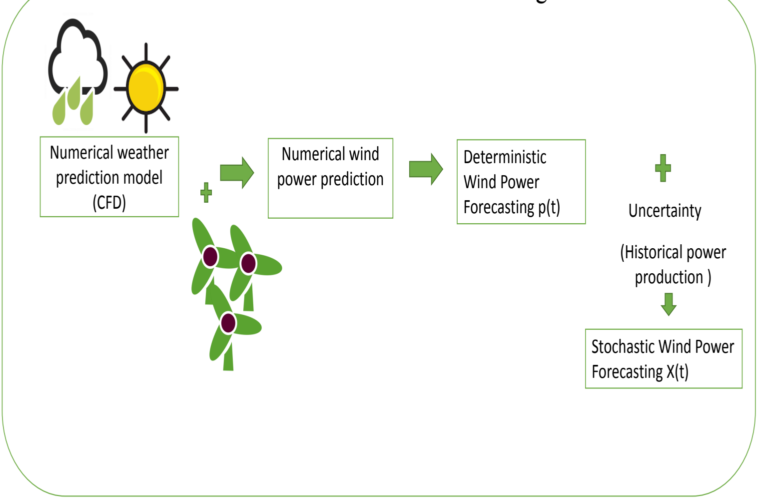
Probabilistic Framework of Wind Power Generation by Itô Stochastic Differential Equations

Soumaya Elkantassi

Computer, Electrical and Mathematical Sciences & Engineering division (CEMSE)
King Abdullah University of Science and Technology

August 21, 2017

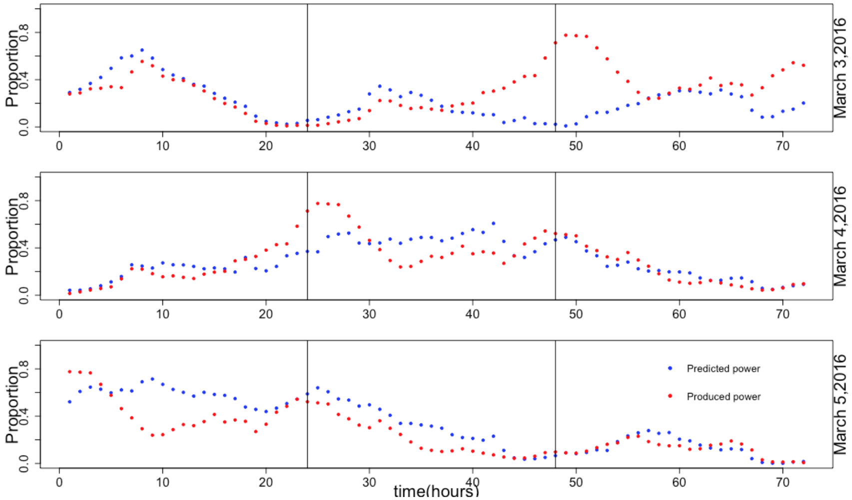
Stochastic Wind Power Forecasting



- Provide stochastic forecast analogue to p(t) $\Rightarrow$ scenarios of wind power forecasting with confidence bands.
- Stochastic optimization problems: management of electricity costs.

Outline

Data



Normalized historical power d_{ji} and numerical forecast p_{ji} in 12 wind farms in Uruguay from March 01 to May 31, 2016.

Itô SDE Modeling

Model 1

Let, $X(t)$, denote the stochastic forecast analogue to $p(t)$.
Assume it is the solution of the parametrized SDE

$$\text{Model 1}\ \begin{cases} dX(t) = b(X(t);\boldsymbol{\theta})dt + \sigma_1(X(t);\boldsymbol{\theta})dW(t), & t > 0\,, \\ X(0) = x_0 \end{cases}$$

$X(t) \in [0,1]$ and $\boldsymbol{\theta}$ denotes a vector of parameters.

Itô SDE Modeling

Model 1

Let, $X(t)$, denote the stochastic forecast analogue to $p(t)$.
Assume it is the solution of the parametrized SDE

$$\text{Model 1} \begin{cases} dX(t) = b(X(t); \boldsymbol{\theta})dt + \sigma_1(X(t); \boldsymbol{\theta})dW(t), & t > 0\,, \\ X(0) = x_0 \end{cases}$$

$X(t) \in [0, 1]$ and $\boldsymbol{\theta}$ denotes a vector of parameters.

- **Form of the drift:** $X(t)$ is 'close' to $p(t)$. In expectation, $X(t)$ tends to $p(t)$ in stationary regime

$$b(X(t), t; \boldsymbol{\theta}) = -\theta(t)\Big(X(t) - p(t)\Big)$$

Itô SDE Modeling

Model 1

Let, $X(t)$, denote the stochastic forecast analogue to $p(t)$.
Assume it is the solution of the parametrized SDE

$$\text{Model 1} \begin{cases} dX(t) = b(X(t);\boldsymbol{\theta})dt + \sigma_1(X(t);\boldsymbol{\theta})dW(t), & t > 0\,, \\ X(0) = x_0 \end{cases}$$

$X(t) \in [0,1]$ and $\boldsymbol{\theta}$ denotes a vector of parameters.

