

## INFERENCE AND SENSITIVITY IN STOCHASTIC WIND POWER FORECAST MODELS.

Soumaya Elkantassi<sup>1</sup>, Evangelia Kalligiannaki<sup>1</sup> and Raul Tempone<sup>1</sup>

<sup>1</sup>Applied Mathematics and Computational Science, CEMSE,  
King Abdullah University of Science & Technology  
Thuwal 23955, Saudi Arabia

e-mail: {soumaya.kantassi, evangelia.kalligiannaki, raul.tempone}@kaust.edu.sa

**Keywords:** Indirect inference, wind power, probabilistic forecasting, model selection, sensitivity.

**Abstract.** *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.*

## 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<sup>1</sup>. To incorporate the increased amount of generated power in to the supply system, reliable forecasting models are crucial.

Many statistical forecast methods have been developed, for example persistence models [1], Kalman filters [4], and Autoregressive and Moving Average Model (ARMA) [?]. 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). 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 it possible for decision makers to control electricity costs in the market.

Recently, a probabilistic forecast model based on stochastic differential equations (SDEs), [7] 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 [7] 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 Section 2 we formulate the SDEs and propose two models. In Section 3 we describe the available datasets and the preprocessing procedure. Then, numerical results for each model are presented in Section 4. We conclude in Section 5 by discussing our results.

## 2 Indirect Inference method with Stochastic Differential Equations

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 the following parametrized Ito diffusion SDE

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

and assume that the drift,  $b(X(t), t; \boldsymbol{\theta})$ , and the diffusion,  $\sigma(X(t), t; \boldsymbol{\theta})$ , coefficients are such that a solution exists and is unique, [2]. Here,  $\boldsymbol{\theta}$  denotes the set of parameters for each model.  $W(t)$  is a one-dimensional Brownian motion. As we expect the process,  $X(t)$ , to align with the numerical forecast,  $p(t)$ , we choose the drift to be a linear polynomial of the state

$$b(x, t; \boldsymbol{\theta}) = -\theta(t)(x - p(t)), \quad (1)$$

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

<sup>1</sup>The energy mix statistics are sourced from the Global Wind Energy Council and the World Resources Institute.

We distinguish two major models (Model 1 and Model 2) based on the choice of the diffusion coefficient. The purpose of this distinction is to explore the need for the transformation to have a state independent diffusion coefficient.

**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

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

The first proposed model–Model 1–is

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

and

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

The inference approach for the parameters is based on the two moment equations for the stochastic 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 co-variance functions respectively. The equation of the mean is

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

with the solution

$$\mu_X(t) = e^{-\int_0^t \theta(s)ds} \left( \int_0^t \theta(s)p(s)e^{\int_0^s \theta(u)du} ds + \mu_0 \right). \quad (3)$$

The equation of the variance is

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

from which we have that

$$v_X^\theta(t) = e^{-2(1+\alpha)\int_0^t \theta(s)ds} \left( \int_0^t 2\theta(s)\alpha\mu_X(s)(1 - \mu_X(s))e^{2(1+\alpha)\int_0^s \theta(u)du} ds \right). \quad (4)$$

The co-variance between two times of  $X(t)$  is given by

$$v_X^\theta(t, s) = v_X^\theta(s)e^{-\int_0^t \theta(u)du}, \quad \forall t > s. \quad (5)$$

## Model 2

In this section, we transform equation (2) 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 both simulation and estimation, [9]. Numerical schemes sampled from an SDE with additive noise are more stable, and usually of higher order than those sampled from an SDE with multiplicative noise. Moreover, [10] and [7] suggested that the transformed process is better approximated by a Gaussian distribution.

