediction, $p(t)$. In Fig (4.8), we present the empirical bands based on 2^8 paths. These figures show the similarity between the approximate and the empirical confidence intervals which proves the consistency of the Gaussian approximation.

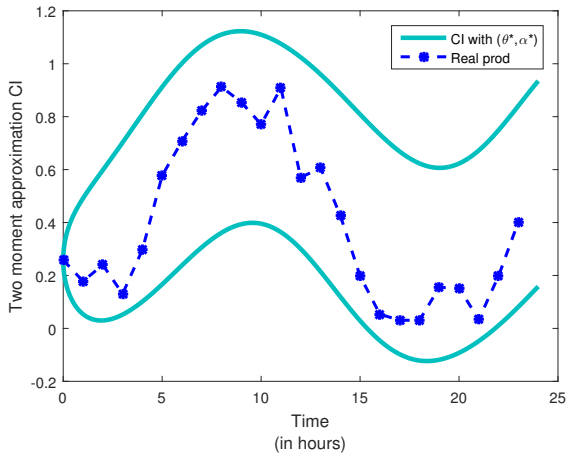

Figure 4.7: 95% Confidence interval obtained with the two moment approximation.

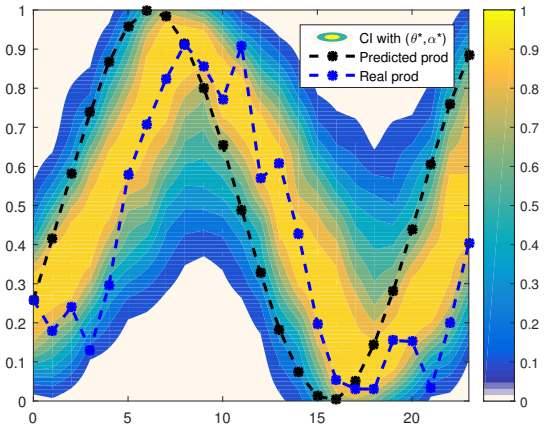

Figure 4.8: Empirical confidence bands with 2^8 synthetic paths.

The empirical bands are plotted using the density of the simulated paths of (Model 1) using the MLE estimators and Euler-Maruyama discretization scheme. For both figures, the empirical bands and the paths look periodic and this comes from the nature of the even and periodic forecast function $\sin^2(Bt)$. We can see that the confidence interval based on the Gaussian approximation exceeds $\{0, 1\}$ and this is a result of the approximation, we also observe that both paths fall inside the 95% of the approximate and empirical confidence intervals.

Synthetic data from real forecast

This benchmark example is closer to our problem and we expect the numerical results to be different from those obtained in the previous case. First, for each $p_i(t)$, we sample one path $X_i(t)$ to mimic the real data (the real forecast is affected by the weather conditions, the topology of each site, etc). Most importantly, for this example the number of available sets for 24 hours time horizon, is limited to 73 which is less that 2^8 paths that are previously sampled.

In this testing case, we generate data from real deterministic forecast paths, $p_i, i = 1, \ldots, J$. As the forecast is recorded every hour, we need to interpolate between the observation in order to obtain a continuous curve representing $p_j(t), j = 1, \ldots, J$. We use natural cubic splines interpolant where the number of knots used is equal to the number of observations per set (e.g., 24 in the case of 24 hours time horizon).

Figures (4.9-4.12) show that the true values of $(\theta, \alpha, \phi) = (0.2, 0.2, 0.05)$ are contained in the 95% confidence interval of MLE1. Although the confidence intervals are relatively larger than those obtained previously (as we just use 73 paths) and the estimated values, $(0.1865, 0.1970, 0.0549)$, are not exactly the true ones, MLE1 still provide reliable parameter estimates.

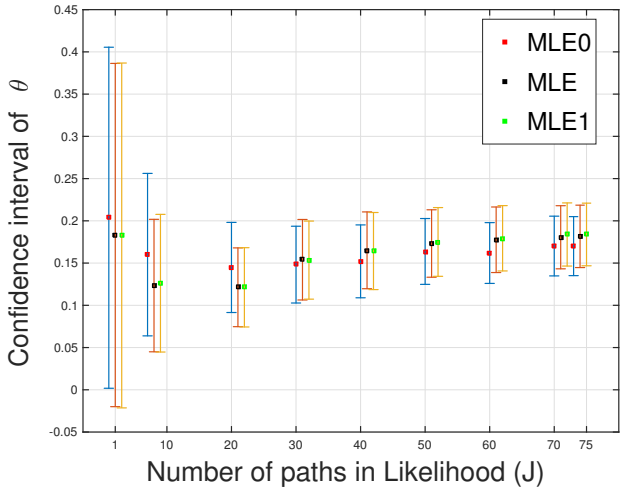

Figure 4.9: Convergence of θ for synthetic data with real forecast-Model 1

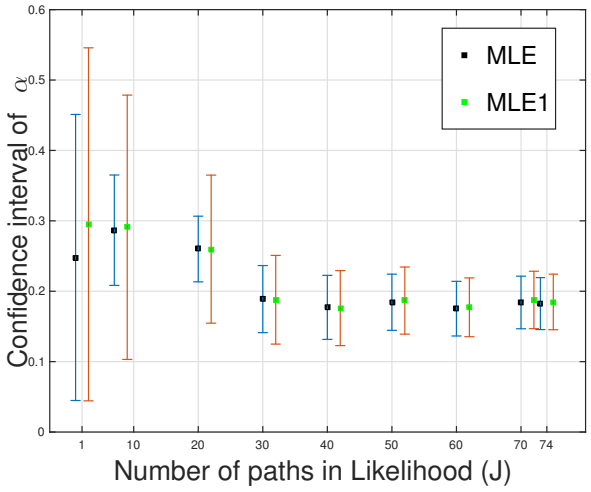

Figure 4.10: Convergence of α for synthetic data with real forecast-Model 1

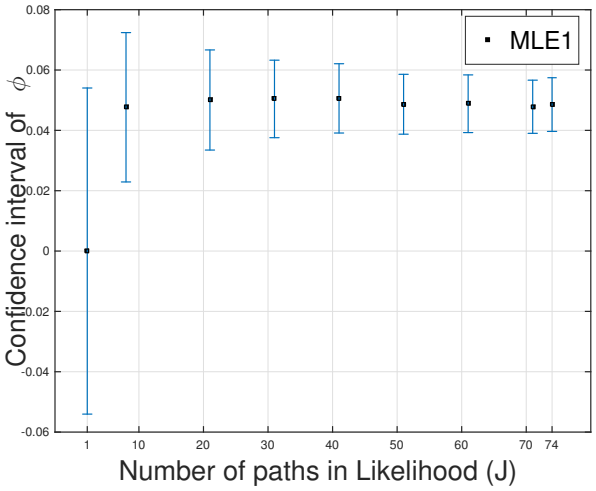

Figure 4.11: Convergence of ϕ for synthetic data with real forecast-Model 1

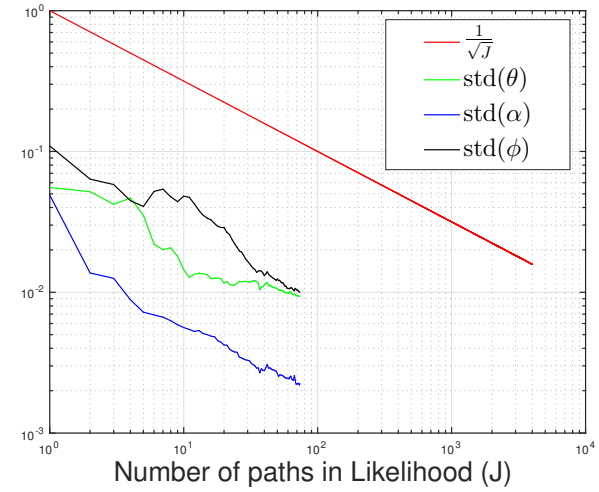

Figure 4.12: Rate of convergence of the maximum likelihood estimates

In Fig (4.13) and (4.14), we can see that the curves of the mean and the variance obtained with (θ, α, ϕ) are almost coincident with those obtained with $(\theta^*, \alpha^*, \phi^*)$.

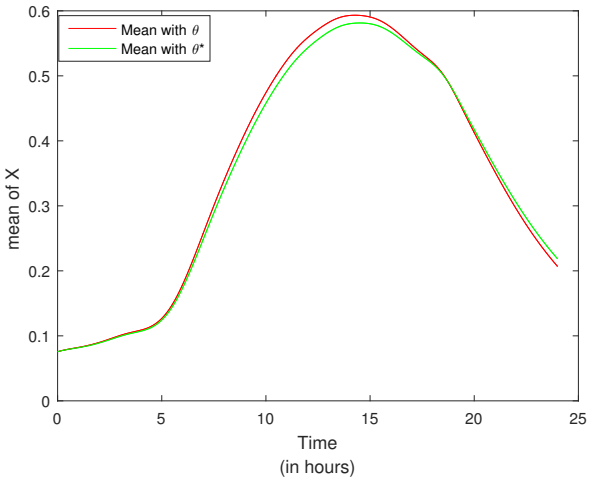

Figure 4.13: Mean with true and optimal parameters (March 15, 2016)

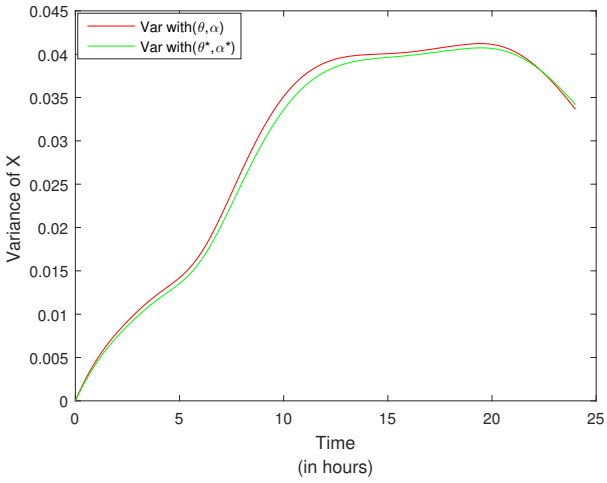

Figure 4.14: Variance with true and optimal parameters (March 15, 2016)

4.1.2 Model 2

For Model 2, we use the same true values, initial points and the number of paths. The only difference with Model 1 is the way we sample the synthetic data.

Trigonometric forecast

If we compare Figures (4.15-4.18) of Model 2 to Figures (4.1-4.4) of Model 1, we can see that for the same number of paths, Model 2 provides more precise confidence intervals and the asymptotic convergence of the estimates is perfectly shown in Fig (4.18).

