Wind project inference

Khaoula Ben Chaabane

Ecole Polytechnique de Tunisie

28/05/2020

In these slides, I will explain the inference process with which we estimate the parameters of our model.

The wind power production is modeled as follows, where X_t is the normalized real production :

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

We may introduce the following model for the forcecast error of the normalized wind power production where X_t is the real production, p_t the forecast and $V_t = X_t - p_t$ is the error :

$$\begin{cases} dV_t = -\theta_t V_t dt + \sqrt{2\alpha\theta_0 (V_t + p_t) (1 - V_t - p_t)} dW_t, & t \in [0, T] \\ V_0 = v_0 \in [-1 + \varepsilon, 1 - \varepsilon] \end{cases}$$

$$(1)$$

Model

To guarantee a unique solution for the process X_t, θ_t needs to be bounded for $t \in [0, T]$. We have that :

In my latest results, I removed the factor 2 from the equation of theta_t. When you have time, please, remove it and find the optimal parameters theta_0 and alpha again. We may reach the same values. However, what you did is NOT wrong since you can choose any theta_t that satisfies the conditions of existence for the SDE.

$$\theta_t = \max\left(\theta_0, \frac{\alpha\theta_0 + |2\dot{p}_t|}{2\min(1-p_t, p_t)}\right)$$

This is not true for θ_t if $p_t \to 0^+$ or $p_t \to 1^-$. Therefore we need to ensure that $p_t \in [\varepsilon, 1-\varepsilon]$ for some $0 < \varepsilon < \frac{1}{2}$, $\forall\ t \in [0, T]$.

Model

We define then the corrected forecast :

$$p_t^\varepsilon = \begin{cases} \varepsilon & \text{if} & p_t < \varepsilon \\ p_t & \text{if} & \varepsilon \le p_t < 1 - \varepsilon \\ 1 - \varepsilon & \text{if} & p_t \ge 1 - \varepsilon \end{cases}$$

and the corrected (and bounded) drift coefficient is therefore :

$$\theta_t^\varepsilon = \max\left(\theta_0, \frac{\alpha\theta_0 + 2|\dot{p}_t^\varepsilon|}{2\min\left(1 - p_t^\varepsilon, p_t^\varepsilon\right)}\right)$$

Likelihood

We sample each of our M continuous-time Itô process $V = (V_t)_{t \in [0,T]}$ at $N + 1$ equidistant discrete points with a given length interval Δ. $V^{M,N+1} = \left\{ V_{t_1}^{N+1}, V_{t_2}^{N+1}, \ldots, V_{t_{M+1}}^{N+1} \right\}$ denotes this random sample, with $V_{t_j}^{N+1} = \left\{ V_{t_j + i\Delta}, i = 0, \ldots, N \right\}, \forall j \in \{1, \ldots, M\}$.

Let $\rho(v | v_{j,i-1}; \boldsymbol{\theta})$ be the conditional probability density of $V_{t_j + i\Delta} \equiv V_{j,i}$, given $V_{j,i-1}$ where $\boldsymbol{\theta} = (\theta_0, \alpha)$ are the unknown model parameters.

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

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

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

Likelihood approximation

In order to compute the exact likelihood function, we need a closed-form expression of the transition probability of V which can be found using the Fokker-Planck equation :

$$\frac{\partial f}{\partial t}\rho\left(v,t|v_{j,i-1},t_{j,i-1};\theta\right) = -\frac{\partial}{\partial v}\left(-\theta_t v\rho\left(v,t|v_{j,i-1},t_{j,i-1};\theta\right)\right)$$
$$+\frac{1}{2}\frac{\partial^2}{\partial v^2}\left(2\theta_0\alpha\left(v+p_t\right)\left(1-v-p_t\right)\rho\left(v,t|v_{j,i-1},t_{j,i-1};\theta\right)\right)$$

However, solving this equation is not always possible and is computationally costly. For this reason, we approximated the likelihood using a proxy distribution for V.

In our case we used a Beta distribution (ξ_1, ξ_2) as for the family of diffusion term in our SDE (1) (Pearson diffusion), it has been proved to be the best approximation. In order to find the parameters (ξ_1, ξ_2) of this proxy distribution we will match its first and second moments with the ones of the exact distribution deduced from the SDE (1).