Let us denote by  $Z(t) = \Psi(X(t)) = \int \frac{1}{\sqrt{x(1-x)}} dx \Big|_{x=X_t}$  the transformed process that satisfies an SDE with state-independent diffusion. For the definition of the Lamperti transform see [9]. 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  $\alpha$

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

Here 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  $\alpha$ . This suggests the form of the Model 2

$$\begin{cases} dY(t) = -\theta(Y(t) - p(t))dt + \sqrt{2\theta\alpha p(t)(1 - p(t))Y(t)(1 - Y(t))}dW_t, & t > 0, \\ Y(0) = x_0, \end{cases} \quad (6)$$

and

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

The Lamperti transformation of the process  $Y(t)$  defines the process

$$Z(t) = \Psi(Y(t)) = \int \frac{1}{\sqrt{x(1-x)}} dx \Big|_{x=Y(t)} = \arcsin(2Y(t) - 1),$$

which satisfies the following SDE

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

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

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

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

$$\mu_Z(t) = \arcsin \left\{ e^{-\int_0^t h(u)du} \left( \int_0^t \theta(s)(2p(s) - 1)e^{\int_0^s h(u)du} ds + \sin(Z(0)) \right) \right\}, \quad (7)$$

where  $h(t) = \theta(t)[1 - 2\alpha p(t)(1 - p(t))]$ . Using the inverse of  $\Psi$  given by  $Y(t) = \Psi^{-1}(Z(t)) = \frac{1}{2}(1 + \sin(Z(t)))$ , we obtain the following approximation for the mean  $\mu_Y(t) = \mathbb{E}[Y(t)]$  after using the Lamperti transform

$$\tilde{\mu}_Y(t) = \Psi^{-1}(\mu_Z(t)) = \frac{1}{2}(1 + \sin(\mu_Z(t))). \quad (8)$$

The equation of the variance is given by

$$\frac{dv_Z(t)}{dt} = 2A(t; \boldsymbol{\theta})v_Z(t) + \tilde{\sigma}^2(t; \boldsymbol{\theta}), \quad v_Z(0) = 0,$$

and an exact solution for this first-order ordinary differential equation (ODE) is given by

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

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

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

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

$$\tilde{v}_Y(t) = \frac{1}{4} \cos^2(\mu_Z(t)) v_Z(t). \quad (9)$$

and

$$\tilde{v}_Y(t, s) = \frac{1}{4} \cos(\mu_Z(t)) v_Z(t, s) \cos(\mu_Z(s)). \quad (10)$$

## 2.1 Approximate Likelihood

We assume that the observation variables  $\mathbf{D}^{(j)} = (D^{(j)}(t_1), \dots, D^{(j)}(t_N))$ ,  $j = 1, \dots, J$ , satisfy

$$\mathbf{D}^{(j)} = \mathbf{X}^{(j)} + \boldsymbol{\epsilon}^{(j)} \quad (11)$$

where  $\mathbf{X}^{(j)} = (X^{(j)}(t_1), \dots, X^{(j)}(t_N))$  is the solution of the SDE in equations (2) or (6) for  $p(t) = p^{(j)}(t)$ , and  $\boldsymbol{\epsilon}^{(j)}$  are the measurement errors that are generated from a Gaussian distribution  $\mathcal{N}(0, \Sigma^{\epsilon, (j)})$ .

The approximate likelihood based on the two-moment expansion of the process  $X(t)$  and  $Y(t)$  is

$$L(\boldsymbol{\theta}; D_{nm}) = \prod_{j=1}^J (2\pi)^{-N/2} |\mathbf{V}^{(j)}(\boldsymbol{\theta})|^{-1/2} \exp [(\mathbf{d}^{(j)} - \boldsymbol{\mu}^{(j)}(\boldsymbol{\theta}))^T [\mathbf{V}^{(j)}(\boldsymbol{\theta})]^{-1} (\mathbf{d}^{(j)} - \boldsymbol{\mu}^{(j)}(\boldsymbol{\theta}))] \quad (12)$$