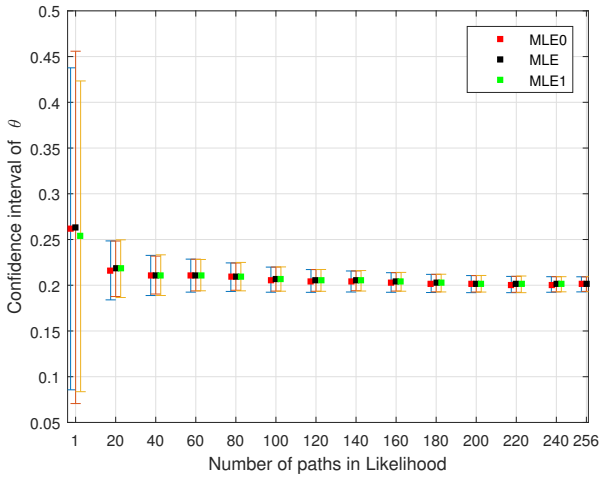

Figure 4.15: Convergence of θ for synthetic data with trigonometric forecast-Model 2

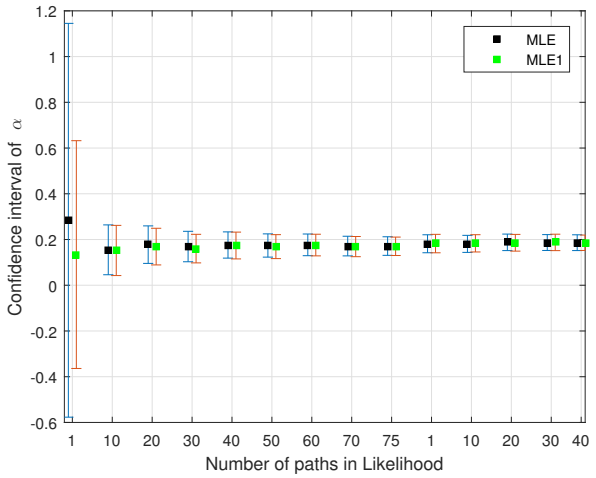

Figure 4.16: Convergence of α for synthetic data with trigonometric forecast-Model 2

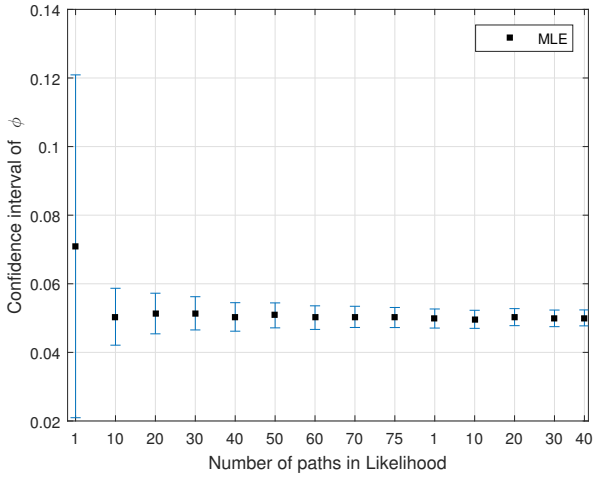

Figure 4.17: Convergence of ϕ for synthetic data with trigonometric forecast-Model 2

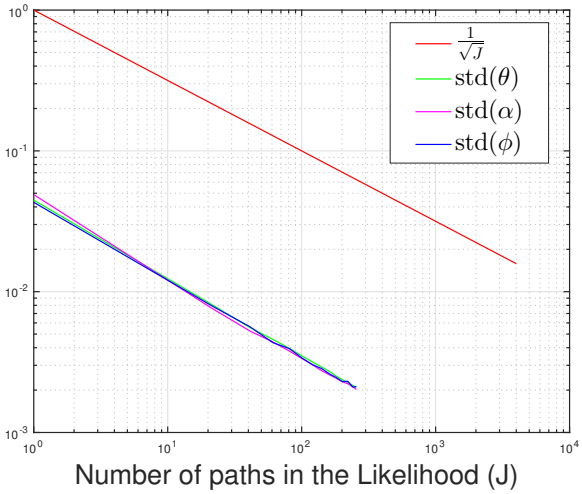

Figure 4.18: Rate of convergence of the likelihood estimates

Synthetic data from real forecast

In this benchmark example, we add the compensator parameter of Model 2 and we consider the maximum likelihood estimates of $(\theta, \alpha, \phi, c)$. We want to check the stability of the optimization problem and if it is able to recover the true values, $(0.2, 0.2, 0.05, 0.5)$, from the synthetic 73 paths.

Fig (4.19) and (4.21) prove that θ and ϕ fall within the 95% confidence intervals

which are shrinking every time we include more paths and centered around the true ones. However, MLE estimates of α and c are behaving differently, Fig (4.20) and (4.22) show that with 45 paths, the maximum likelihood estimates are almost equal to the true values. Then, after adding more paths the estimates depart farther and farther from the true values.

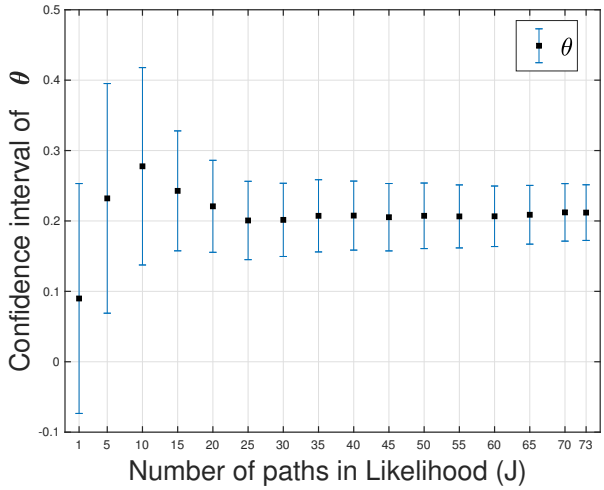

Figure 4.19: Convergence of θ for synthetic data with real forecast-Model 2

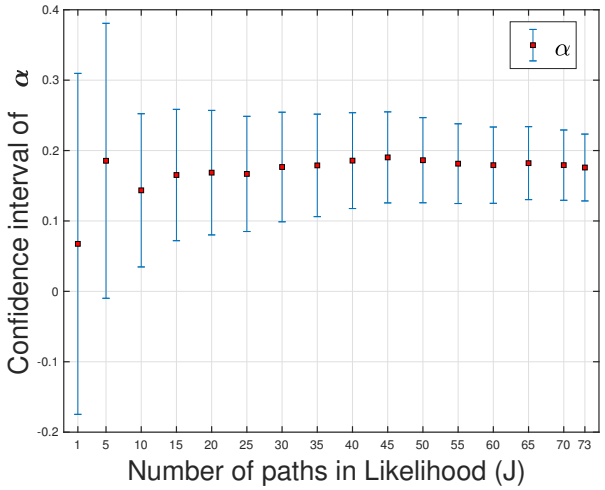

Figure 4.20: Convergence of α for synthetic data with real forecast-Model 2

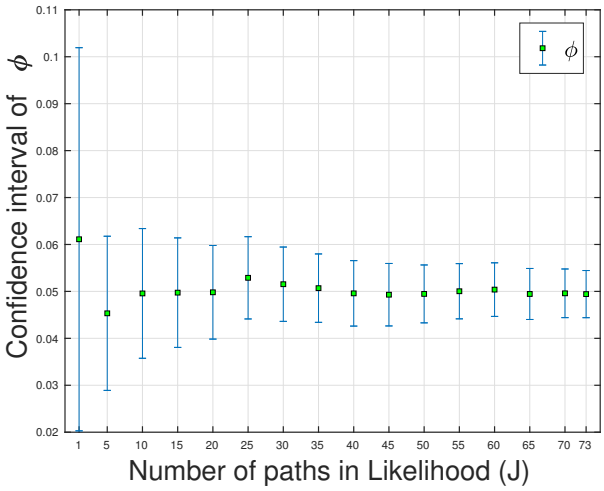

Figure 4.21: Convergence of ϕ for synthetic data with real forecast-Model 2

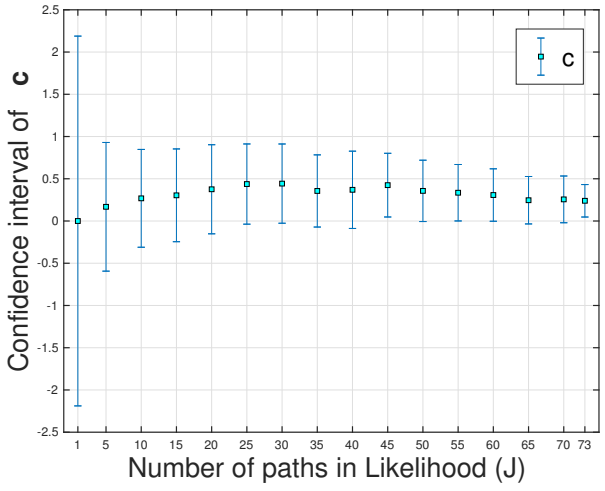

Figure 4.22: Convergence of c for synthetic data with real forecast-Model 2

4.2 Conclusion

The numerical results of the benchmark example based on the trigonometric forecast reveal that although both models are able to retrieve the true values, Model 2 is numerically more stable in terms of the length of the confidence intervals, rate of convergence, number of paths needed to converge to the true values.

The benchmark example based on the fitted numerical forecast show that the confidence intervals are relatively larger than those obtained in the previous example. This is a direct result of the restricted number of paths, but the maximum likelihood estimates are still reliable.

We also observe that the optimization problem based on Model 2 confuses between the compensator c and the diffusion parameter α. However, we checked for the case where $c = 0$, that this model is perfectly able to recover the rest of the true values (see appendix **??**).

Chapter 5

Stochastic forecasting of wind power in Uruguay

5.1 Introduction

In this chapter, we present the application of the models described in chapter 2 to the datasets described in chapter 1.

For both models, we consider a parametrization with respect to time-independent parameters. Next, we introduce potential trends of θ and select the best model according to the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) criteria. Then, to check variability of the estimates of the selected model, we use the bootstrapping method to compute optimal parameters with respect to randomly selected subsets of the data and present the bootstrap confidence intervals. To validate the models, we split the data into training and testing sets. We use the training sets for the optimization problem and use the optimal values to generate Empirical Confidence Bands (ECB). Finally, we validate the best model by verifying that the test set observations fall within the generated confidence bands. We also present a sensitivity analysis of the model parameters based on the Fisher Information Matrix (FIM) for the approximate likelihood and the Finite Difference method (FD) .

Preprocessing of the data

For the real data, if we choose a time horizon that is greater than 24 hours, we observe redundancy between successive sets. In fact, every 2 successive sets share the same values for a specific lag of time. This contradicts the independence assumption while

computing the likelihood function. For this reason, some sets should be removed, this is illustrated in Fig (5.1) for the case of 48 hours time horizon.