- **Form of the drift:** $X(t)$ is 'close' to $p(t)$. In expectation, $X(t)$ tends to $p(t)$ in stationary regime

$$b(X(t),t;\boldsymbol{\theta}) = -\theta(t)\Big(X(t) - p(t)\Big)$$

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

Itô SDE Modeling

Model 1

Let, $X(t)$, denote the stochastic forecast analogue to $p(t)$.
Assume it is the solution of the parametrized SDE

$$\text{Model 1} \quad \begin{cases} dX(t) = b(X(t);\boldsymbol{\theta})dt + \sigma_1(X(t);\boldsymbol{\theta})dW(t), & t > 0\,, \\ X(0) = x_0 \end{cases}$$

$X(t) \in [0,1]$ and $\boldsymbol{\theta}$ denotes a vector of parameters.

- **Form of the drift:** $X(t)$ is 'close' to $p(t)$. In expectation, $X(t)$ tends to $p(t)$ in stationary regime

$$b(X(t),t;\boldsymbol{\theta}) = -\theta(t)\Big(X(t) - p(t)\Big)$$

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

- **Form of the diffusion:** the diffusion vanishes at the boundaries

$$\sigma_1(X(t);\boldsymbol{\theta}) = \sqrt{2\alpha\theta(t)X(t)\Big(1 - X(t)\Big)}$$

Approximate likelihood

Suppose we have the following datasets

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

$x^{(j)}(t_i)$ is a realization of $X^{(j)}(t)$, solution of Model 1 with $p(t) = p^{(j)}(t)$

$p^{(j)}(t)$ is natural cubic-splines interpolation of $\mathbf{p}^{(j)} = \left\{ p_{ji} = p^{(j)}(t_i) \right\}$

Approximate likelihood

Suppose we have the following datasets

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

$x^{(j)}(t_i)$ is a realization of $X^{(j)}(t)$, solution of Model 1 with $p(t) = p^{(j)}(t)$
$p^{(j)}(t)$ is natural cubic-splines interpolation of $\mathbf{p}^{(j)} = \{p_{ji} = p^{(j)}(t_i)\}$

Observation equation:

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

- $\mathbf{G}^{(j)} = (G^{(j)}(t_1), \ldots, G^{(j)}(t_N))$
  $G^{(j)}(t)$ Gaussian approximation of $X^{(j)}(t)$ uniquely defined by:
  - $\mu^{(j)}(t) = \mathbb{E}[X^{(j)}(t)]$,
  - $v^{(j)}(t) = \mathbb{E}[\big(X^{(j)}(t) - \mu_j(t)\big)^2]$
  - $v^{(j)}(t,s) = \mathbb{E}[\big(X^{(j)}(t) - \mu_j(t)\big)\big(X^{(j)}(s) - \mu_j(s)\big)]$
- $\epsilon^{(j)} \sim \mathcal{N}(0, \mathbf{\Sigma}^{\epsilon,(j)})$, denotes the measurement error

KAUST

Approximate likelihood

The approximate likelihood function is

$$L(\boldsymbol{\theta}; \mathbf{d}_{JN}) = \prod_{j=1}^{J} (2\pi)^{-\frac{N}{2}} |\boldsymbol{\Sigma}^{(j)}(\boldsymbol{\theta})|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{d}^{(j)} - \boldsymbol{\mu}^{(j)}(\boldsymbol{\theta}))^T [\boldsymbol{\Sigma}^{(j)}(\boldsymbol{\theta})]^{-1} (\mathbf{d}^{(j)} - \boldsymbol{\mu}^{(j)}(\boldsymbol{\theta}))}$$

- $\boldsymbol{\mu}^{(j)}(\boldsymbol{\theta}) = (\mu^{(j)}(t_1), \ldots, \mu^{(j)}(t_N))$,
- $\boldsymbol{\Sigma}^{(j)}(\boldsymbol{\theta}) = \mathbf{V}^{(j)}(\boldsymbol{\theta}) + \boldsymbol{\Sigma}^{\epsilon,(j)}(\boldsymbol{\theta})$,
  - $\mathbf{V}^{(j)}(\boldsymbol{\theta})$ is the covariance matrix, $[\mathbf{V}^{(j)}(\boldsymbol{\theta})]_{kl} = v^{(j)}(t_k, t_l), k, l = 1, \ldots N$,
  - $\boldsymbol{\Sigma}^{\epsilon,(j)}(\boldsymbol{\theta}) = \phi^2 \mathbb{I}$

$|\boldsymbol{\Sigma}|$ denotes the determinant of the matrix $\boldsymbol{\Sigma}$.

$\boldsymbol{\theta} = (\theta(\cdot), \alpha, \phi)$ denotes the vector of parameters

Two-moment equations of Model 1

The equation of the mean is given by

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

The equation of the variance is

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

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

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

Equations are derived using Itô's Lemma.