for a given set of observations  $d_{NJ} = \{\mathbf{d}^{(j)} = (d^{(j)}(t_1), \dots, d^{(j)}(t_N)), j = 1, \dots, J\}$ . Here  $\boldsymbol{\mu}^{(j)}(\boldsymbol{\theta}) = (\mu(t_1), \dots, \mu(t_N))$ ,  $\mathbf{V}^{(j)}(\boldsymbol{\theta}) = \mathbf{V}_s^{(j)}(\boldsymbol{\theta}) + \Sigma^{\epsilon, (j)}$ , and  $\mathbf{V}_s^{(j)}(\boldsymbol{\theta})$  is an  $n \times n$  matrix with elements  $[\mathbf{V}_s^{(j)}]_{kl} = v(t_k, t_l)$ ,  $k, l = 1, \dots, N$ . Note that  $\mu(t)$ ,  $v(t)$  and  $v(t, s)$  are given in equations (3) -(5) of Model 1 and in equations (8)- (10) of Model 2. Finally, the maximum likelihood estimates (MLE) of the parameters,  $\boldsymbol{\theta}$ , are obtained by maximizing the function in equation (12).

### 3 Description of the data

Application of our models to the wind power problem is built on two datasets. The first dataset consists of 72 hours of numerical predictions of wind power production and the second dataset includes the actual wind power produced during the same 72 hours (one observation every one hour) in Uruguay. The data correspond to the aggregate power ( MWatts ) in the following wind farms: ARTI, CAPE, FACE, CPPP, FLO1, JPTE, KENT, LRLM, MIN1, MWIN, RSUR and TDMA. The nominal capacity of this set of parks is 793.7 MWatts. Before proceeding, we normalize each of the datasets with the maximum installed capacity to obtain values between  $[0, 1]$ . We denote by  $p_j(t_n)$ ,  $d^{(j)}(t_n)$ ,  $j = 1, \dots, J$ ,  $J = 81$ ,  $n = 1, \dots, 72$  the normalized numerical prediction and historical prediction at the  $j$ th set at time  $t_n$ .