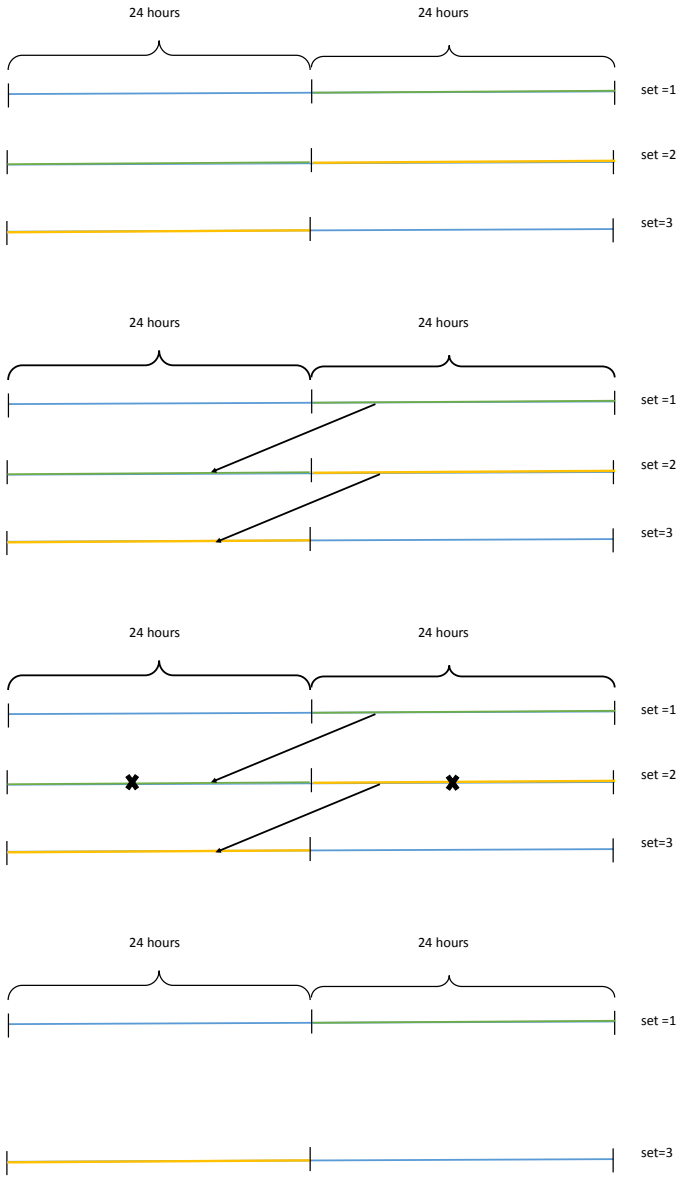

Figure 5.1: Transforming the data into independent sets

In our model, we expect the deterministic forecast to capture some features of the observed real production. Thus, we opt to discard the time series where a big difference

in absolute values between d_{ji} and p_{ji} (i.e. more than 20%) is observed, especially at time ($t = 1$). We consider them[1] as outliers with respect to our models, including them affects the quality of the parameter estimates and produces instabilities in the likelihood maximization problem (see appendix A).

5.1.1 Proposed models and model selection

The different parametrizations of Model 1 and Model 2 are summarized in Table 5.1.

Table 5.1: Proposed models with parametrization of the rate $\theta(t)$

Parametric Model	$\theta(t)$	Number of parameters $(\theta(t), \alpha, \phi)$
SDE0	θ_0	3
SDE1	$\theta_1(t) = \theta_0 e^{-\theta_1 t}$	4
SDE2	$\theta_2(t) = \theta_0 e^{-\theta_1 t} + \theta_2$	5
SDE3	$\theta_0 e^{-\theta_1 t} + \theta_2 e^{\theta_3 t}$	6

We assume that the observed production has non-correlated Gaussian measurement error for each set, $j = 1, \ldots, J$, $\epsilon^{(j)} \sim N(0, \phi^2 \mathbb{I})$, and we consider $\phi > 0$ a parameter of the model.

[1] These days are (1,2,4) of Mars 2016, (1,8) of April 2016 and (8,14, 17) of May 2016

Model 1

Table 5.2: Optimal parameter values for real data and different time horizons for Model 1.

Horizon	Parameters	Without noise	$\phi = 0.05$	$\phi = 0.02$	$\phi = 0.01$
12 h	$(\theta^*, \alpha^*, \phi^*)$	$(0.177, 0.082)$	$(0.256, 0.106, 0.036)$	$(0.200, 0.094, 0.004)$	$(0.184, 0.087, 0)$
	standard deviation	$10^{-2}(0.83, 0.41)$	$10^{-2}(1.75, 0.7, 0.25)$	$10^{-2}(0.91, 0.44, 0.44)$	$10^{-2}(0.87, 0.44, 0.23)$
18 h	$(\theta^*, \alpha^*, \phi^*)$	$(0.140, 0.089)$	$(0.213, 0.102, 0.038)$	$(0.163, 0.102, 0.010)$	$(0.1458, 0.094, 0)$
	standard deviation	$10^{-2}(0.83, 0.41)$	$10^{-2}(1.2, 0.59, 0.19)$	$10^{-2}(0.68, 0.45, 0.19)$	$10^{-2}(0.58, 0.40, 0.23)$
24h	$(\theta^*, \alpha^*, \phi^*)$	$(0.140, 0.087)$	$(0.196, 0.118, 0.036)$	$(0.160, 0.098, 0.002)$	$(0.143, 0.091, 0)$
	standard deviation	$10^{-2}(0.47, 0.31)$	$10^{-2}(0.96, 0.57, 0.17)$	$10^{-2}(0.57, 0.37, 0.48)$	$10^{-2}(0.49, 0.33, 0.24)$
36h	$(\theta^*, \alpha^*, \phi^*)$	$(0.121, 0.109)$	$(0.172, 0.147, 0.034)$	$(0.139, 0.125, 0.004)$	$(0.130, 0.111, 0)$
	standard deviation	$10^{-2}(0.47, 0.47)$	$10^{-2}(0.96, 0.57, 0.17)$	$10^{-2}(0.52, 0.51, 0.36)$	$10^{-2}(0.50, 0.47, 0.20)$
48h	$(\theta^*, \alpha^*, \phi^*)$	$(0.114, 0.105)$	$(0.154, 0.156, 0.035)$	$(0.126, 0.124, 0.004)$	$(0.126, 0.111, 0)$
	standard deviation	$10^{-2}(0.39, 0.39)$	$10^{-2}(0.73, 0.76, 0.17)$	$10^{-2}(0.44, 0.47, 0.32)$	$10^{-2}(0.41, 0.40, 0.16)$

In Table 5.2, we present the optimal values of the parametric model SDE0, (θ, α, ϕ) for the time horizons $\{12, 18, 24, 36, 48\}$ based on Model 1. In column termed 'Without noise' we report the output of the minimization of the negative logarithm of the likelihood function (2.4) using the real data to test whether these values can be retrieved from the optimization problem.

We observe that Model 1 is not able to recover the value of ϕ, and the output is always zero. Thus we simply report values of θ and α. To investigate the effect of this observation, we added artificial Gaussian noise to the data with $\phi = \{0.05, 0.02, 0.01\}$. We can see that the values of θ and α change noticeably and appear to be sensitive to the measurement error. For $\phi = 0.05$, the likelihood is able to retrieve a value of order 0.036 and the value of α jumps from 0.082 to 0.106, which suggests that it is difficult for our model to distinguish between the measurement error and the noise coming from the SDE through α. Part of the noise, ϕ, is indeed absorbed by α, which explains the jump in its value. This observation is confirmed by the last two columns of Table 5.2, where the likelihood is no more able to retrieve ϕ with noise on the order of 0.02 and 0.01.

We observe that θ is decreasing with time, which indicates that the predictability of the model becomes less reliable for increasing time horizons. This motivates the use of different exponentially decreasing parametric functions for θ in Table 5.1. SDEi, $i = \{0, 1, 2, 3\}$ denote the different parametrizations $(\theta_i(t), \alpha, \phi)$ such that $\theta_0(t) = \theta_0$ is a constant rate, $\theta_1(t)$ is a one-term exponential function, $\theta_2(t)$ has an asymptotic positive lower bound given by θ_2 and $\theta_3(t)$ is a linear combination of exponentially decreasing and increasing functions to better fit the data (we expect θ_2 to be very small such that the increasing term with dominate $\theta(t)$).

To obtain initial values, that match the behavior of $\theta(t)$ for the optimization with time varying $\theta(t)$ (i.e SDE1, SDE2, and SDE3). We used the fitted values of constant θ^* obtained in table 5.2 with the 'Curve Fitting Toolbox'in MATLAB. The interactive tool allows us to interpolate between the points, generate a fit and evaluate the quality of the resulting fit using metrics such as adjusted R squared. We then use the fitted coefficients of $(\theta_0, \theta_1, \theta_2, \theta_3)$ as initial points for the maximum likelihood estimators of $(\theta(t), \alpha, \phi)$. The MLE parameters for all models are presented in Table 5.3 for the 24-hour time horizon.

Table 5.3: Optimal parameters for a 24-hour time horizon for the parametric SDE in Table 5.1 for Model 1

Model	Initial values pf $\boldsymbol{\theta}$	Optimal values	Standard deviation
SDE0	$(0.2, 0.4)$	$(0.140, 0.087)$	$10^{-2}(0.47, 0.31)$
SDE1	$(0.202, 0.013, 0.4)$	$(0.199, 0.034, 0.141)$	$10^{-2}(0.68, 0.15, 0.53)$
SDE2	$(0.124, 0.047, 0.099, 0.4)$	$(0.137, 0.108, 0.087, 0.125)$	$10^{-2}(2.09, 1.33, 0.34, 0.46)$
SDE3	$(0.2, 0.01, 0.002, 0.02, 0.1)$	$(0.28, 0.103, 10^{-3}, 0.363, 0.018)$	$10^{-2}(2.06, 1.04, 4 \cdot 10^{-4}, 0.16, 0.05)$

Among the four candidate models, Table 5.4 indicates that SDE1 is the 'best'model as it has the smallest AIC and BIC values.

Table 5.4: Model selection

Criterion	df	AIC	BIC
SDE0	3	−8418	−8411
SDE1	4	−8427	−8418
SDE2	5	−8426	−8414
SDE3	6	−8273	−8259

To visualize the behavior of the estimated parameters of the selected model while including more paths in the likelihood , we present Figures (5.2,5.3,5.4). We can see that the value of θ_0 is relatively more stable than θ_1 and α. In fact, θ_0 converges to the value 0.2 within 40 paths, which suggests that the first 40 sets are enough informative about θ_0. In contrast, θ_1 and α have a tendency to increase and including more data may make a difference in the estimation.

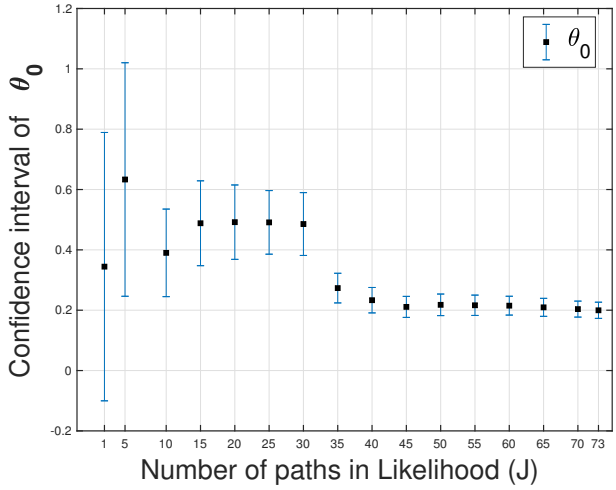

Figure 5.2: Convergence w.r.t to number of paths of θ_0 in SDE1.

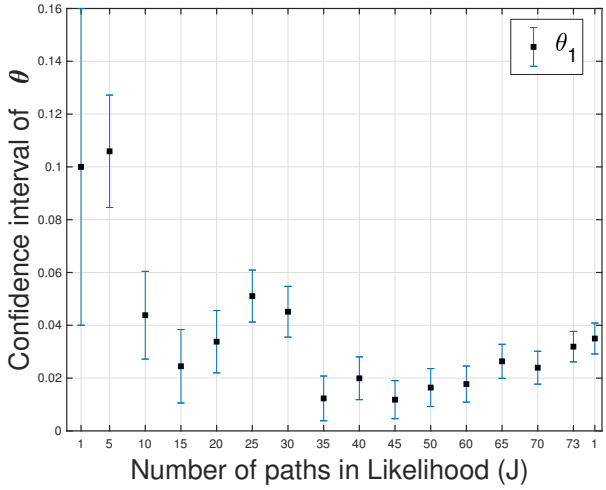

Figure 5.3: Convergence w.r.t to number of paths of θ_1 in SDE1.