Exact solutions

Mean of $X(t)$

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

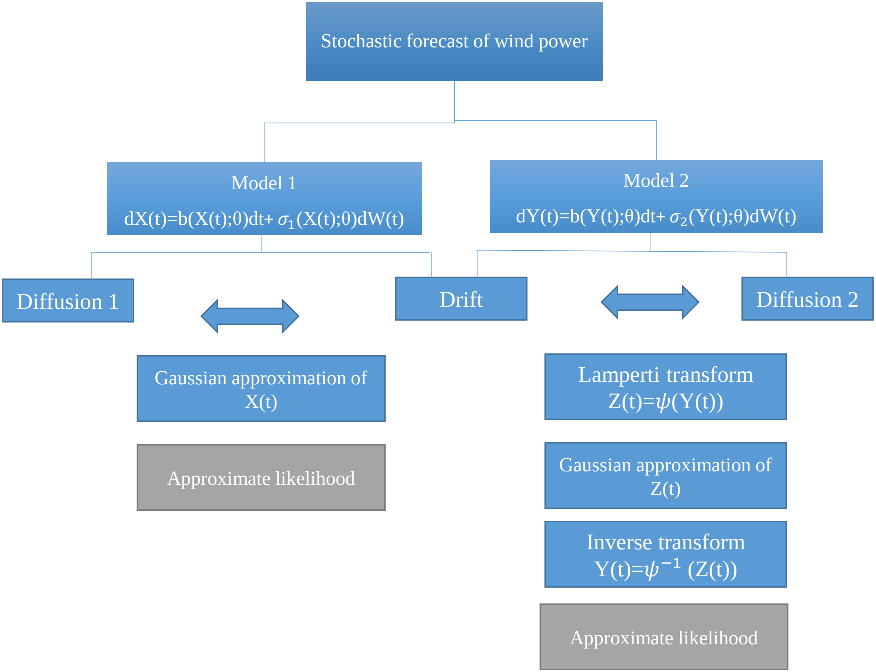
Variance $X(t)$

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

Covariance $X(t)$

$$v_X(t,s) = v_X(s)e^{-\int_0^t \theta(u)du}, \text{ for } t > s$$

KAUST



Stochastic forecast of wind power
Model 1
dX(t)=b(X(t);θ)dt+ σ_1(X(t);θ)dW(t)
Model 2
dY(t)=b(Y(t);θ)dt+ σ_2(Y(t);θ)dW(t)
Diffusion 1
Drift
Diffusion 2
Gaussian approximation of X(t)
Approximate likelihood
Lamperti transform Z(t)=ψ(Y(t))
Gaussian approximation of Z(t)
Inverse transform Y(t)=ψ^{-1} (Z(t))
Approximate likelihood

Model 2

Motivation: State independent diffusion.

How? Lamperti transform

$$\Psi(X(t)) = \int \frac{1}{\sigma_1(x)} dx \Big|_{x=X(t)}$$

If we transform Model 1 to a state-independent diffusion, the solution of the mean $\mathbb{E}\left[\Psi(X(t))\right]$ is not unique (i.e. likelihood is not well defined).

Condition: $\alpha \leq \min(p(t), 1-p(t)) \Rightarrow \alpha(t) = \tilde{\alpha} p(t)\Big(1-p(t)\Big)$
later ($\tilde{\alpha} = \alpha$)

Model 2

$$\text{Model 2} \begin{cases} dY(t) = b(Y(t);\boldsymbol{\theta})dt + \sigma_2(Y(t);\boldsymbol{\theta})dW(t), & t > 0\,, \\ Y(0) = y_0 \end{cases}$$

- **Form of the drift**:

$$b(Y(t), t; \boldsymbol{\theta}) = -\theta(t)(Y(t) - p(t))$$

- **Form of the diffusion**: The diffusion vanishes at the boundaries and the mean of the transformed process exists and is unique

$$\sigma_2(Y(t), t; \boldsymbol{\theta}) = \sqrt{2\theta(t)\alpha p(t)\Big(1 - p(t)\Big)Y(t)\Big(1 - Y(t)\Big)}$$

- Inference is based on the approximate likelihood using $\mu_Y(t)$, $v_Y(t)$ and $v_Y(t,s)$

Lamperti transform

The Lamperti transform of the process $Y(t)$ defines the process $Z(t)$ given by

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

and satisfies the following SDE:

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

where

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

and

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

Approximate two-moment equations

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

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

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

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

The equation of the covariance,
$v_Z(t,s) \approx \mathbb{E}\Big[\big(Z(t) - \mu_Z(t)\big)\big(Z(s) - \mu_Z(s)\big)\Big]$, is

$$\begin{cases} \frac{dv_Z(t,s)}{dt} = \tilde{b}'(\mu_Z(t))v(t)\,, \\ v_Z(s,s) = v_Z(s) \end{cases}$$

Mean of $Z(t)$

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

where $h(t) = \theta[1 - 2\alpha p(t)(1 - p(t))]$

Variance of $Z(t)$

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

Covariance of $Z(t)$

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

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

KAUST

Two-moment equations of Model 2

Using the inverse function of Ψ, the approximate mean, variance and covariance functions of $Y(t)$ are

Mean of $Y(t)$

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

Variance of $Y(t)$

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

Covariance of $Y(t)$

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

- Bias ($\tilde{\mu}_Y(t) \approx \mu_Y(t)$) introduced in the mean of Model 2 due to the use of the inverse Lamperti transform.

Plan

Benchmark examples

Steps

Benchmark examples

Steps

- Choose values for $\boldsymbol{\theta}$ and a time horizon, $T \in \{12, 24, 36, 48\}$.

Benchmark examples

Steps

- Choose values for $\boldsymbol{\theta}$ and a time horizon, $T \in \{12, 24, 36, 48\}$.
- Choose the predefined forecast function $p(t) \in [0, 1]$.

Benchmark examples

Steps

- Choose values for $\boldsymbol{\theta}$ and a time horizon, $T \in \{12, 24, 36, 48\}$.
- Choose the predefined forecast function $p(t) \in [0, 1]$.
- Sample M paths from X or paths from Z and transform them to Y using Ψ^{-1}.

Benchmark examples

Steps

- Choose values for $\boldsymbol{\theta}$ and a time horizon, $T \in \{12, 24, 36, 48\}$.
- Choose the predefined forecast function $p(t) \in [0, 1]$.
- Sample M paths from X or paths from Z and transform them to Y using Ψ^{-1}.
- Add noise to the data, $e_i \sim N(0, \Sigma_i^e)$, $\Sigma_i^e = \phi^2 \mathbb{I}$.