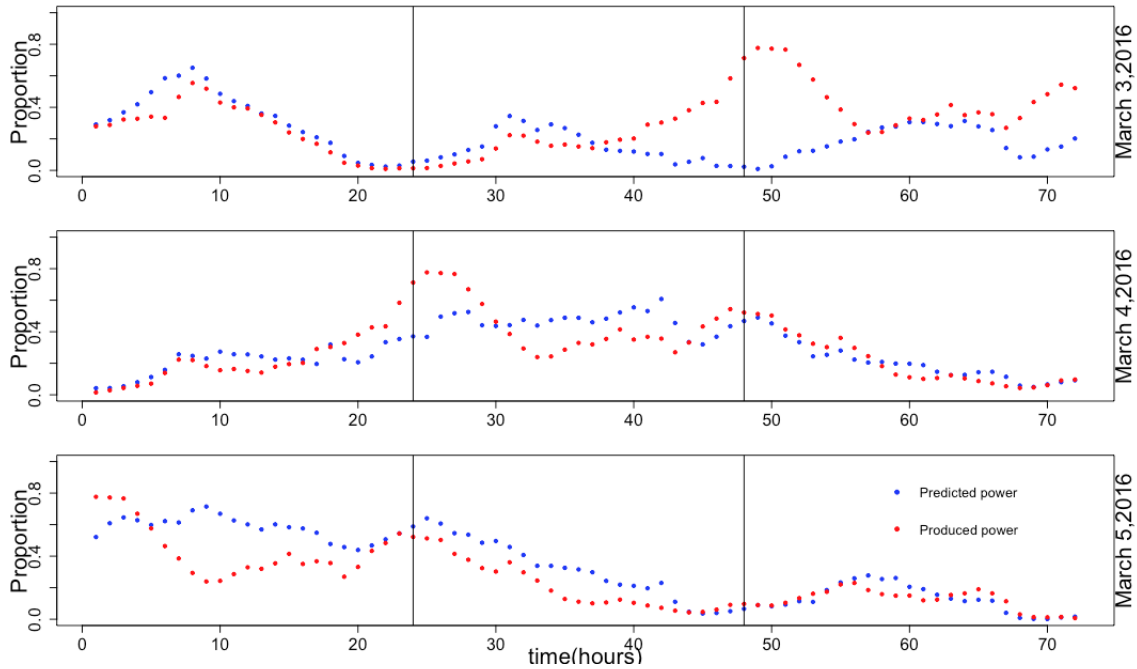


Figure 1: Numerical forecast and real production comparison

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 between  $x(t)$  and  $p(t)$  (i.e., more than 20%) is observed especially at time ( $t = 1$ ). We consider them as outliers with respect to our models. Including them affects the quality of the parameter estimates and produces instabilities in the likelihood maximization problem. If we choose a time horizon that is greater than 24 hours, we observe redundancy between successive sets of actual power production, which contradicts the independence assumption while computing the likelihood function. For this reason, some sets are discarded ( $J = 73$ ).

### 4 Results

In this section, we present the application of the models described in Section 2 to the datasets described in Section 3. For both models, we consider a parametrization with respect to time-independent parameters. Next, we introduce potential trends of  $\theta$  and select the best model according to the Akaike Information Criterion (AIC) and the Bayesian Information Criterion

(BIC). Then, to check variability in the estimates from the selected model, we use the bootstrapping method to compute optimal parameters with respect to randomly selected subsets of 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. Finally, we validate the best model by verifying that the test set observations fall within the generated confidence bands. We also present the sensitivity of the model parameters based the Fisher Information Matrix (FIM) for the approximate likelihood.

#### 4.1 Proposed models and model selection

The different parametrizations of Model 1 (2) and Model 2 (6) are summarized in Table 4.1.

Parametric Model	$\theta(t)$	Number of parameters $(\theta(t), \alpha, \phi)$
SDE0	$\theta_0$	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

Table 1: Proposed models with parametrizations of the rate  $\theta(t)$

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

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)	(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	$10^{-2}(0.39, 0.39)$	$10^{-2}(0.73, 0.76, 0.17)$	$10^{-2}(0.44, 0.47, 0.32)$	$10^{-2}(0.41, 0.40, 0.16)$

Table 2: Model 1. Optimal parameter values for actual data and different time horizons.

In Table 2, we present the optimal values of the parametric model SDE0  $(\theta, \alpha, \phi)$  for the time horizons

$\{6, 12, 18, 24, 36, 48\}$  based on Model 1. In the column termed 'Without noise', we report the output of the minimization of the negative logarithm of the likelihood function (12) using the actual data. For the numerical optimization, we used the fmincon function in MATLAB. We observe that Model 1 is not able to recover the value of  $\phi$ , and the output is always zero ( thus we simply report values of  $\theta$  and  $\alpha$ ). To investigate the effect of this observation, we added an artificial Gaussian noise to the data with  $\phi = \{0.05, 0.02, 0.01\}$ . We can see that the values of  $\theta$  and  $\alpha$  change noticeably and appear to be sensitive to the measurement error. For  $\phi = 0.05$ , the likelihood is 0.036 while  $\alpha$  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  $\alpha$ . Part of the noise,  $\phi$ , is indeed absorbed by  $\alpha$ , which explains the jump in its value. This observation is confirmed in the last two columns of Table 2, where the likelihood cannot retrieve  $\phi$  for noise on the order of 0.02 and 0.01.

We observe that  $\theta$  is decreasing with time, which indicates that the predictability of the model becomes less reliable with increasing time horizons. This motivates the use of the exponentially decreasing parametric functions for  $\theta$  in Table 4.1, that define the models SDE0, SDE1, SDE2, and SDE3. The models are parametrized by  $(\theta_i(t), \alpha, \phi)$  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 lower bound given by  $\theta_2$  to avoid converging to zero and  $\theta_3(t)$  is a linear combination of exponentially decreasing and increasing functions to better fit the data (see table 4.1).