For a time $s \in [t_n, t_{n+1}]$ the exact first moment $m_1(s)$ deduced from the SDE (1) is the solution of the following ODE :

$$\begin{cases} \mathrm{d}m_1(s) = [-m_1(s)\theta(s)] \, \mathrm{d}s \\ m_1(t_{n-1}) = v_{t_{n-1}} \end{cases}$$

We want to compute $m_1(t_n)$:

▶ If $\theta(t_n) = \theta(t_{n+1}) = \theta$ then the exact solution is :
$m_1(t_n) = m_1(t_{n-1}) \exp(-\theta(t_n - t_{n-1}))$

▶ else, we compute a linear approximation of $\theta(s)$ and approximate the ODE using Forward-Euler :

$$m_1(s_n) = m_1(s_{n-1})(1 - \theta(s_{n-1})\Delta s)$$

Moment matching : Second moment

Using Ito's formula, we find, for a time $s \in [t_n, t_{n+1}]$, the exact second moment $m_2(s)$ deduced from the SDE (1) is the solution of the following ODE :

$$
\begin{cases}
\mathrm{d}m_2(s) = [-2m_2(s)(\theta(s) + \alpha\theta_0) + 2\alpha\theta_0 m_1(s)(1 - 2p(s)) \\
+2\alpha\theta_0 p(s)(1 - p(s))\mathrm{d}s \\
m_2(t_{n-1}) = v_{t_{n-1}}^2
\end{cases}
\quad .
$$

We compute a linear interpolation for the functions $\theta(s)$ and $p(s)$. After, we solve the ODE using Forward-Euler :
$m_2(s_n) = m_2(s_{n-1}) + [-2m_2(s_{n-1})(\theta(s_{n-1}) + \alpha\theta_0)$
$+2\,\alpha\theta_0 m_1(s_{n-1})(1 - 2p(s_{n-1})) + 2\alpha\theta_0 p(s_{n-1})(1 - p(s_{n-1}))\Delta s$

We use the same discretization points for both $m_1(s)$ and $m_2(s)$.

Moment matching

V is approximated by a new proxy random variable : $V = a + (b - a)X$ with support in $[a, b] = [-1, 1]$, where $X \sim \beta(\xi_1, \xi_2)$ and PDF $f_V(v)$. We find the two first moments :

- $\mathbb{E}[V] = a + (b - a)\mathbb{E}[X] = a + (b - a)\frac{\xi_1}{\xi_1 + \xi_2} = \mu_V$
- $\mathbb{V}[V] = (b - a)^2\mathbb{V}[X] = \frac{(b-a)^2\xi_1\xi_2}{(\xi_1 + \xi_2)^2(\xi_1 + \xi_2 + 1)} = \sigma_V^2$

We want the first two moments of the true random variable and its approximation to be equal $\forall t$.
Therefore, $\mu(t) = m_1(t)$ and $\sigma^2(t) = m_2(t) - m_1^2(t)$.

For each measurement $V_{t_{n-1}}$, we can find the analytical moments at time t_n solving the ODEs from the previous slides. We can then find the parameters ξ_1 and ξ_2 of the proxy.

Evaluation of (ξ_1, ξ_2)

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

all evaluated at time t_n

Log-density of the proxy random variable V

We want to compute the PDF $f_V(v)$ of the random variable :
$V = a + (b-a)X$.

For $[a, b] = [-1, 1]$, we have that :
$f_V(v) = f_X\left(g^{-1}(v)\right)\left|\frac{d}{dv}g^{-1}(v)\right|$ where $f_X(x) = \text{Beta}\left(\xi_1, \xi_2\right)$
and $g(x) = a + (b-a)x$.

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

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

Log-likelihood

We introduce the number of paths (days) M, and the number of measurements per path $N + 1(N$ transitions). We have a total of $M \times N$ samples. The log-likelihood is :

$$\mathfrak{L}\left(\{V\}_{M,N}\right) = \sum_{i=1}^{M} \sum_{j=2}^{N+1} \log\left[\rho_{i,j}\left(V_{i,j}|V_{i,j-1}\right)\right]$$

where $\rho_{i,j}\left(V_{i,j}|V_{i,j-1}\right) = \rho_{i,j}\left(V_{i,j}|V_{i,j-1}; \xi_{1,j}, \xi_{2,j}\right)$.

Initial Estimation of the parameters

In order to evaluate the initial parameters of our model we apply the least square method on the forecast error V_t.

We consider the transition $\Delta V_i = V_{i+1} - V_i$ with $\Delta t = t_{i+1} - t_i$.