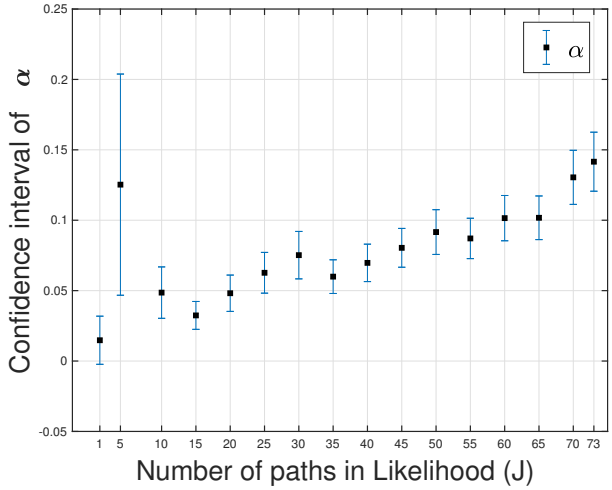

Figure 5.4: Convergence w.r.t to number of paths of α in SDE1.

Model 2

Next, we present the best model among those described in Table 5.1 based on Model 2 following the same steps as for Model 1.

Table 5.5: Optimal parameter values for real data and different time horizons for the parametric SDE in Table 5.1 for Model 2.

Horizon	Parameters	Estimates
12 h	$(\theta^*, \alpha^*, \phi^*)$	$(0.138, 0.108, 0.087)$
	standard deviation	$10^{-2}(0.79, 1.87, 0.07)$
18 h	$(\theta^*, \alpha^*, \phi^*)$	$(0.127, 0.569, 0.003)$
	standard deviation	$10^{-2}(0.41, 1.49, 0.04)$
24h	$(\theta^*, \alpha^*, \phi^*)$	$(0.115, 0.602, 0.03)$
	standard deviation	$10^{-2}(0.28, 1.16, 0.03)$
36h	$(\theta^*, \alpha^*, \phi^*)$	$(0.095, 1.008, 0.019)$
	standard deviation	$10^{-2}(0.39, 4.52, 0.12)$
48h	$(\theta^*, \alpha^*, \phi^*)$	$(0.087, 0.993, 0.016)$
	standard deviation	$10^{-2}(0.31, 3.98, 0.09)$

From Table 5.5, we see that θ exhibits the same behavior as in Model 1; it decreases with longer time horizon. We also observe that Model 2 is able to capture the measurement error, using the same datasets, and the optimal values are consistent with our findings from Model 1 (i.e. the result align with the anticipated order of α for 24 hours). The estimated value of θ_2 in SDE2 and SDE3 is exactly zero which

leads us again to SDE1.

Table 5.6: Optimal parameters for a 24-hour time horizon for Model 2

Model	Initial values	Optimal values	Standard deviation
SDE0	$(0.4, 0.2, 0.1)$	$(0.115, 0.602, 0.03)$	$10^{-2}(0.28, 1.16, 0.03)$
SDE1	$(0.161, 0.013, 0.2, 0.1)$	$(0.218, 0.033, 0.453, 0.033)$	$10^{-2}(0.79, 0.33, 1.27, 0.05)$
SDE2	$(0.114, 0.013, 0.061, 0.2, 0.1)$	$(0.218, 0.033, 0, 0.453, 0.033)$	$10^{-2}(0.79, 0.32, 0.57, 1.26, 0.05)$
SDE3	$(0.1673, 0.0157, 6 \cdot 10^{-8}, 0.245, 0.2, 0.1)$	$(0.218, 0.033, 0, 0.018, 0.453, 0.033)$	$10^{-2}(0.81, 0.32, 0.02, 1.39, 1.04, 0.05)$

Table 5.7 confirms that SDE1 is the best model with respect to the AIC and BIC criterion.

Table 5.7: Model selection

Criterion	AIC	BIC
SDE0	-6410	-6403
SDE1	-7311	-7302
SDE2	-7309	-7297
SDE3	-7310	-7296

Figures (5.5 - 5.6) represent the estimated parameters of SDE1 with respect to the number of paths in the likelihood function. A comparison with the figures obtained in Model 1 shows that θ_0 exhibits the same behavior for both models (i.e. after 40 sets its value is almost constant 0.2). Although Model 1 and 2 end up with the same estimation of $\theta_1 \approx 0.03$, the trajectory of the the latter with Model 2 looks more stable (i.e. we are able to retrieve the same value using fewer paths). For ϕ and α, we observe that their curves reach a certain range ($\alpha \approx 0.5, \phi \approx 0.02$), then, a jump occur for the last two added paths.

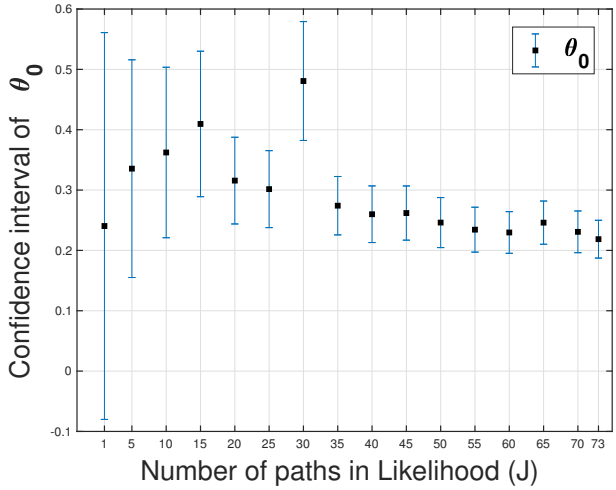

Figure 5.5: Convergence w.r.t to number of paths of θ_0 in SDE1.

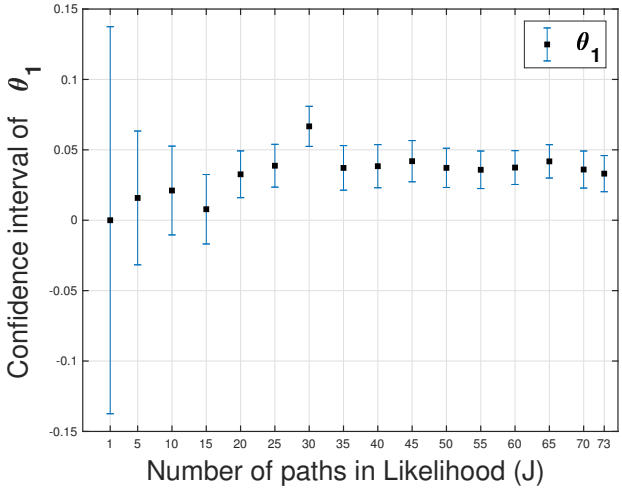

Figure 5.6: Convergence w.r.t to number of paths of θ_1 in SDE1.

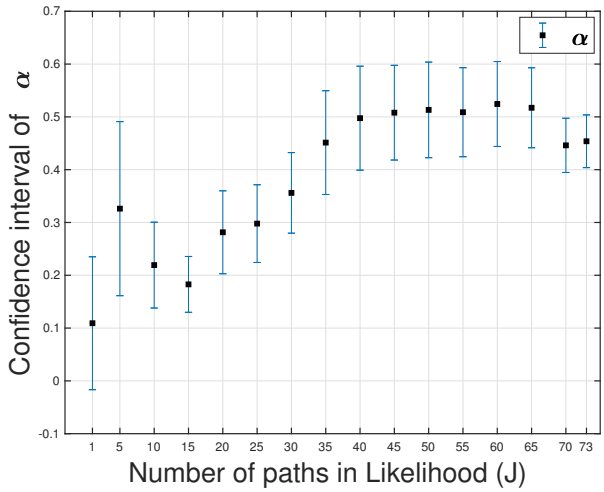

Figure 5.7: Convergence w.r.t to number of paths of α in SDE1.

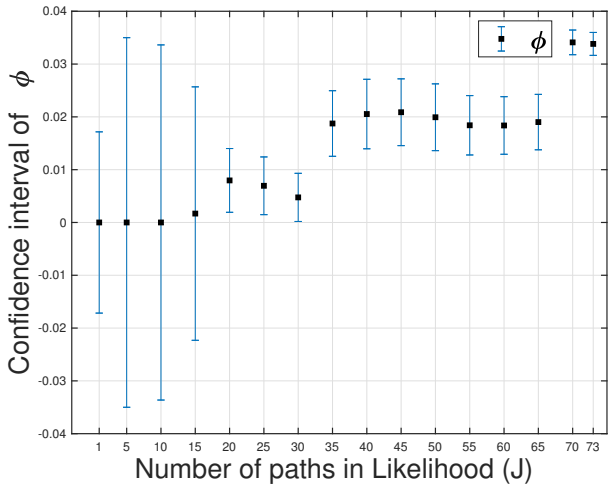

Figure 5.8: Convergence w.r.t to number of paths of ϕ in SDE1.

5.1.2 Variability, validation and sensitivity analysis

An important issue to explore is the variability of the datasets. Here, we use the bootstrapping method to select randomly 80% of the data and compute statistics for

the estimated parameters from the bootstrap samples. We compute the sample mean and the bootstrap confidence intervals of each parameter. Figures (5.9) and (5.10) show that our data are rich enough to predict a reliable set of estimates such that the bootstrap histograms are centered around the MLE estimates. Also, we notice that both models provide similar values for the sample mean and sample confidence intervals.

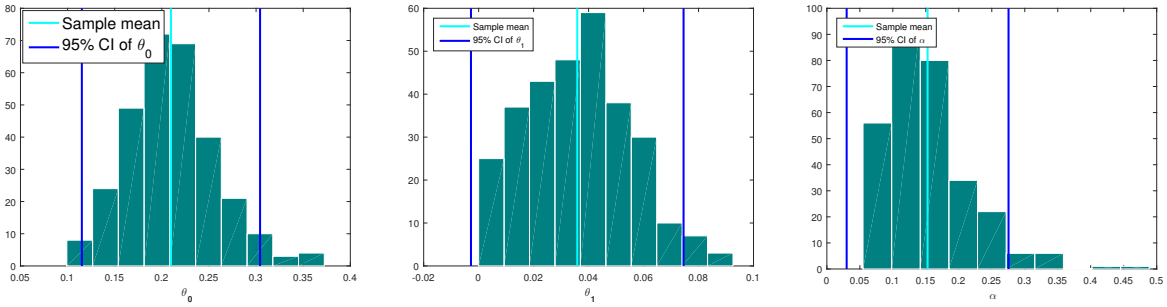

Figure 5.9: Bootstrap histograms of θ_0, θ_1 and α based on $n = 200$ iterations for Model 1