To obtain good initial values for the optimization for the time varying  $\theta(t)$ , in SDE1, SDE2, and SDE3 models, we fit the optimal values for the constant  $\theta^*$  of the SDE0 model obtained in the first column of Table 2 using the ‘Curve Fitting Toolbox’ of MATLAB. 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 3 for the  $T = 24$ -hour time horizon.

Model	Initial values	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, $2 \cdot 10^{-3}$ , 0.02, 0.1)	(0.28, 0.103, $10^{-3}$ , 0.363, 0.018)	$10^{-2}(2.06, 1.04, 4 \cdot 10^{-4}, 0.16, 0.05)$

Table 3: Model 1. Optimal parameters for a 24-hour time horizon for the parametric SDE’s i Table 4.1

Among the four candidate models, data in Table 4.1 confirm that SDE1 is the best model as it has the smallest AIC and BIC values.

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

Table 4: Model selection

## Model 2

Next, we present the best model among those described in Table 4.1 based on Model 2, described in equation (6), following the same steps as for Model 1.



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, 0.03)
	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	$10^{-2}(0.31, 3.98, 0.09)$

Table 5: Optimal parameter values for actual real data and different time horizons for Model 2 using the Lamperti-transformed process.

From Table 5, we see that  $\theta$  exhibits the same behavior as in Model 1, it decreases with longer time horizons. 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.

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, 0.018, 0.453, 0.033)	$10^{-2}(0.81, 0.32, 0.02, 1.39, 1.04, 0.05)$

Table 6: Optimal parameters for 24-hour time horizon using the Lamperti transformed process.

Table 7 confirms that SDE1 is the best model with respect to AIC and BIC.

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

Table 7: Model selection

## 4.2 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 parameters from the bootstrapped samples. We compute the sample mean and the bootstrapped confidence intervals of each parameter. Figures 2 and 3 show that our data predict a reliable set of estimates that both models agree on.

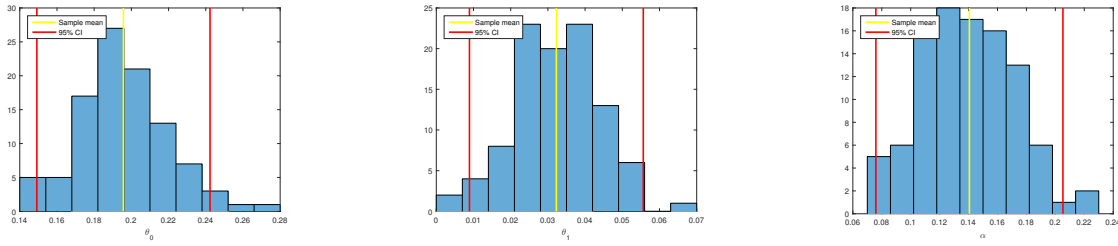


Figure 2: Model 1. Bootstrap histograms of  $\theta_0$ ,  $\theta_1$  and  $\alpha$  based on  $n = 100$  iterations

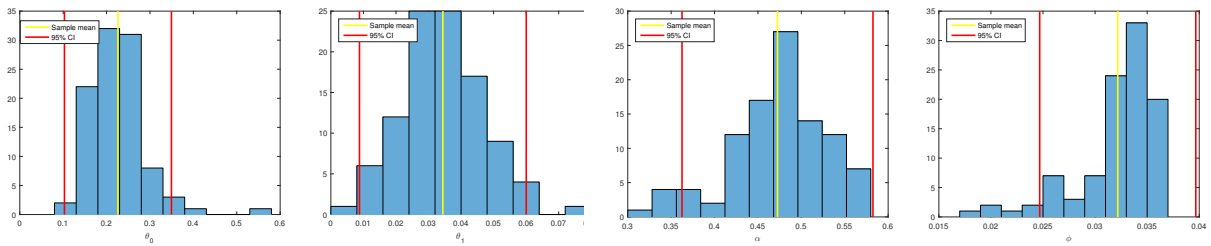


Figure 3: Model 2. Bootstrap histograms of  $\theta_0$ ,  $\theta_1$ ,  $\alpha$  and  $\phi$  based on  $n = 100$  iterations

Figure 4 shows that the prediction interval of  $\theta(t)$  of Model 1 falls inside the prediction interval of Model 2, suggesting that both predict the same rate.