$(V_{i+1}|V_i)$ is a random variable which conditional mean can be approximated by the solution of the following system :

$$\begin{cases} \mathrm{d}\mathbb{E}[V] = -\theta_t^\varepsilon \mathbb{E}[V]\mathrm{d}t \\ \mathbb{E}\left[V\left(t_i\right)\right] = V_i \end{cases}$$

evaluated in t_{i+1} (i.e., $\mathbb{E}\left[V\left(t_{i+1}\right)\right]$).

Then, the random variable $(V_{i+1} - \mathbb{E}\left[V\left(t_{i+1}\right)\right])$ has a mean equal to 0 approximately.

If we assume that $\theta_t^{\varepsilon} = c \in \mathbb{R}^+$ for all $t \in [t_i, t_{i+1}]$, then $\mathbb{E}\left[V\left(t_{i+1}\right)\right] = V_i e^{-c\Delta t}$.

If we have a total of n transitions, we can write the regression problem for the conditional mean with L^2 loss function as :

$$c^* \approx \arg\min_{c \geq 0} \left[\sum_{i=1}^{n} (V_{i+1} - \mathbb{E}\left[V\left(t_{i+1}\right)\right])^2\right]$$

$$= \arg\min_{c \geq 0} \left[\sum_{i=1}^{n} \left(V_{i+1} - V_i e^{-c\Delta t}\right)^2\right] \qquad (2)$$

Least Square Minimization : LSM

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

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

and introduce it in equation (1). We get

$$c^* \approx \arg\min_{c \geq 0} \underbrace{\left[\sum_{i=1}^{n}\left(V_{i+1} - V_i(1 - c\Delta t)\right)^2\right]}_{=f(c)}$$

As $f(c)$ is convex in c, solving (5) (finding c^*) is equivalent to solving

$$\frac{\partial f}{\partial c}(c^{**}) = 0$$

and choosing $c^* = \max\{0, c^{**}\}$

Remove c^{**} and the definition with the maximum. It has no sense to choose c^*=0 if c^{**}<0 because we do not want a drift that makes the SDE scape from the forecast (I know I put that in the slides, but I do not like it anymore).

Least Square Minimization : LSM

$$\frac{\partial f}{\partial c} = \sum_{i=1}^{n} 2\left(-V_i\right)\left(-\Delta t\right)\left(V_{i+1} - V_i\left(1 - \theta_0 \Delta t\right)\right)$$

$$= \sum_{i=1}^{n} 2V_i \Delta t\left(V_{i+1} - V_i(1 - c\Delta t)\right)$$

$$= \sum_{i=1}^{n} 2V_{i+1}V_i \Delta t - 2V_i^2 \Delta t + 2V_i^2 \Delta t^2 c$$

Then, c^{**} satisfies the following :

$$c^{**} \approx \frac{\sum_{i=1}^{n} V_i\left(V_i - V_{i+1}\right)}{\Delta t \cdot \sum_{i=1}^{n}\left(V_i\right)^2}$$

Quadratic variation

We approximate the SDE by its E-M scheme. In particular, we approximate the Itô quadratic variation with the discrete one :

- ▶ Itô process quadratic variation : $[V]_t = \int_0^t \sigma_s^2 \mathrm{d}s$
- ▶ Discrete process quadratic variation : $[V]_t = \Sigma_{0 < s \le t} (\Delta V_s)^2$

Then, considering Δt the time between the measurements, we approximate :

$$\theta_0^* \alpha^* \approx \frac{\sum_{i=1}^n (\Delta V_i)^2}{2\Delta t \sum_{i=1}^n (V_i + p_i)(1 - V_i - p_i)}$$

In this section, we will use the approximation made previously to estimate the parameters $(\theta_0, \alpha, \varepsilon)$ of the SDE. Let us define $(\theta_0^*, \alpha^*, \varepsilon^*)$ as their estimators.

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

I = $\{i \in \{1, \ldots, n\} :$ the LSM estimation will estimate $\theta_0\}$
J = $\left\{j \in \{1, \ldots, n\} :$ the LSM estimation will estimate $\frac{\theta_0 \alpha}{\varepsilon}\right\}$
We will proceed then to approximate these sets in order to estimate our parameters.