Benchmark examples

Steps

- Choose values for $\boldsymbol{\theta}$ and a time horizon, $T \in \{12, 24, 36, 48\}$.
- Choose the predefined forecast function $p(t) \in [0, 1]$.
- Sample M paths from X or paths from Z and transform them to Y using Ψ^{-1}.
- Add noise to the data, $e_i \sim N(0, \Sigma_i^e)$, $\Sigma_i^e = \phi^2 \mathbb{I}$.
- Compute the corresponding mean, variance and covariance functions.

Benchmark examples

Steps

- Choose values for $\boldsymbol{\theta}$ and a time horizon, $T \in \{12, 24, 36, 48\}$.
- Choose the predefined forecast function $p(t) \in [0, 1]$.
- Sample M paths from X or paths from Z and transform them to Y using Ψ^{-1}.
- Add noise to the data, $e_i \sim N(0, \Sigma_i^e)$, $\Sigma_i^e = \phi^2 \mathbb{I}$.
- Compute the corresponding mean, variance and covariance functions.
- Use the sampled synthetic data to maximize the likelihood function, compute the confidence interval of the optimal parameters.

Benchmark examples

Steps

- Choose values for $\boldsymbol{\theta}$ and a time horizon, $T \in \{12, 24, 36, 48\}$.
- Choose the predefined forecast function $p(t) \in [0, 1]$.
- Sample M paths from X or paths from Z and transform them to Y using Ψ^{-1}.
- Add noise to the data, $e_i \sim N(0, \Sigma_i^e)$, $\Sigma_i^e = \phi^2 \mathbb{I}$.
- Compute the corresponding mean, variance and covariance functions.
- Use the sampled synthetic data to maximize the likelihood function, compute the confidence interval of the optimal parameters.

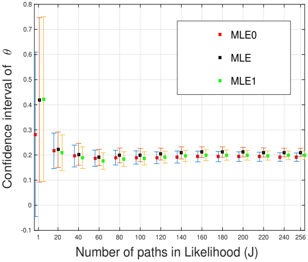
Numerical implementation

- Euler-Maruyama $\delta t = 0.01$.
- Trapezoidal quadrature with time step $\Delta t = 0.01$.
- 'fmincon' on MATLAB for the minimization procedure.

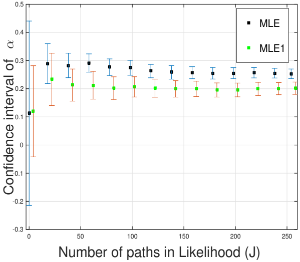
Benchmark example 1 (Model 1)

The trigonometric function $p^{(j)}(t) = \sin^2(t/6)$, $j = 1, \ldots, J = 2^8$,
$\boldsymbol{\theta} = (\theta, \alpha, \phi) = (0.2, 0.2, 0.05)$

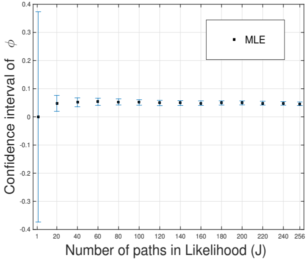
- MLE0: maximum likelihood estimate of θ for fixed $(\alpha, \phi) = (0.3, 0.02)$.
- MLE: maximum likelihood estimates of (θ, α) for fixed $\phi = 0.02$.
- MLE1: maximum likelihood estimates of (θ, α, ϕ).



Convergence of θ.



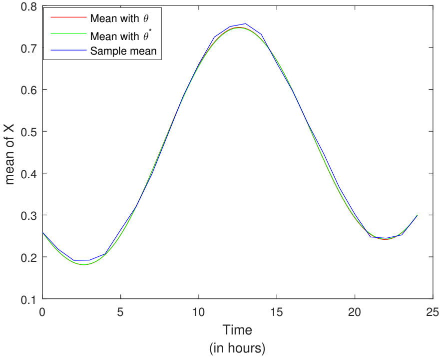
Convergence of α.



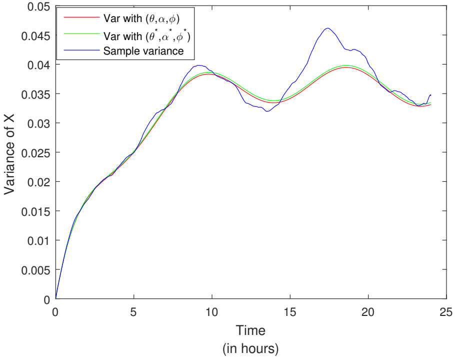
Convergence of ϕ.

- MLE1 is able to recover the true values $(\hat{\theta}, \hat{\alpha}, \hat{\phi}) = (0.1997, 0.2006, 0.0507)$

Benchmark example 1



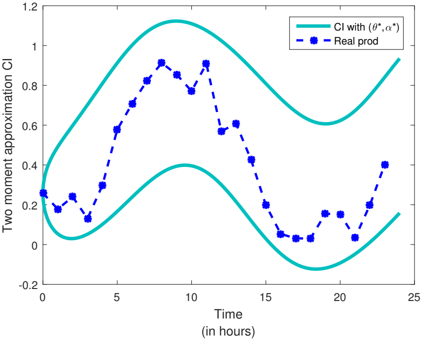
Mean vs sample mean with 2^8 synthetic paths.