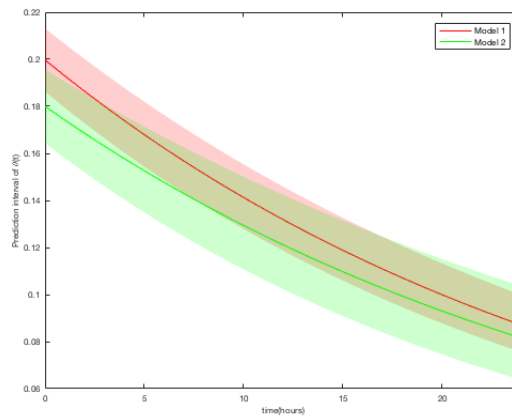


Figure 4: Best fit for SDE1 for the two models.

To validate the selected models we present the empirical path density and confidence bands for two days in the training set, as shown in figures 5 - 8. We generate the empirical path density from simulated paths of SDE's (2) and (6) for the MLE optimal parameters for which we use a Euler-Maruyama discretization scheme. On the same graphs, we present the actual wind energy production and the available numerical forecast. The figures verify that the actual production falls into the 95% confidence band for most times.

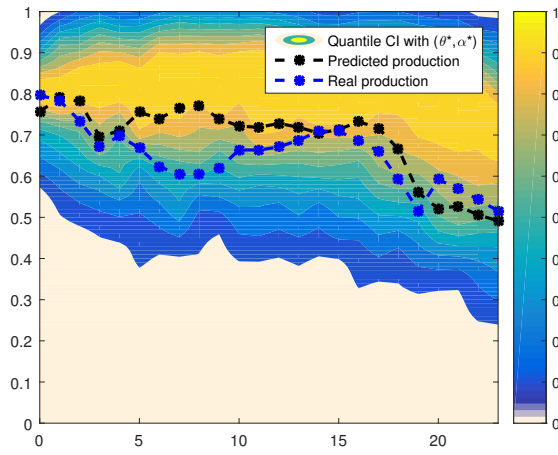


Figure 5: Empirical Confidence Interval (CI) and real wind energy production corresponding to May 16, 2016 (12 hours)

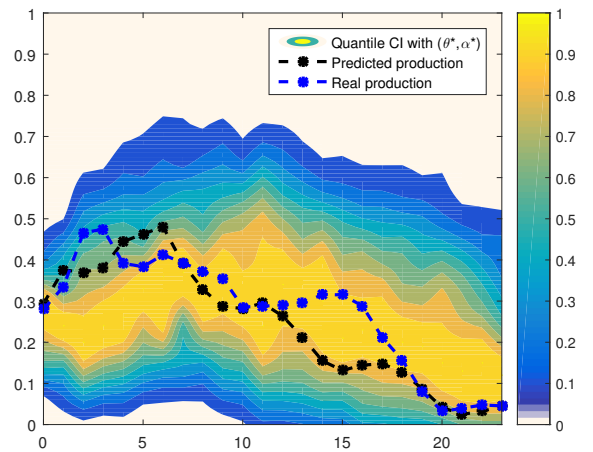


Figure 6: Empirical CI and real wind energy production corresponding to May 22, 2016 (24 hours)