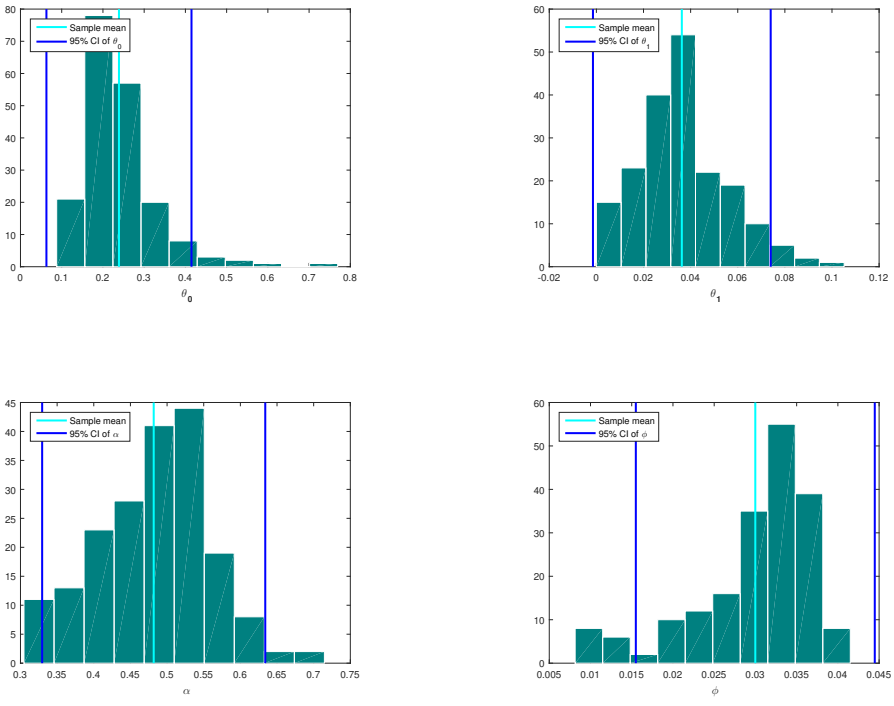

Figure 5.10: Bootstrap histograms of θ_0, θ_1, α and ϕ based on $n = 200$ iterations for Model 2

Fig (5.11) shows the curves of $\theta(t)$ obtained with the optimal values (θ_0^*, θ_1^*) of Model 1 and Model 2. We also plot the resulting 95% confidence intervals (i.e computed using the hessian of the log likelihood function). We can see that the predicted curve of Model 1 falls inside the prediction interval of Model 2 for most of the time, confirming that both models predict the same rate.

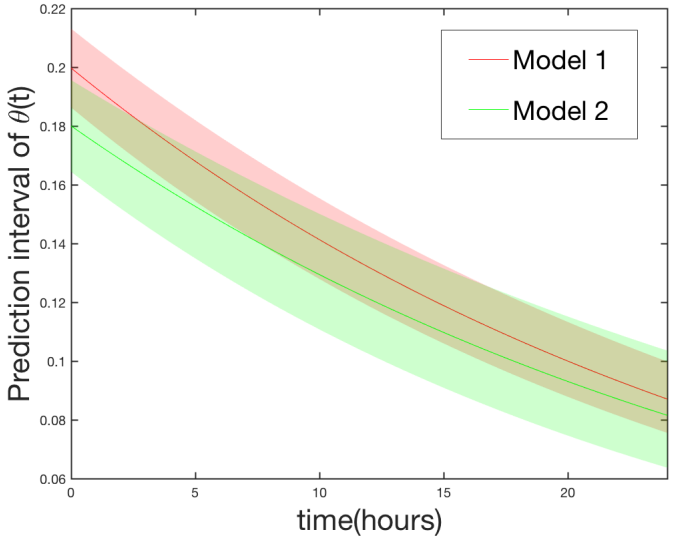

Figure 5.11: Best fit for SDE1 for the two models.

To validate the selected models we present the empirical path density and the confidence bands for two days in the training set as shown in Figures (5.12 - 5.15) We generate the empirical path density from the simulated paths of (Model 1) and (Model 2) using the MLE estimators and Euler-Maruyama discretization scheme. On the same graphs, we present the real (historical) wind energy production and the numerical forecast. The figures verify that the real production falls into the 95% confidence band for most of the sets.

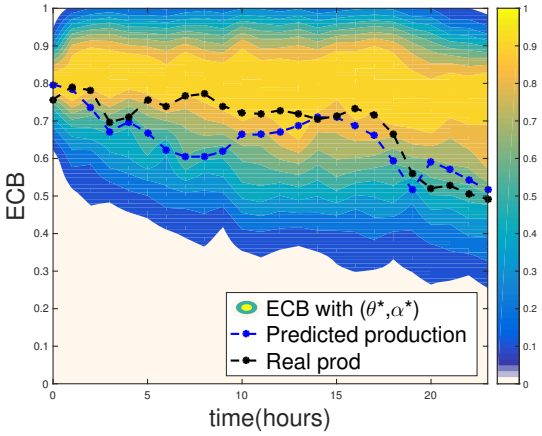

Figure 5.12: ECB and real production on May 16, 2016 for Model 1

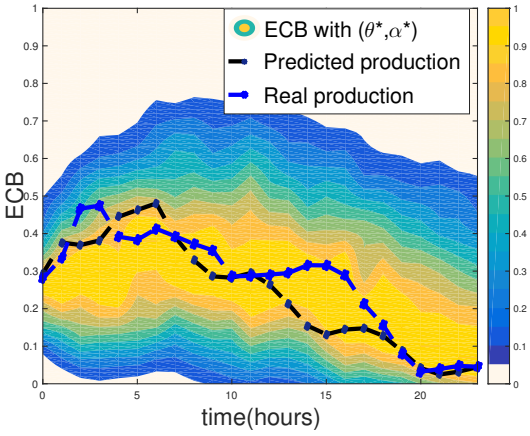

Figure 5.13: ECB and real production on May 22, 2016 for Model 1

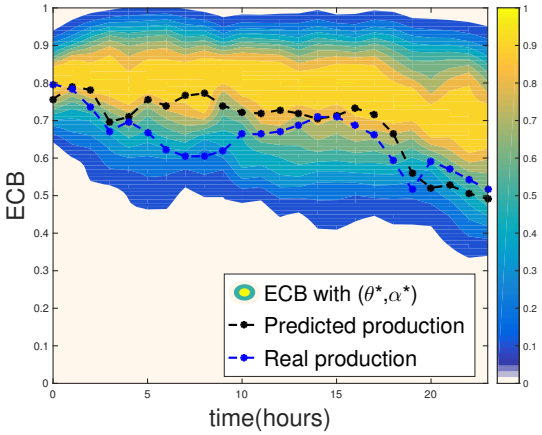

Figure 5.14: ECB and real production on May 16, 2016 for Model 2

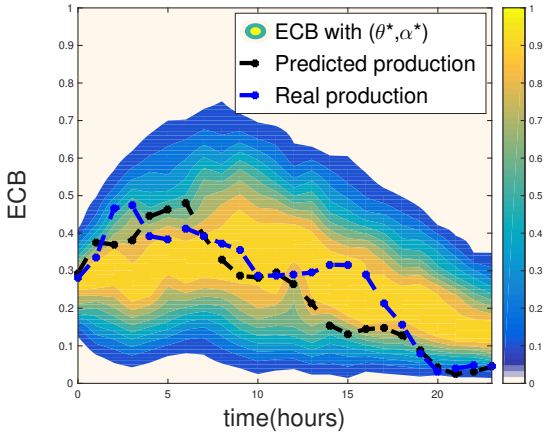

Figure 5.15: ECB and real production on May 22, 2016 for Model 2

Next, we analyze the sensitivity of both models to the parameters $(\theta_0, \theta_1, \alpha)$ using the Finite Difference indices S_{BD}, S_{FD} and S_{CFD} [14]. A direct sensitivity analysis of the models based on the stochastic differential equations (Model 1) and (Model 2) demonstrate that the most sensitive parameter for both models is θ_1, which controls the rate of decay of $\theta(t)$ with respect to the time horizon. Note that the least sensitive parameter in the two models is α.

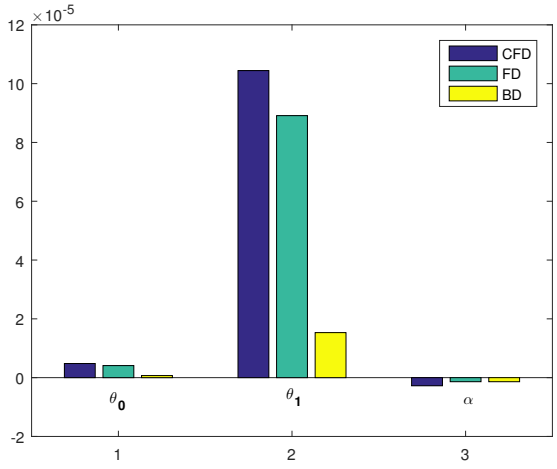

Figure 5.16: Sensitivity of parameters for Model 1 (FD).

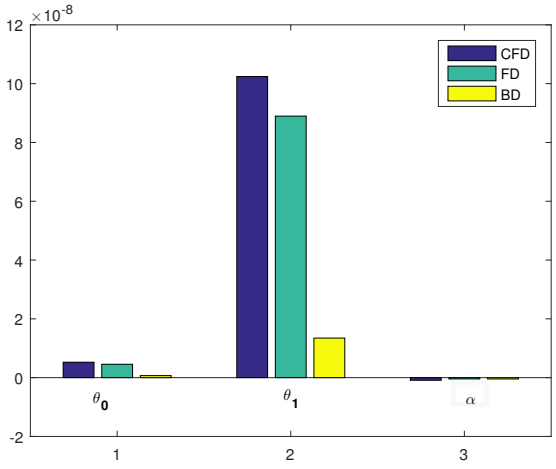

Figure 5.17: Sensitivity of parameters for Model 2 (FD).

5.2 Conclusion

The motivation of our work is the lack of accuracy in the deterministic forecast of wind power generation, in particular the datasets we are provided with. In fact, the dynamic of the wind as a source of energy is very complex, which makes a reliable prediction extremely difficult if not impossible. Our purpose is to generate a probabilistic forecast that includes the uncertainty inherited in the deterministic prediction which allows us to generate many scenarios with estimated confidence bands. This type of application is crucial for stochastic optimization problems where decisions makers need to do a trade of between different energy sources, in order to minimize the cost of generating electricity in the market .

It is important to point out that we based our study on the benchmark examples in chapter 4. The numerical results of these examples showed that Model 2 is numerically more stable.

With real data, Model 1 is not able to capture the value of ϕ for all $\theta_i(t),\ i \in \{1, 2, 3\}$ and its parameters are very sensitive to an artificial measurement error. On the other hand Model 2 provides an estimate of the measurement error and we assume that this is related to the stability of Model 2 due to the use of Lamperti-transformed process. However, it is remarkable that both models predict the same rate.

The variability analysis verifies that the available datasets are rich enough to predict an informative model. Moreover, for both models the most sensitive parameter is θ_1 which controls the decay. The least sensitive parameter is α, the measurement noise parameter.

Chapter 6

Partially observed model stochastic process

Introduction

In this chapter we derive the Gaussian approximation for a multidimensional stochastic process and we introduce the notion of partially observed process. The parameters inference is then based on the marginal observed process. This approach will be applied to our problem of Probabilistic Forecast of Wind Power Production where we assume that we can not keep track of the aggregated power production in all the installed farms in the country. We do not present numerical results for this chapter because we do not have access to real datasets.

6.1 Partially Observed model

We say that a model is fully observed, if all the components of the process, $X = (X_1, X_2, \cdots, X_N)$, are observed. In a partially observed model, we assume that we have a subset of missing data within realization of the process, X.