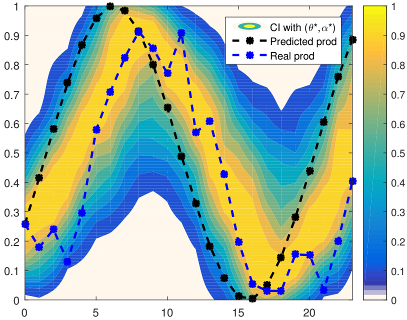
Variance vs sample variance with 2^8 synthetic paths.

- Gaussian approximation is consistent.
- Small perturbations of $\boldsymbol{\theta}$ effects the curves sightly.

Benchmark example 1



95% Confidence interval obtained with the two-moment equations.



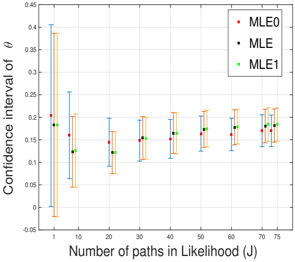
Empirical confidence bands with 2^8 synthetic paths.

- Initial value $x_0 = 0.258$.
- The real production path falls within the 95% Empirical Confidence Bands.

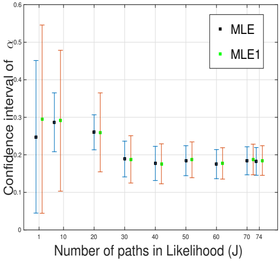
Benchmark example 2 (Model 1)

The functions $p^{(j)}(t),\ j = 1, \ldots, J = 73$, are fitted from the available numerical forecast data.

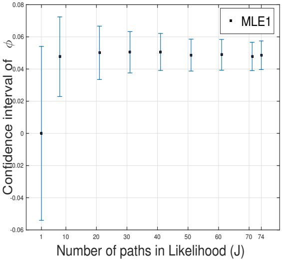
$\boldsymbol{\theta} = (\theta, \alpha, \phi) = (0.2, 0.2, 0.05)$



Convergence of θ.



Convergence of α.

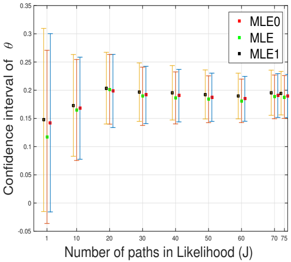


Convergence of ϕ.

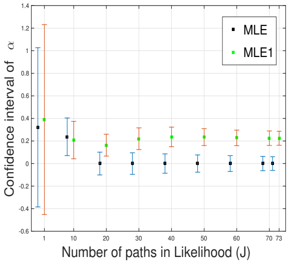
- $(\hat{\theta}, \hat{\alpha}, \hat{\phi}) = (0.1865, 0.1970, 0.0549)$
- The confidence intervals are relatively larger than those obtained in benchmark example 1 as a result of lack of data.

Benchmark example 2 (Model 2)

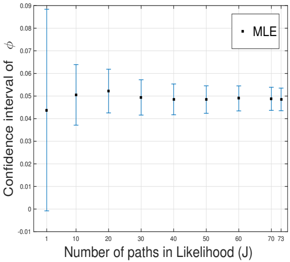
$$\boldsymbol{\theta} = (\theta, \alpha, \phi) = (0.2, 0.2, 0.05)$$



Convergence of θ.



Convergence of α.

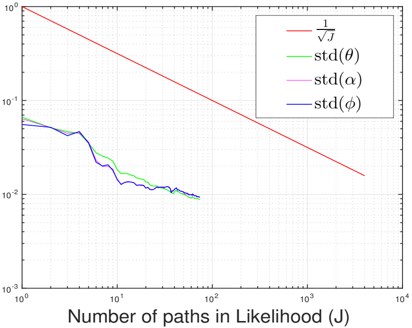


Convergence of ϕ.

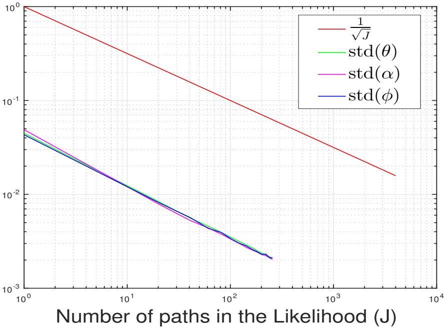
- $(\hat{\theta}, \hat{\alpha}, \hat{\phi}) = (0.188, 0.224, 0.048)$
- Model 2 retrieves the true values with fewer paths compared to Model 1.

Comparison of the models

Trigonometric forecast: comparison of the rates of convergence for maximum likelihood estimators of Model 1 and Model 2.



Model 1 convergence of the MLE estimators.

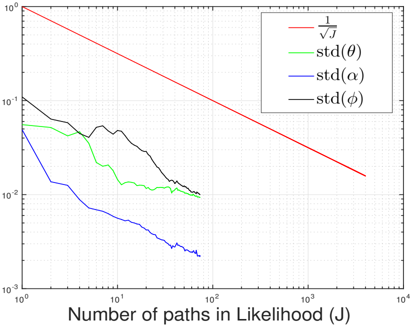


Model 2 convergence of the MLE estimators.