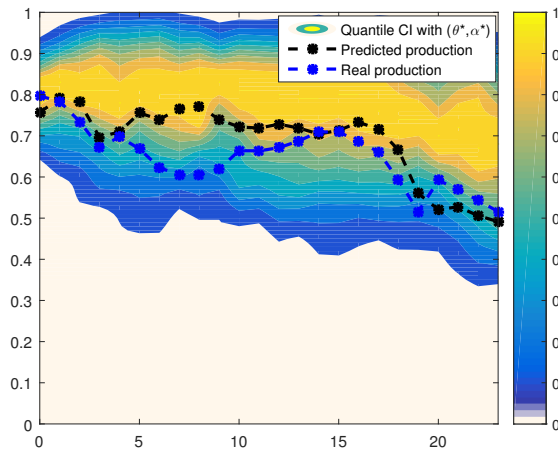


Figure 7: Empirical CI and real wind energy production corresponding to May 16, 2016 (12 hours)

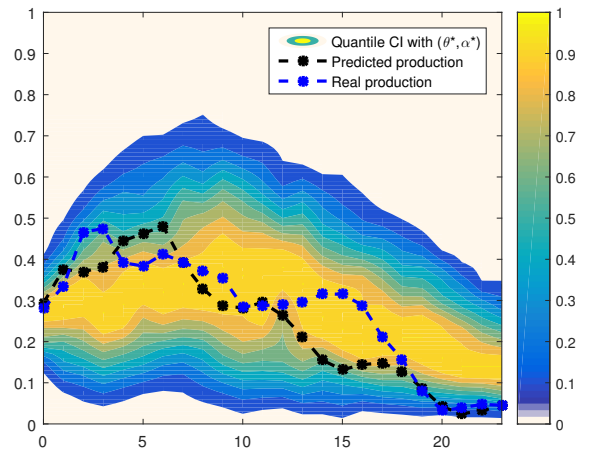


Figure 8: Empirical CI and real wind energy production corresponding to May 22, 2016 (Horizon 24 hours)

Next, we analyze the sensitivity of both models to the parameters  $(\theta_0, \theta_1, \alpha)$ . The sensitivity index is given by the diagonal elements of the FIM that corresponds to the approximate likelihood for each model, [14]. A direct sensitivity analysis of the models based on (2), and (6) is the subject ongoing work, [16]. Figures 9 and 10 demonstrate that the most sensitive parameter for both models is  $\theta_1$ , 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 different; for Model 1, it is  $\alpha$  while for Model 2, it is  $\theta_0$ .

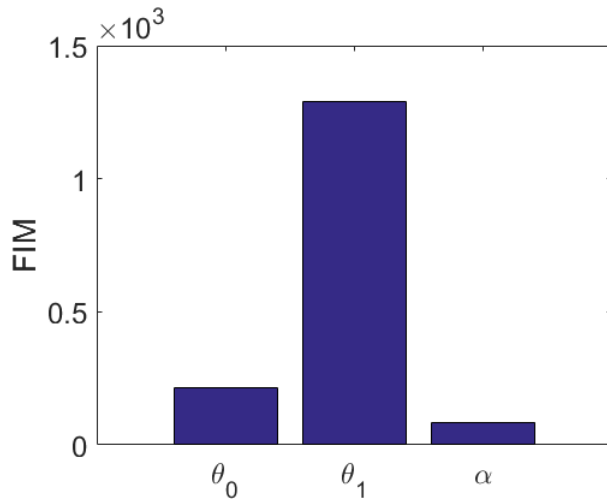


Figure 9: Sensitivity of parameters in Model 1. The most sensitive parameter is  $\theta_1$  and the less sensitive is  $\alpha$ .

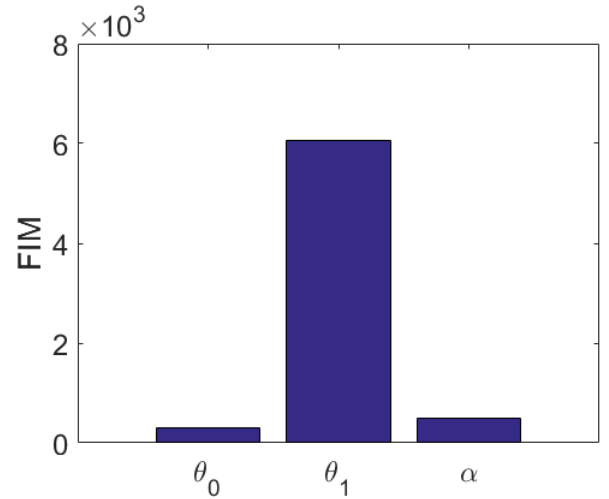


Figure 10: Sensitivity of parameters in Model 2. The most sensitive parameter is  $\theta_1$  and the less sensitive is  $\theta_0$ , in contrast to Model 1.