Estimation of $(\theta_0, \alpha, \varepsilon)$

To use the LSM estimation, we assumed that $\theta_t^\varepsilon = c \in \mathbb{R}^+$, and we defined θ_t^ε :

$$\theta_t^\varepsilon = \max\left(\theta_0, \frac{\alpha\theta_0 + 2\left|\dot{p}_t^\varepsilon\right|}{2\min\left(1 - p_t^\varepsilon, p_t^\varepsilon\right)}\right)$$

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

$$\mathsf{J} \approx \mathsf{J} = \left\{j \in \{1, \ldots, n\} : p_j^\varepsilon \in \{\varepsilon, 1 - \varepsilon\}\right\}$$

and θ_t^ε, we have that it is more likely that $\theta_t^\varepsilon = \theta_0$ if $p_t^\varepsilon \approx \frac{1}{2}$. Then, we can approximate I by

$$\mathsf{I} \approx \tilde{\mathsf{I}} = \left\{i \in \{1, \ldots, n\} : p_i \in (\gamma, 1 - \gamma)\right\}, \gamma \approx \frac{1}{2}, \gamma < \frac{1}{2}$$

With the previous approximation made of the quadratic variation we can estimate $\theta_0 * \alpha* = 0.094$ therefore, with our given estimation of θ_0* we find that : $\alpha* = 0.08$

Estimation of ε^*

Now that we have an approximated value of $\theta_0 \alpha$, if we can estimate $\frac{\theta_0 \alpha}{\varepsilon}$, then we can estimate ε. We showed previously that for $\varepsilon << 1$, the LSM estimation using indexes from J is an estimator for $\frac{\theta_0 \alpha}{\varepsilon} =: k$
The goal is to find values for ε that satisfy $\varepsilon << 1$. For that we start by randomly choosing a small initial value for ε (that we will call ε_0), and iterating we aim to converge to some local minimum. We proceed with the following steps :

▶ We sample ε_0 from $U[0.01, 0.1]$ and load $\varepsilon \leftarrow \varepsilon_0$

▶ We create $\tilde{J}$ and use the LSM estimation to find k.

  ▶ If $k < \theta_0^*$, then the assumption $\theta_t^\varepsilon = c \in \mathbb{R}^+$ is wrong and we reduce the value of ε, i.e., $\varepsilon \leftarrow \varepsilon * 0.999$.

  ▶ If $k \geq \theta_0^*$, we load
    $\varepsilon \leftarrow \frac{\theta_0^* \alpha^*}{k}$ (we allow a maximum relative change of 1%).

  We repeat this step 100 times.

▶ We repeat steps 1 and 2, 50 times.

Initial parameters estimation

To conclude, the estimations of the SDE parameters that we found are : $(\theta_0^*, \alpha^*, \varepsilon^*) = (1.25, 0.08, 0.018)$.

The code computing this process can be found in the file **Wind_project_intial_guess.ipynb**.

Log-likelihood optimization

I minimized the negative log-likelihood using the function fmin from the library scipy.optimize in Python to estimate the parameters (θ_0, α). I found the following results for the different datasets provided :

Data providers	θ_0	α	$\theta_0\alpha$
Complete	1.161	0.0718	0.083
UTEP5	1.357	0.0809	0.108
MTLOG	1.175	0.0856	0.100
AWSTEP	1.196	0.0846	0.101

The code of this optimization can be found in the file
Wind_project_optimization.ipynb

In order to have $E(X) = p$ at all times, our SDE (1) needs to have a starting point where $V = 0$ ($X = p$).

However, we noticed that with our data, the error V at $t = 0$ is not 0. Therefore, to still unsure we have this property, we suppose that there existed a time $t_\delta < 0$ in the past where the condition $X = p$ was verified. (This can be justified as we assume that the forecast providers had at some point perfect information on the data).

.

Time parameter δ

To do this, we will linearly interpolate the forecast to a time t_δ $(t_0 - t_\delta = -\delta)$, and for the transition $V_{t_0}^i | V_{t_\delta}^i \ \forall i \in \{1, \ldots, M\}$ we will solve the following problem :

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

We will approximate again the density with a distribution and redo the same steps as before of moment matching to find its parameters.

We finally find that $\delta = 0.0837 = 120$ minutes.

Excellent.