In our application, we are not always able to have access to the real aggregated production and the numerical forecast of all wind power plants in the country. For example some plants are under construction, non equipped with measurement devices or new plants that are not working with full capacity. Thus, we consider a two dimensional process, $X = \{Y, Z\}$, only a part $\{Y\}$ of which is actually observed.

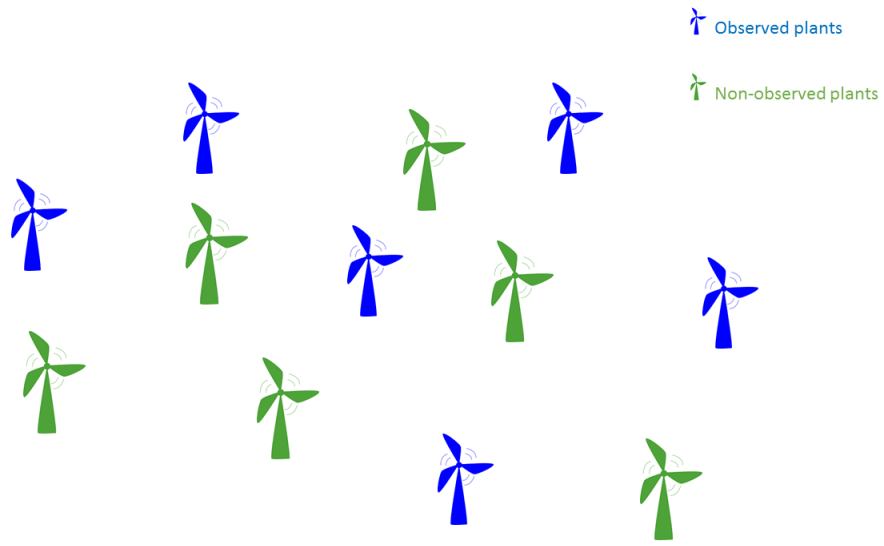

Figure 6.1: Observed and non observed plants

Assume that $Y(t)$ and $Z(t)$ satisfies respectively the following SDEs

$$dY(t) = b_1(Y(t);\boldsymbol{\theta})dt + \sigma_1(Y(t);\boldsymbol{\theta})dW_1(t)$$
$$dZ(t) = b_2(Z(t);\boldsymbol{\theta})dt + \sigma_2(Z(t);\boldsymbol{\theta})dW_2(t)$$

The coupling of the two SDEs is done through the diffusion term. However both SDEs have specific drift with different rates $\{\theta_1(t)\ \theta_2(t)\}$ and numerical forecast $\{p_1(t)\ p_2(t)\}$. By analogy to the the uni-dimensional problem, the drift functions are given by

$$b_1(Y(t);\boldsymbol{\theta}) = -\theta_1(t)(Y(t) - p_1(t))$$
$$b_2(Z(t);\boldsymbol{\theta}) = -\theta_2(t)(Z(t) - p_2(t))$$

and the diffusion functions are

$$\sigma_1(Y(t);\boldsymbol{\theta}) = \sqrt{2\theta_1(t)\alpha Y(t)(1-(Y(t)))}$$
$$\sigma_2(Z(t);\boldsymbol{\theta}) = \sqrt{2\theta_2(t)\alpha Z(t)(1-(Z(t)))}$$
$$\sigma(X(t),Y(t)) = \begin{bmatrix} \sigma_1(Y(t)) & 0 \\ 0 & \sigma_2(Z(t)) \end{bmatrix}$$

where $dW_1(t)\cdot dW_2(t) = \rho dt$ and $\boldsymbol{\theta} = (\theta_1(t), \theta_2(t), \alpha_1, \alpha_2, \rho)$. Inference about partially observed model is based on the marginal distribution of Y and we expect, the optimal value of $\boldsymbol{\theta}$ given by the MLE of the partially observed data to be different than the values obtained from the full model.

6.2 Two-moment equations for multidimensional process

In this section we will derive the Gaussian approximation for a multidimensional process, $X(t) = (X_1(t), \cdots, X_N(t)) \in \mathbb{R}^N$ where the i-th component satisfies the following SDE

$$\begin{cases} dX_i(t) = b_i(X(t);\boldsymbol{\theta})dt + \sum_{j=1}^{N} \sigma_{ij}(X(t),\boldsymbol{\theta})dW_j(t), \text{ for all } i \in \{1,\cdots,N\} \\ X_i(0) = x_0(i) \end{cases} \tag{6.1}$$

where $b_i(X(t);\boldsymbol{\theta})$ and $\sigma_{ij}(X(t),\boldsymbol{\theta})$ are two known regular functions such that a solution of (6.1) exists and is unique. $W_i(t)$ are independent Brownian motions , for all $i \in \{1,\cdots,N\}$.

Define, $\mu_i(t) = \mathbb{E}[X_i(t)]$, $v_i(t) = \mathbb{E}[\big(X_i(t)-\mu_i(t)\big)^2]$ and $v_{ij}(t) = E[\big(X_i(t)-\mu_i(t)\big)\big(X_j(t)-\mu_i(t)\big)]$ to be the mean, variance and covariance, respectively, for all $i,j \in \{1,\cdots,N\}$. For ease of notation, in the derivation of these equations we remove the dependence

on time and reconsider it in the final formulation, we also use $b_i(X(t)), \sigma_{ij}(X(t))$ and assume implicitly that they depend on $\boldsymbol{\theta}$.

The equation of the mean is given by

$$\frac{d\mathbb{E}[X_i]}{dt} = \mathbb{E}[b_i(X)]$$

Expanding $b_i(X)$ around the mean μ gives

$$\begin{aligned} b_i(X) = b_i(\mu) + \sum_j \frac{\partial b_i(\mu)}{\partial X_j}(X_j - \mu_j) + 0.5 \sum_{j,k} \frac{\partial^2 b_i(\mu)}{\partial X_j \partial X_k}(X_j - \mu_j)(X_k - \mu_k) \\ + \sum_{j,k,l} \rho_{ijkl}(X_j - \mu_j)(X_k - \mu_k)(X_l - \mu_l) \end{aligned} \tag{6.2}$$

where $\rho_{ijkl} = 0.5 \int_0^1 (1-\xi)^2 \frac{\partial^3 b_i}{\partial X_j \partial X_k \partial X_l}(\mu + \xi(X - \mu))d\xi$, the remainder of the Taylor series in integral form.

Taking the expectation of (6.2), we obtain the following equation of the mean

$$\frac{d\mu_i(t)}{dt} = \mathbb{E}[b_i(X)] = b_i(\mu(t)) + 0.5 \sum_{j,k} \frac{\partial^2 b_i(\mu(t))}{\partial X_j \partial X_k} v_{jk}(t) + \eta_1 \tag{6.3}$$

where $\eta_1 = \sum_{j,k,l} \rho_{ijkl} \mathbb{E}[(X_j - \mu_j)(X_k - \mu_k)(X_l - \mu_l)]$.

To obtain the equation of the covariance, we apply Dynkin's formula to $f(X) = \big(X_i - \mu_i\big)\big(X_j - \mu_j\big)$ for all $i, j = 1..n$, (appendix, A.2.3).

For that, we need to compute the following derivatives

$$\begin{aligned} &\frac{\partial f}{\partial X_l} = \big(X_i - \mu_i\big)\delta_{jl} + \big(X_j - \mu_j\big)\delta_{il}, \text{ with } \delta_{ij} = 1 \text{ for } i = j, \ 0 \text{ otherwise} \\ &\frac{\partial^2 f}{\partial X_k \, \partial X_l} = \delta_{ik}\delta_{jl} + \delta_{jk}\delta_{il} \end{aligned}$$

Then, equation (A.2.3) gives

$$\mathcal{L}f(X) = b_j(X)(X_i - \mu_i) + b_i(X)(X_j - \mu_j) + \frac{1}{2}\big[\sigma\sigma^\top\big]_{i,j}(X) + \frac{1}{2}\big[\sigma\sigma^\top\big]_{j,i}(X)$$

Since, $\big[\sigma\sigma^\top\big]_{i,j}(X) = [\sigma\sigma^\top]_{j,i}(X)$, for all $i, j \in \{1, \cdots, N\}$, we end up with

$$\mathbb{E}[\mathcal{L}f(X)] = \mathbb{E}[b_j(X)(X_i - \mu_i)] + \mathbb{E}[b_i(X)(X_j - \mu_j)] + \mathbb{E}\Big[\big[\sigma\sigma^\top\big]_{i,j}(X)\Big]$$

As we are interested in the second-order moment of X we use the first-order expansion of b

$$b_i(X) = b_i(\mu) + \sum_k \frac{\partial b_i(\mu)}{\partial X_k}(X_k - \mu_k) + \sum_{i,k,l} \rho_{ikl}(X_k - \mu_k)(X_l - \mu_l) \tag{6.4}$$

where $\rho_{ikl} = \int_0^1 (1-\xi)^2 \frac{\partial^2 b_i}{\partial X_k \partial X_l}(\mu + \xi(X - \mu))d\xi$

Multiplying (6.4) by $(X_j - \mu_j)$ and taking the expectation we obtain

$$\mathbb{E}[b_i(X)(X_j - \mu_j)] = \sum_k \frac{\partial b_j}{\partial X_k}\Big(\mu\Big)v_{kj} + \eta_2$$
$$\mathbb{E}[b_j(X)(X_i - \mu_i)] = \sum_k \frac{\partial b_j}{\partial X_k}\big(\mu\big)v_{ki} + \eta_3, \text{ derived in a similar way}$$

where

$$\eta_2 = \sum_{k,l} \rho_{ikl}\mathbb{E}[(X_j - \mu_j)(X_k - \mu_k)(X_l - \mu_l)]$$
$$\eta_3 = \sum_{k,l} \rho_{jkl}\mathbb{E}[(X_i - \mu_i)(X_k - \mu_k)(X_l - \mu_l)]$$

Moreover, expanding the function $\big[\sigma\sigma^\top\big]_{i,j}(X)$ and taking its expectation, we obtain

$$\mathbb{E}\Big[\big[\sigma\sigma^\top\big]_{i,j}(X)\Big] = \big[\sigma\sigma^\top\big]_{i,j}(\mu) + \frac{1}{2}\sum_{kl} \frac{\partial^2\big[\sigma\sigma^\top\big]_{i,j}}{\partial X_k X_l}\Big(\mu\Big) \cdot v_{kl} + \eta_4$$

where

$$\eta_4 = \sum_{klf} R_{ij,klh}\mathbb{E}[(X_k\mu_k)(X_l-\mu_l)(X_h-\mu_h)]$$

$$R_{ij,klh} = 0.5\int_0^1 (1-\xi)^2 \frac{\partial^3\left[\sigma\sigma^\top\right]_{i,j}}{\partial X_k \partial X_l \partial X_h}(\mu+\xi(X-\mu))d\xi$$

Hence the equation of the covariance can be written in the following form

$$\frac{dv_{ij}}{dt} = \left[\sigma\sigma^\top\right]_{i,j}(\mu)+\sum_k \frac{\partial b_j}{\partial X_k}\Big(\mu\Big)v_{kj}+\sum_k \frac{\partial b_j}{\partial X_k}\Big(\mu\Big)v_{ki}+\frac{1}{2}\sum_{kl}\frac{\partial^2\left[\sigma\sigma^\top\right]_{i,j}}{\partial X_k X_l}\Big(\mu\Big)v_{kl}+\eta_1+\eta_2+\eta_4 \tag{6.5}$$