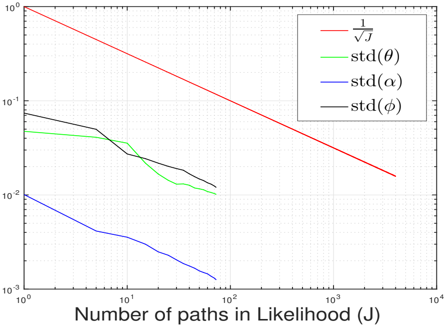
J =size of the training set, N=number of observations per set.
Asymptotic normality of the estimators $\hat{\boldsymbol{\theta}}$, $J \to \infty$, $\sqrt{J}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim \mathcal{N}(0, \mathcal{I}^{-1}(\boldsymbol{\theta}))$
$\mathcal{I}(\boldsymbol{\theta}) = -\mathbb{E}[\frac{\partial^2 \log L(\boldsymbol{\theta}; \boldsymbol{d}_{JN})}{\partial^2 \theta}]$, the Fisher Information.

Comparision of the models

Fitted real forecast: comparison of the rates of convergence for maximum likelihood estimators of Model 1 and Model 2.



Model 1 convergence of the MLE estimators.



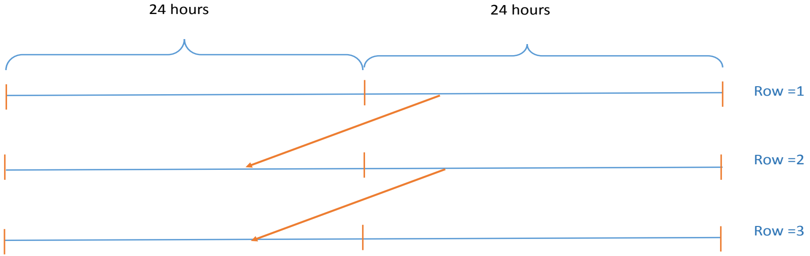
Model 2 convergence of the MLE estimators.

- Model 2 is numerically more stable.

Numerical results using real power production in Uruguay

Preprocessing of the data

- Remove redundant sets to ensure the independence assumption in the likelihood



- Discard time series where the difference between d_{ji} and p_{ji} is more than 20%.

Real data

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

Horizon	Parameters	Model 1	Model 1 (artificial noise, $\phi = 0.05$)	Model 2
12 h	$(\theta^*, \alpha^*, \phi^*)$ std	$(0.177, 0.082)$ $10^{-2}(0.83, 0.41)$	$(0.256, 0.106, 0.036)$ $10^{-2}(1.75, 0.7, 0.25)$	$(0.138, 0.108, 0.087)$ $10^{-2}(0.79, 1.87, 0.07)$
18 h	$(\theta^*, \alpha^*, \phi^*)$ std	$(0.140, 0.089)$ $10^{-2}(0.83, 0.41)$	$(0.213, 0.102, 0.038)$ $110^{-2}(1.2, 0.59, 0.19)$	$(0.127, 0.569, 0.003)$ $10^{-2}(0.41, 1.49, 0.04)$
24 h	$(\theta^*, \alpha^*, \phi^*)$ std	$(0.140, 0.087)$ $10^{-2}(0.47, 0.31)$	$(0.196, 0.118, 0.036)$ $10^{-2}(0.96, 0.57, 0.17)$	$(0.115, 0.602, 0.03)$ $10^{-2}(0.28, 1.16, 0.03)$
36 h	$(\theta^*, \alpha^*, \phi^*)$ std	$(0.121, 0.109)$ $10^{-2}(0.47, 0.47)$	$(0.172, 0.147, 0.034)$ $10^{-2}(0.96, 0.57, 0.17)$	$(0.095, 1.008, 0.019)$ $10^{-2}(0.39, 4.52, 0.12)$
48 h	$(\theta^*, \alpha^*, \phi^*)$ std	$(0.114, 0.105)$ $10^{-2}(0.39, 0.39)$	$(0.154, 0.156, 0.035)$ $10^{-2}(0.73, 0.76, 0.17)$	$(0.087, 0.993, 0.016)$ $10^{-2}(0.31, 3.98, 0.09)$

- Fit a decreasing function of time to θ.

Real data

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	$\theta_3(t) = \theta_0 e^{-\theta_1 t} + \theta_2 e^{\theta_3 t}$	6

Model 1

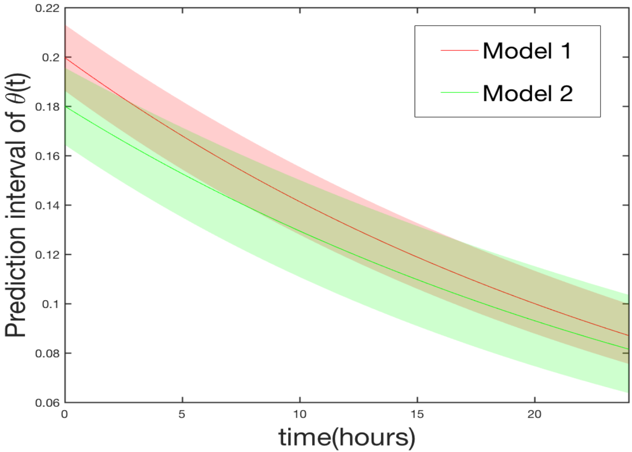
Criterion	AIC	BIC
SDE0	-8418	-8411
SDE1	-8427	-8418
SDE2	-8426	-8414
SDE3	-8273	-8259

Model 2

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

- SDE1 is the “best” model as it has the smallest AIC and BIC values.