## 5 Discussion and Conclusions

It is important to point out that we based our study on a benchmark example in which we generate synthetic data from real deterministic forecast paths,  $p_i(t)$ ,  $i = 1, \dots, J$  using specific values of  $(\theta(t), \alpha, \phi)$ . Then, for each  $p_i(t)$ , we sample one path,  $X_i(t)$ , to mimic the actual data and check whether the MLE parameters are close to the true parameters. The numerical results of the benchmark example reveal that although both models are able to retrieve the true parameter values, Model 2 is numerically more stable (parameters have smaller confidence intervals, and the rate of convergence is more stable).

With actual data, Model 1 is not able to capture the value of  $\phi$  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 the 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, the two models predict the same rate, (see Figure 4) and the most sensitive parameter for both models is  $\theta_1$ , which controls the decay. The least sensitive parameter is different.

To avoid removing datasets that have poor deterministic forecasts or that have extreme observations (i.e., sudden dramatic changes in wind power production that can not be predicted by the deterministic forecast), it may be possible to use regime-switching models.

## REFERENCES

- [1] S. de Mello, G. Cazes, and A. Gutierrez Arce, *Operational wind energy forecast with power assimilation*, 14th International Conference on Wind Engineering, 2014.
- [2] B. Øksendal, *Stochastic Differential Equations*, Springer, 2000.

- 
- [3] M Westphal, and J. Thwaites, *Transformational Climate Finance: An Exploration of Low-Carbon Energy*, World Resources Institute, 2016.
- [4] D. Pan, H. Liu, and Y. Li, *A wind speed forecasting optimization model for wind farms based on time series analysis and Kalman filter algorithm* Power System Technology, 32 (7), 82-86, 2008.
- [5] H. Liu, E. Erdem, 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, pp. 724732, 2011.
- [6] R.G. Kavasseri, and K. Seetharaman, *Day-ahead wind speed forecasting using f-ARIMA models*, Renewable Energy, vol. 34, no. 5, pp. 13881393, 2009.
- [7] M. Kloppenborg, M. Zugno, and H. Madsen, *Probabilistic forecasts of wind power generation by stochastic differential equation models*, vol. 35, no. 3, pp. 189–224, 2016.
- [8] M. Ballesio, *Indirect inference for scalar time-homogeneous stochastic differential equations based on moment expansions*, Masters thesis, Politecnico di Torino, 2015.
- [9] S. M. Iacus, *Simulation and Inference for Stochastic Differential Equations (with R examples)*. Springer Series in Statistics, 2008.
- [10] 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, pp. 297–316, 2002.
- [11] M. Kloppenborg and H. Madsen, *From state dependent diffusion to constant diffusion in stochastic differential equations by the Lamperti transform*, Technical University of Denmark, 2010.
- [12] 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.
- [13] A. Moraes, *Simulation and statistical inference of stochastic reaction networks with applications to epidemic models*, Ph.D. Thesis, KAUST, 2015.
- [14] T. M. Cover, and J. A. Thomas, *Elements of Information Theory*. Wiley Series in Telecommunications and Signal Processing, 2006.
- [15] S. Asmussen, and P. W. Glynn, *Stochastic Simulation: Algorithms and Analysis*, Springer-Verlag New York, 2007.
- [16] Y. Pantazis, and M. Katsoulakis, *A Relative Entropy Rate Method for path space sensitivity analysis of stationary complex stochastic dynamics*. J. Chem. Phys., 138, 054115, 2013