Thus, the Gaussian approximation for a multidimensional process, $X(t)$, can be written as

$$\begin{cases} \frac{d\mu_i(t)}{dt} = b_i(\mu(t)) + 0.5\sum_{j,k}\frac{\partial^2 b_i}{\partial X_j \partial X_k}\Big(\mu(t)\Big)v_{jk}(t) + \eta(t) \\ \frac{dv_{ij}(t)}{dt} = \sum_k \frac{\partial b_i}{\partial X_k}\Big(\mu(t)\Big)v_{kj}(t) + \sum_k \frac{\partial b_j}{\partial X_k}\Big(\mu(t)\Big)v_{ki}(t) + \left[\sigma\sigma^\top\right]_{i,j}\Big(\mu(t)\Big) \\ \qquad +\frac{1}{2}\sum_{kl}\frac{\partial^2\left[\sigma\sigma^\top\right]_{i,j}}{\partial X_k X_l}\Big(\mu(t)\Big)v_{kl} + \eta_2(t) + \eta_3(t) + \eta_4(t) \end{cases}$$

To close our system, we consider a second-order approximation of the mean and the covariance function. The resulting equations are given by

$$\begin{cases} \frac{d\mu_i(t)}{dt} = b_i(\mu(t)) + 0.5\sum_{j,k}\frac{\partial^2 b_i}{\partial X_j \partial X_k}\Big(\mu(t)\Big)v_{jk}(t) \\ \frac{dv_{ij}(t)}{dt} = \left[\sigma\sigma^\top\right]_{i,j}(\mu(t)) + \sum_k \frac{\partial b_i}{\partial X_k}\Big(\mu(t)\Big)v_{kj}(t) + \sum_k \frac{\partial b_j}{\partial X_k}\Big(\mu(t)\Big)v_{ki}(t) \\ \qquad +\frac{1}{2}\sum_{kl}\frac{\partial^2\left[\sigma\sigma^\top\right]_{i,j}}{\partial X_k X_l}\Big(\mu(t)\Big)v_{kl}(t) \end{cases}$$

6.3 Conclusion

In this work we proposed a probabilistic forecast of wind power generation based on parametrized logistic-type SDEs. First, we considered a fully observed model where the available datasets represent the aggregate historical power production and numerical forecast of all farms in the country. Then, we considered a partially observed model, where available datasets correspond to a subset of all farms.

To infer the SDE parameters we considered two approaches: inference based on state-dependent diffusion and inference based on state-independent diffusion through the use of Lamperti transform. Using Gaussian approximation we developed a set of ODEs for the second-order moment and we used the approximate likelihood method. We applied our approaches to two benchmark examples and Uruguayan data for the period of March 1 to May 3, 2016. The numerical results reveal that transforming the SDE to state independent diffusion provide more stable and accurate estimates. In fact, making the term that governs the diffusion independent from state reduces the variance and allows one source of randomness coming from the Brownian motion.

The main contribution of this thesis is investigating the performance of SDE in generating scenario forecasting, carrying a numerical-based comparison between the two inference approaches described above to explore the need of Lamperti transform, proposing a parametric form for the rate parameter and studying the sensitivity of each model with respect to all parameters.

As a future work, in order to confirm our findings, we will fit datasets provided by three different forecasting companies in Uruguay for the period of January 1, 2016 to March 13, 2017. We will also compare how different the fitting can be, based on the quality of the numerical forecast. To complete the study we will develop the numerical results of the partially observed model.

To avoid removing sets that have poor deterministic forecast or those with extreme observations (i.e. sudden dramatic change in wind power production that can not

be predicted by the deterministic forecast), it may be possible to use the following regime-switching SDE where δ is a threshold that should be selected properly

$$dX(t) = -\theta(X(t) - p(t))\mathbb{I}_{\{|X(t)-p(t)|<\delta\}} + f(X(t))\mathbb{I}_{\{|X(t)-p(t)|\geq\delta\}} \tag{6.6}$$

REFERENCES

[1] S. deMello, G.Cazes, and A. Arce, "Operational wind energy forecast with power assimilation," *14th International Conference on Wind Engineering*, 2014.

[2] a. Y. L. D Pan, H Liu, "A wind speed forecasting optimization model for wind farms based on time series analysis and kalman filter algorithm," *Power System Technology*, vol. 7, no. 32, 2008.

[3] E. E. H Liu and J. Shi, "Comprehensive evaluation of arma-garch(-m) approaches for modeling the mean and volatility of wind speed," *Applied Energy*, vol. 88, no. 3, 2011.

[4] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl, *The State-Of-The-Art in Short-Term Prediction of Wind Power: A Literature Overview, 2nd edition.* ANEMOS.plus, 2011, project funded by the European Commission under the 6th Framework Program, Priority 6.1: Sustainable Energy Systems.

[5] K. Moller, M. Zugno, and H. Madsen, "Probabilistic forecasts of wind power generation by stochastic differential equation models," *Journal of Forecasting*, vol. 35, no. 3, pp. 189–224, 2016.

[6] S. M. Iacus, *Simulation and Inference for Stochastic Differential Equations (with R examples).* Springer Series in Statistics, 2008.

[7] M. Ballesio, "Indirect inference for scalar time-homogeneous stochastic differential equations based on moment expansions," *Masters Thesis*, 2015.

[8] B. Øksendal, *Stochastic Diffrential Equations.* Springer, 2000.

[9] G. Durham and R. Gallant, "Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes," *Journal of Business & Economic Statistics*, vol. 20, no. 3, 2002.

[10] K. Moller and H. Madsen, "From state dependent diffusion to constant diffusion in stochastic differential equations by the lamperti transform," *Kgs. Lyngby, Denmark: Technical University of Denmark, DTU Informatics, Building 321. (IMM-Technical Report-2010-16)*, 2010.

[11] S. Asmussen and P. Glynn, *Stochastic Simulation: Algorithms and Analysis*, ser. Stochastic Modelling and Applied Probability. Springer New York, 2007. [Online]. Available: https://books.google.com.sa/books?id=vMI2MdxchU0C

[12] P. Dupuis, M. A. Katsoulakis, Y. Pantazis, and P. Plecháč, "Path-space information bounds for uncertainty quantification and sensitivity analysis of stochastic dynamics," *SIAM/ASA Journal on Uncertainty Quantification*, vol. 4, no. 1, pp. 80–111, 2016. [Online]. Available: http://dx.doi.org/10.1137/15M1025645

[13] W. Greene, *Maximum likelihood estimation*, 2010.

[14] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.

[15] A. Moraes, "Simulation and statistical inference of stochastic reaction networks with applications to epidemic models," 2015.

[16] A. Moraes, F. Ruggeri, R. Tempone, and P. Vilanova, "Multiscale modeling of wear degradation in cylinder liners," *Multiscale Model Simulation*, vol. 12, no. 1, pp. 396–409, 2014.

APPENDICES

A Appendix 1

A.1 Preprocessing the datasets

A.1.1 Interpolation of the numerical forecast

As we have hourly recorded observation of the numerical forecast, we need to interpolate between the point prediction in order to obtain a continuous curve representing $p_j(t), j = 1, \ldots, J$. For that, we use natural cubic splines interpolant, constructed of piecewise third-order polynomials which pass through a set of control points called knots. For our cases we chose the number of knots to be equal to the number of observations per set, for example in case of Fig (A.1) we consider 48-hour time horizon.

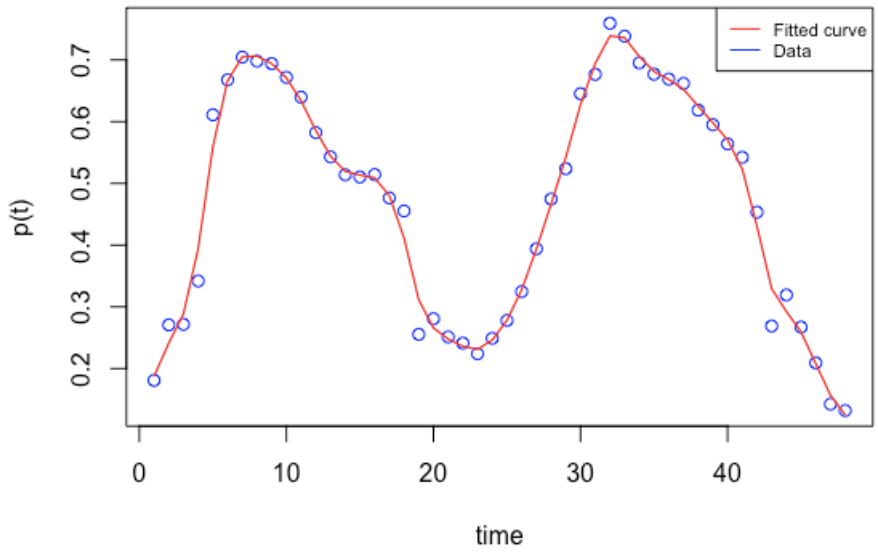

Figure A.1: Interpolation of the numerical forecast , set of March 1 , 2016

A.1.2 Examples of discarded sets

In Figures (A.2), we present three of the 8 discarded sets to show how bad the numerical forecast can be in comparison with the real production. We can see that at time (t=1) there is at least 20% difference in absolute values between the two time series. Also, during the time horizon the numerical forecast is not always able to capture the behavior of the numerical forecast.

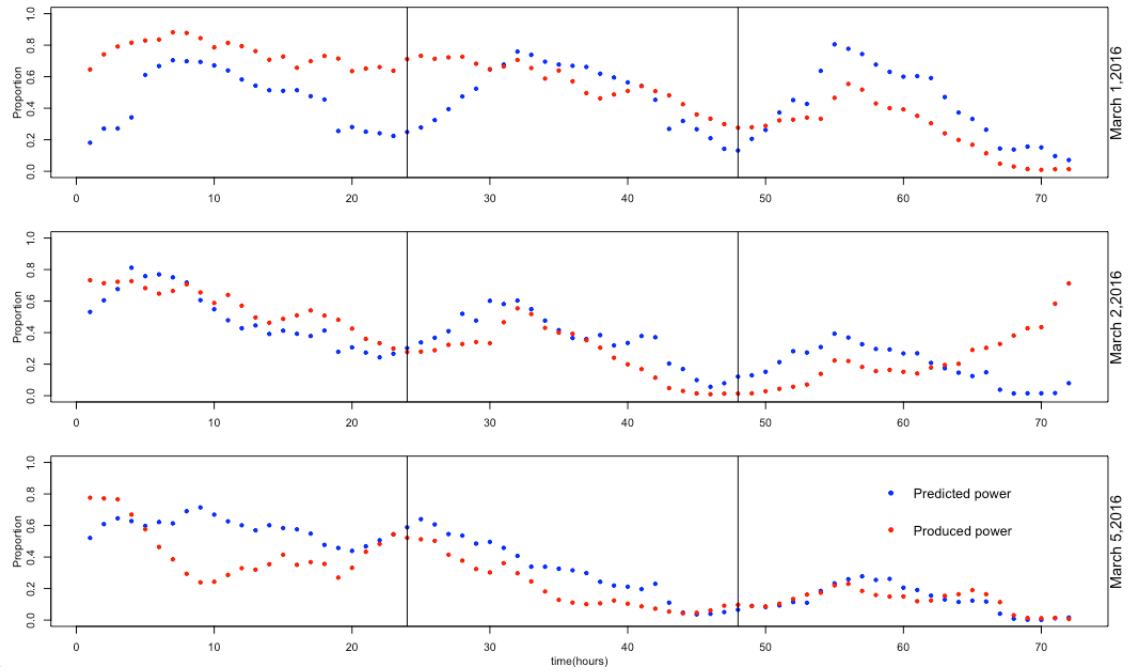

Figure A.2: Numerical forecast and real production

A Appendix 2

A.2 Background tools

A.2.1 Itô formula

Assume $X(t)$ is an Itô drift-diffusion process that satisfies the stochastic differential equation:

$$dX(t) = b(X(t))dt + \sigma(X(t))dW(t)$$

where $W(t)$ is a Wiener process.