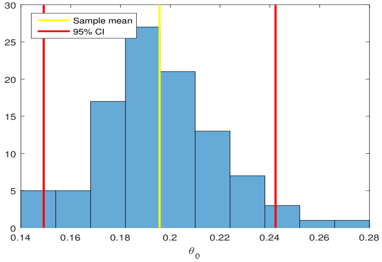
Real data

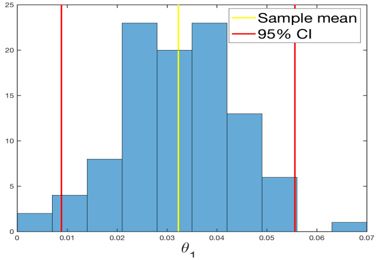


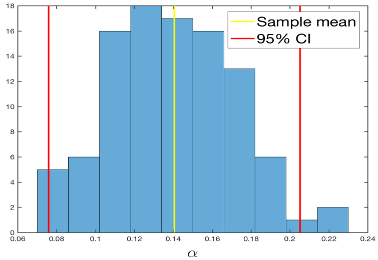
Best fit for SDE1 for the two models.

- The prediction curve of Model 1 falls inside the prediction interval of Model 2, confirming that both predict the same rate.

Variability of Model 1



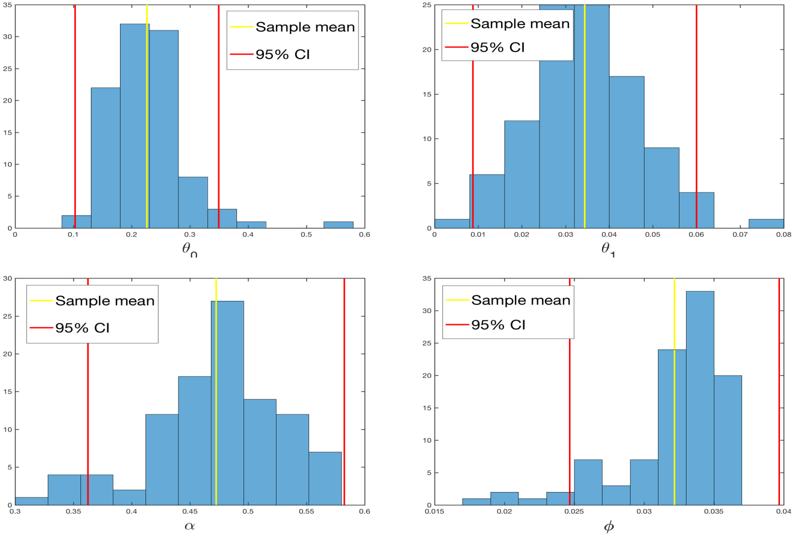




Histograms of θ_0, θ_1 and α based on $n = 100$ bootstrap samples.

- $(\hat{\theta}_0, \hat{\theta}_1, \hat{\alpha}) = (0.199, 0.034, 0.141)$

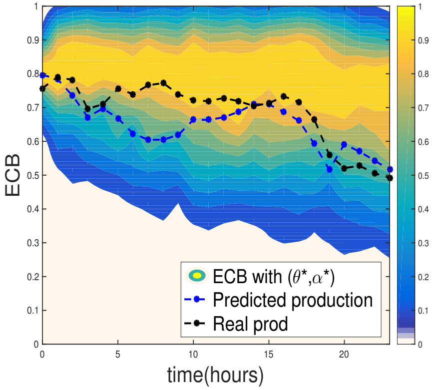
Variability of Model 2



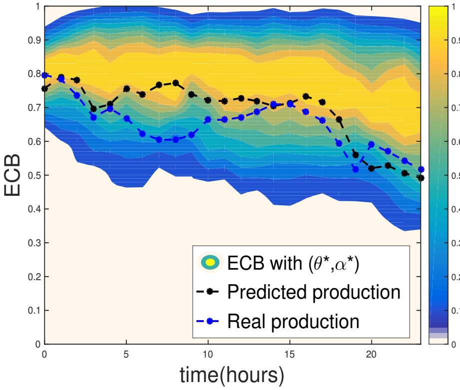
Histograms of θ_0, θ_1, α and ϕ based on $n = 100$ bootstrap samples.

- $(\hat{\theta}_0, \hat{\theta}_1, \hat{\alpha}, \hat{\phi}) = (0.218, 0.033, 0.453, 0.033)$

Validation of the models



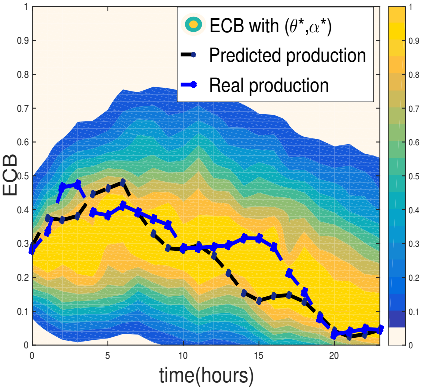
ECB and real production on May 16,2016 for Model 1.



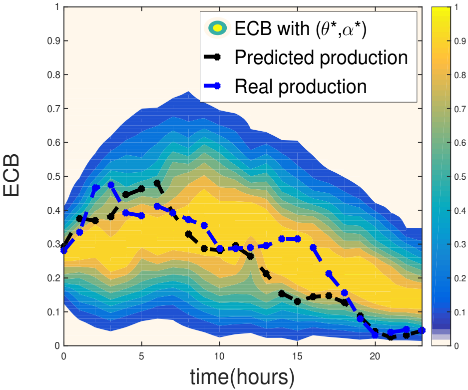
ECB and real production on May 16,2016 for Model 2.

- Empirical Confidence Bands (ECB) sampled from estimated parameters of the training sets, 80% of the data and real production from the testing sets.

Validation of the models



ECB and real production on May 22,2016 for Model 1.