Path simulation

I plotted the path for the real production, the forecast and some simulations of the production using the model.
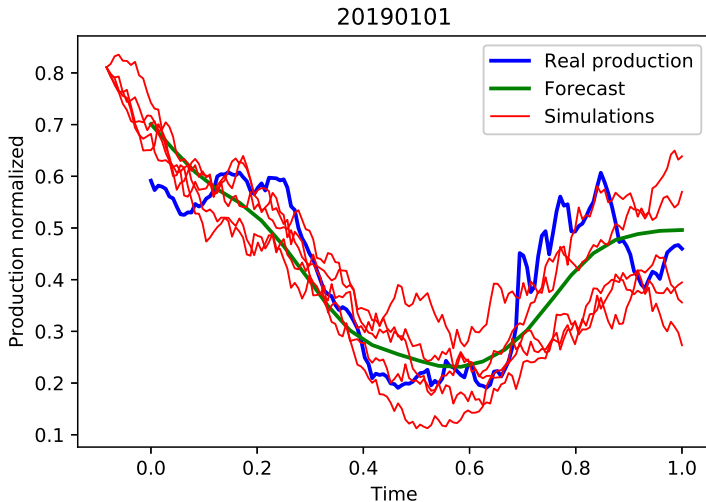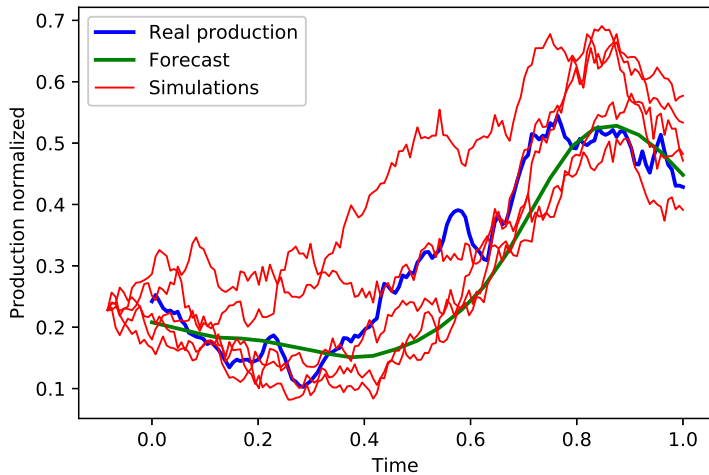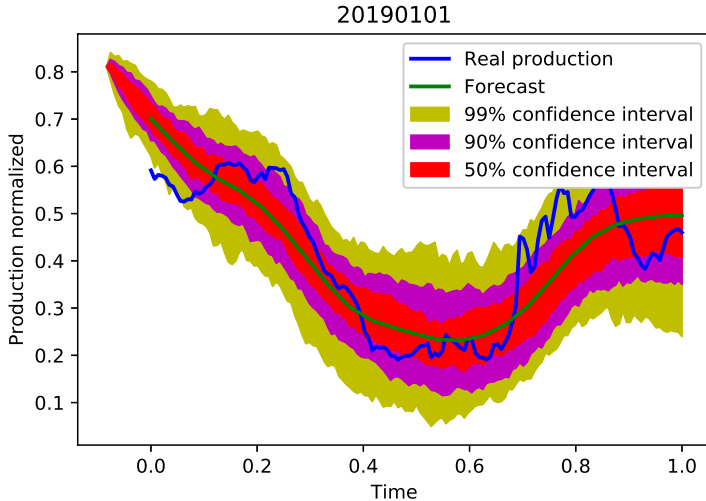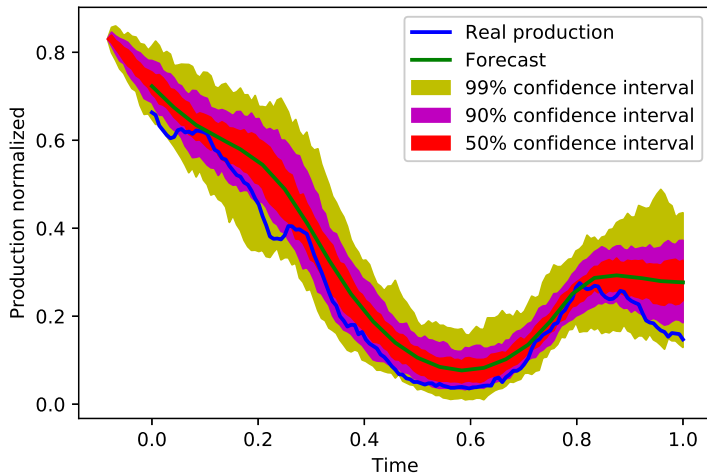
20190218
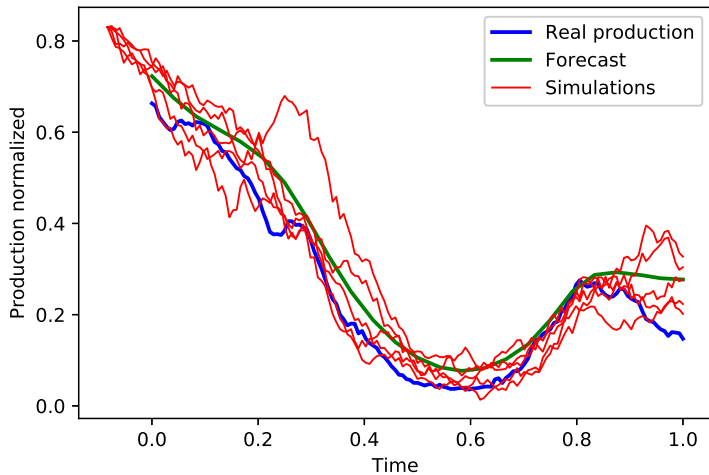
20190326
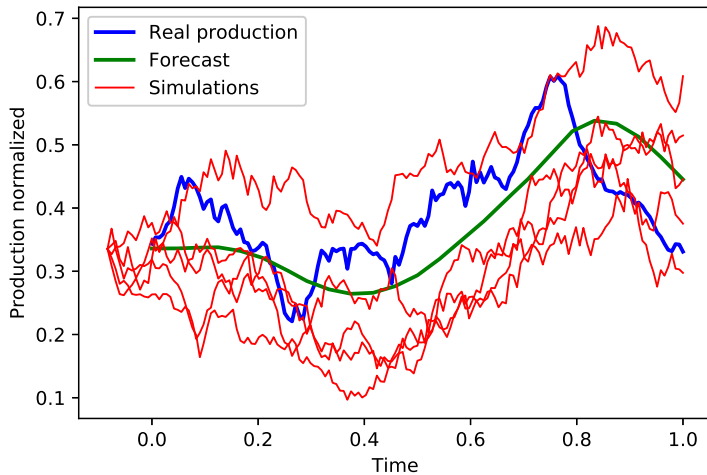
20190422