If g(t,x) is a twice-differentiable scalar function with respect to x and once-differentiable with respect to t, its Taylor series expansion is given by

$$dg = \frac{\partial g}{\partial t}\,dt + \frac{\partial g}{\partial x}\,dx + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}\,dx^2 + \mathcal{O}(dx^3).$$

Substituting $X(t)$ for x and therefore $b(X(t))dt + \sigma(X(t))dw(t)$ for dx gives

$$dg = \frac{\partial g}{\partial t}\,dt + \frac{\partial g}{\partial x}(b(X(t))dt + \sigma(X(t))dW(t)) + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}\left(b^2(X(t))\,dt^2 + 2b(X(t))\sigma(X(t))\,dt\,dW(t)\right)$$
$$+\sigma^2(X(t))\,dW^2(t) + \mathcal{O}(dX(t)^3)$$

In the limit as dt tends to 0, the terms dt^2 and $dtdW(t)$ tend to zero faster than $dW^2(t)$. Thus, the equation of g satisfies

$$dg = \left(\frac{\partial g}{\partial t} + b(X(t))\frac{\partial g}{\partial x} + \frac{\sigma^2(X(t))}{2}\frac{\partial^2 g}{\partial x^2}\right)dt + \sigma(X(t))\frac{\partial g}{\partial x}\,dW(t)$$

A.2.2 Euler-Maruyama method

Euler method, is a numerical approximation of the solution of a stochastic differential equation (SDE). Consider the stochastic differential equation

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

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 Y recursively defined on $\Delta t > 0 : 0 = \tau_0 < \tau_1 < \cdots < \tau_N = T$ and $\Delta t = T/N$; by

$$Y_{n+1} = Y_n + b(Y_n)\,\Delta t + \sigma(Y_n)\,\Delta W_n,$$

for $1 \leq n \leq N$ where $\Delta W_n = W_{\tau_{n+1}} - W_{\tau_n}$.

A.2.3 Dynkin's formula and the infinitesimal generator operator

Dynkin's formula is defined as

$$\frac{d\mathbb{E}[f\big(X(t)\big)]}{dt} = \mathbb{E}[\mathcal{L}f\big(X(t)\big)] \tag{A.1}$$

The operator $\mathcal{L}$ is the infinitesimal generator of the process, X, defined as

$$\mathcal{L}f(x) = \lim_{h \downarrow 0} \frac{\mathbb{E}^x[f(X(t+h))|X(t) = x] - f(x)}{h} \tag{A.2}$$

it is defined for the set of functions for which the limit of the right hand side of (A.2)

exists. It is proven, [6], that the infinitesimal generator operator can be written as

$$\mathcal{L}f(X) = \sum_k b_k(X)\frac{\partial f}{\partial X_k}(X) + \frac{1}{2}\sum_{k,l}\left[\sigma\sigma^\top\right]_{k,l}(X)\frac{\partial^2 f}{\partial X_k\,\partial X_l}(X)$$

where $\sigma^\top$ denotes the transpose of σ, the diffusion term.

A.3 Lamperti Transformation for the proposed model

The diffusion term of the process $Y(t)$ is given by:

$$\sigma(Y(t)) = \alpha(t)\beta(Y(t)) = \sqrt{2\alpha\theta p(t)\big(1-p(t)\big)}\sqrt{Y(t)\big(1-Y(t)\big)}$$

In this special case it will be easier to use the Lamperti transform on $\beta(X(t))$ and leave the diffusion time dependent. Define:

$$\begin{aligned}
Z(t) = \Psi(Y(t)) &= \int \frac{1}{\beta(x)}dx\Big|_{x=Y(t)} \\
&= \int \frac{1}{\sqrt{x(1-x)}}dx\Big|_{x=Y(t)} \\
&\text{change of variable } x = \frac{1}{2}(\sin(u)+1), dx = \frac{\cos(u)}{2}du \\
&= \int \frac{\cos(u)}{\sqrt{1-\sin(u)^2}}du \\
&= \int du = u = \arcsin(2x-1)\Big|_{x=Y(t)} \\
&= \arcsin(2Y(t)-1)
\end{aligned}$$

Then

$$Z(t) = \Psi(Y(t)) = \arcsin(2Y(t)-1)\,.$$

A.3.1 Equation of the transformed process

According to Lamperti transformation of $X(t)$, the process $Z(t)$ follows the following SDE:

$$\begin{aligned}
dZ(t) &= \tilde{b}(Z(t))dt + \alpha(t)dW_t \\
&= (\frac{b(Y(t))}{\beta(Y(t))} - \frac{1}{2}\beta'(Y(t))\alpha^2(t))dt + \alpha(t)dw_t \\
&= (\frac{b(\Psi^{-1}(Z(t)))}{\beta(\Psi^{-1}(Z(t)))} - \frac{1}{2}\beta'(\Psi^{-1}(Z(t)))\alpha^2(t))dt + \alpha(t)dw_t \\
&= (\frac{-2\theta[\frac{1+\sin(Z(t))}{2} - p(t)]}{\sqrt{1-sin(Z(t))^2}} - \frac{1}{2}\frac{(1-2\frac{(1+\sin(Z(t)))}{2})}{\sqrt{1-sin(Z(t))^2}}\alpha^2(t))dt + \alpha(t)dw_t \\
&= (\frac{-2\theta[\frac{1+\sin(Z(t))}{2} - p(t)]}{\cos(Z(t))} + \frac{\sin(Z(t))}{\cos(Z(t))}\alpha^2(t))dt + \alpha(t)dw_t \\
&= \frac{-\theta\Big(1+\sin(Z(t)) - 2p(t)) - \alpha\sin(Z(t))p(t)(1-p(t))\Big)}{\cos(Z(t))}dt + \alpha(t)dw_t
\end{aligned}$$

A.3.2 Bias introduced in Model 2

Define the process $\bar{Z}$ as the process with minimal bias in the mean, $\bar{Z}$ satisfies the following SDE

$$\frac{d\bar{Z}(t)}{dt} = F((\bar{Z}(t))) + \sqrt{2\theta\alpha p(t)\Big(1-p(t)\Big)}dW_t \tag{A.3}$$

where F is giving by :

$$F(\bar{Z}(t)) = \frac{-\theta(1+\sin(\bar{Z}(t)) - 2p(t) + (2c-\alpha)\sin(\bar{Z}(t))p(t)(1-p(t)))}{\cos(\bar{Z}(t))}dt$$

The equation of the mean and variance of $\bar{Z}$ are given by

$$\begin{cases}
\frac{dm(t)}{dt} = \Big(m^{\theta}(t) - p(t) - (\frac{\alpha}{2} - c)p(t)(1-p(t))(1-2m(t))\Big) \\
\frac{dv(t)}{dt} = e^{\int_0^t g(s)ds}\Big(\int_0^t 2\alpha p(s)(1-p(s))e^{-\int_0^t g(u)du}ds\Big)
\end{cases}$$

where $g(t) = \frac{-2\theta(1-(\alpha-2c)p(t)(1-p(t))+\sin(\bar{Z}(t))(1-2p(t))}{\cos^2(\bar{Z}(t))}$

A.3.3 Numerical results of Model 2 benchmark example 2

Model 2 is able to retrieve the true values, figures (??-??) show that the estimated values are $(\hat{\theta}, \hat{\alpha}, \hat{\phi}) = (0.188, 0.224, 0.048)$ and they indicate that Model 2 retrieves the true values with fewer paths compared to Model 1.

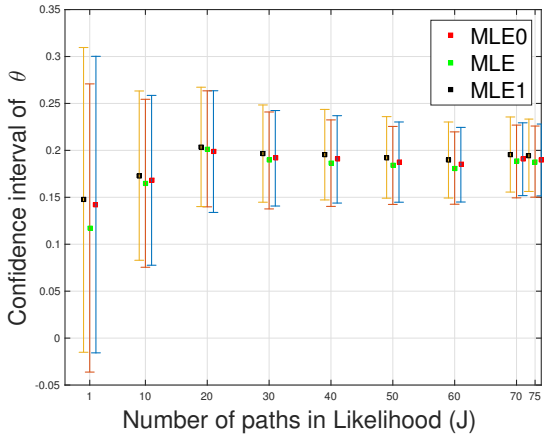

Figure A.3: Convergence of θ.

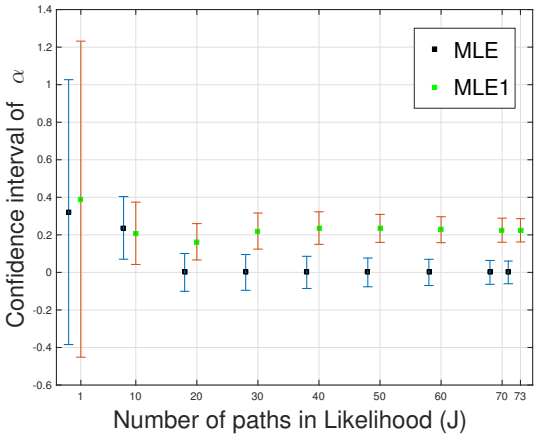

Figure A.4: Convergence of α.

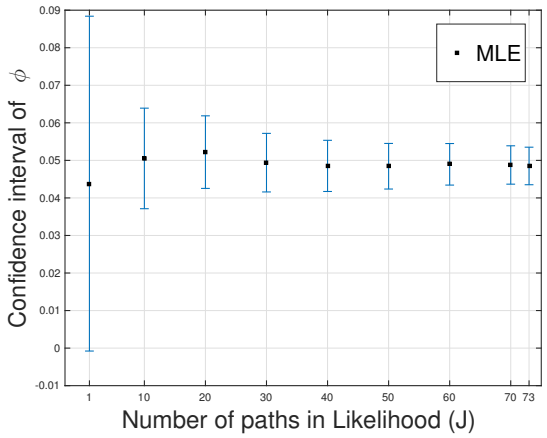

Figure A.5: Convergence of ϕ.

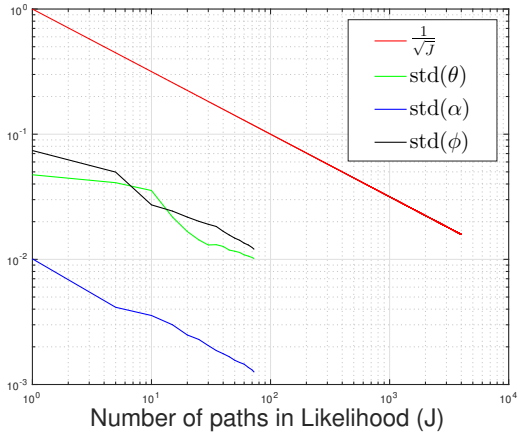

Figure A.6: Rate of convergence of the likelihood