ECB and real production on May 22,2016 for Model 2.

- Compared to Model 1, Model 2 provides more precise Empirical Confidence Bands (ECB) .

Sensitivity

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

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

$F(X(t);\theta) = X(T;\theta),$ is the normalized power production at time T.
The sensitivity of the proposed models with respect to the parameter θ through the derivative, $f'(\theta)$.

- The Backward Difference approximation

$$S_{\text{BD}} = f'(\theta) \approx \frac{f(\theta) - f(\theta-\delta)}{\delta}.$$

- The Forward Difference approximation

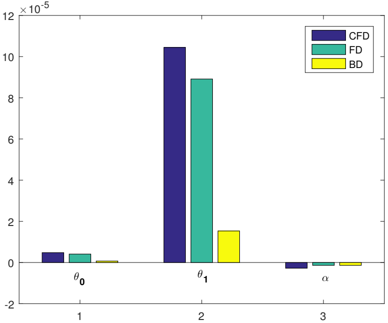
$$S_{\text{FD}} = f'(\theta) \approx \frac{f(\theta+\delta) - f(\theta)}{\delta}.$$

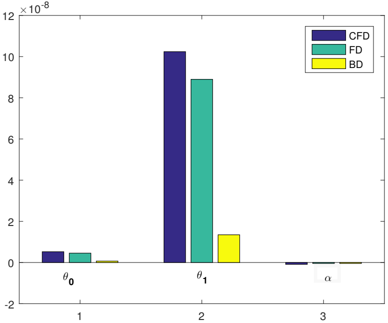
- The Central Finite Difference approximation

$$S_{\text{CFD}} = f'(\theta) \approx \frac{f(\theta+\delta) - f(\theta-\delta)}{2\delta}$$

Sensitivity

Finite Differences





Sensitivity of parameters for Model 1 Sensitivity of parameters for Model 2

- δ is 10% change in the value of $\hat{\theta}$.
- 10^4 sampled paths from real $p^{(j)}(t)$ and common random numbers.
- The most sensitive parameter for both models is θ_1.
- The least sensitive parameter is α.

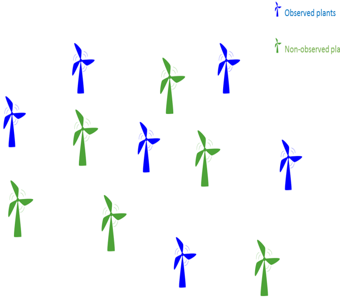
Conclusion

- Deterministic forecast + historical data of wind power production $\Rightarrow$ Confidence bands for the forecast.
- Model 1 : Inference based on state-dependent diffusion.
- Model 2 : Inference based on state-independent diffusion.
- Model 1 is not able to capture the value of the measurement error for all $\theta_i(t),\ i \in \{1, 2, 3\}$ and its parameters are very sensitive to an artificial error.
- Model 2 provides an estimate of the measurement error.
- Model 2 is more stable, we assume that this is related to the use of Lamperti-transformed process (bias).
- Both models predict the same rate.
- Next step: (a) Fit new datasets of one year (b) Numerical results of the partially observed stochastic forecast.

KAUST

Partially observed model

$Y(t)$ observed plants
$Z(t)$ non-observed plants



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

- $dW_1(t) \cdot dW_2(t) = \rho dt$
- $b_1(Y(t);\boldsymbol{\theta}) = -\theta_1(t)(Y(t) - p_1(t))$
  $b_2(Z(t);\boldsymbol{\theta}) = -\theta_2(t)(Z(t) - p_2(t))$
- $\sigma_1(Y(t);\boldsymbol{\theta}) = \sqrt{2\theta_1(t)\alpha_1 Y(t)(1-(Y(t)))}$
  $\sigma_2(Z(t);\boldsymbol{\theta}) = \sqrt{2\theta_2(t)\alpha_2 Z(t)(1-(Z(t)))}$

$$\boldsymbol{\theta} = (\theta_1(\cdot), \theta_2(\cdot), \alpha_1, \alpha_2, \rho)$$

Thank you for your attention

References

G. C. S. de Mello and A. G. Arce, “Operational wind energy forecast with power assimilation,” *14th International Conference on Wind Engineering*, 2014.

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.

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.

J.K. Moller, M. Zugno, and H. Madsen, ”Probabilistic forecasts of wind power generation by stochastic differential equation models,” vol. 35, no. 3, pp. 189–224, 2016.

- M. Ballesio, “Indirect inference for scalar time-homogeneous,” *Masters Thesis*.
- B. Øksendal, *Stochastic Diffrential Equations*. Springer, 2000.
- S. Asmussen and P. Glynn, “Stochastic Simulation: Algorithms and Analysis,” Springer, 2007.
- S. M. Iacus, *Simulation and Inference for Stochastic Differential Equations (with R examples)*. Springer Series in Statistics, 2008.
- G. Durham and R. Gallant, “Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes,” vol. 20, no. 3, 2002.
- J.K. Moller and H. Madsen, ’From state dependent diffusion to constant diffusion in stochastic differential equations by the Lamperti transform,” 2010. DTU, IMM-Technical Report-2010-16
- A. Moraes, “Simulation and statistical inference of stochastic reaction networks with applications to epidemic models,” 2015.

Appendix

Let $\hat{L} = L(\hat{\theta}; \mathbf{d}_{JN})$ be the likelihood evaluated at $\hat{\theta}$ and k be the number of free parameters in the model.
The AIC and the BIC statistics are defined as, see [**?**]

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

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

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