Confidence intervals

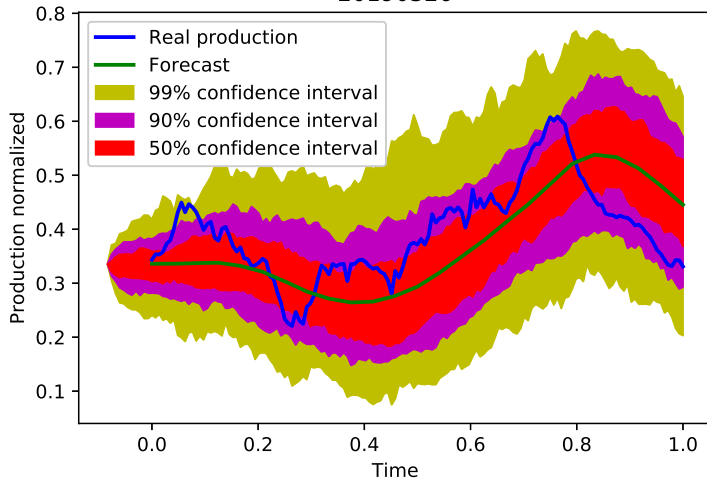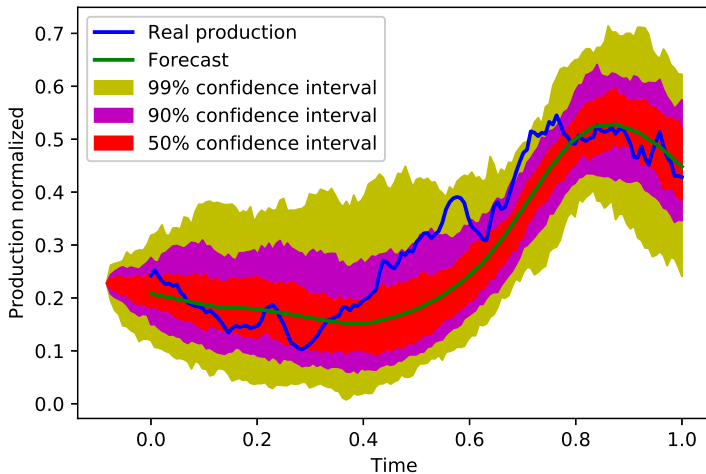I plotted the 99%, 90% and 50% confidence intervals using 100 simulations per day.

20190101

20190218

20190326

20190422

Excellent.