Probabilistic Forecasts of Wind Power Generation by Stochastic Differential Equations Models

Jan Kloppenborg Møller
Emil Banning Iversen
Henrik Madsen

Technical University of Denmark
Department of Applied Mathematics and Computer Science
hmad@dtu.dk

Outline

Outline

Technical University of Denmark DTU

Motivation

An increasing part of electricity supply in Denmark is generated by wind

- Wind power cover about 29% of total system load
- Danish goal: Renewables should cover 50% in 2020 and 100% of total system load in 2035

With the large penetration of wind accurate forecasts (including uncertainties) are needed on all time scales:

- Minutes - few hour: efficient and safe regulation
- 12-36 hour: efficient trading on NordPool
- Days: optimal regulation of large CHP

We focus on horizons from 1-48 hours.

Methods used for Wind Power Forecasting

- Adaptive time series model, using MET-forecasts as input
- Regime models (SETAR, STAR, MSAR)
- Spatio-temporal models
- Combining several MET-forecast
- Corrected MET ensembles
- Time-adaptive quantile regression
- Scenario based forecasting (dependence structure by corr. matrix or copula)
- Stochastic differential equations

Most of these methods are implemented in various modules of the **Wind Power Prediction Tool (WPPT)**.
WPPT is used operationally in Denmark (since 1995).

WPPT point forecasts

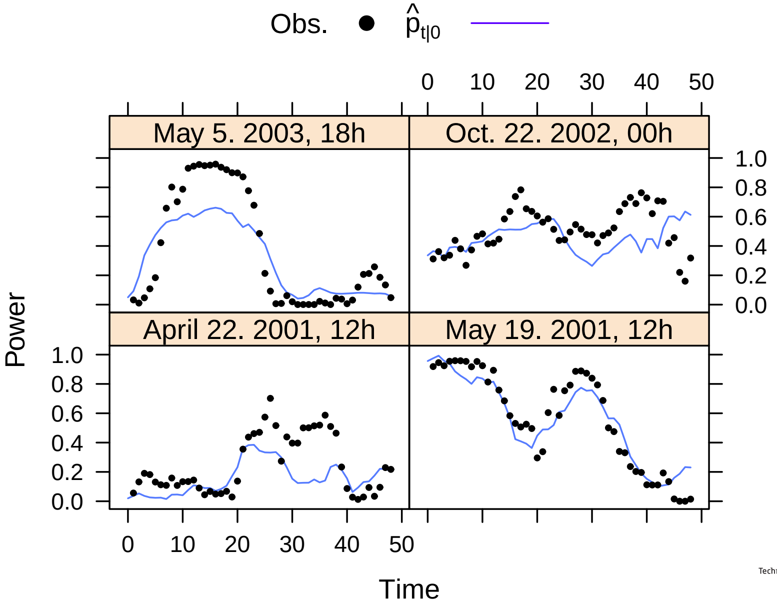
A basic WPPT module is here used to provides point forecast – like

$$\hat{p}_{t+k|t} = \sum_{i=0}^{n_a} a_i p_{t-i} + b\hat{p}^{pc}_{t+k|t}(MET) + f(h_{t+k})$$

where p_t is observed power production, $k \in [1;48)$ prediction horizon, $\hat{p}^{pc}_{t+k|t}(MET)$ is a power curve prediction and h_{t+k} is time of day.

- Parameters are estimated online and adaptively
- WPPT is one af the most widely used forecasting tools for windpower (worldwide)
- WPPT basic point forecasts are used as input to SDE-models

Technical University of Denmark


Obs.
$\hat{p}_{t|0}$
May 5. 2003, 18h
Oct. 22. 2002, 00h
April 22. 2001, 12h
May 19. 2001, 12h
Power
Time
0 10 20 30 40 50
0.0 0.2 0.4 0.6 0.8 1.0

Outline

Technical University of Denmark DTU

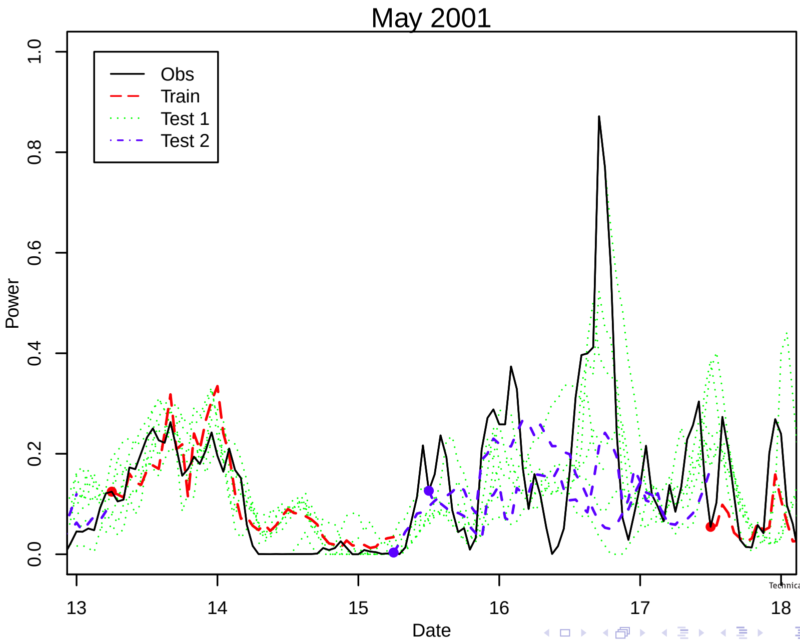
Data

The data set cover the period from January 1. 2001 to May 2. 2003.

- Hourly measurements of actual power production
- 48 hour point-forecast of power production (issued at 00h, 06h, 12h, 18h)
- Only sets where *all point forecasts and all measurements* are non-missing are used in this analysis (2593 complete sets in total)

The dataset is divided into 3 parts

- A training-set: 150 sets of non overlapping data sets
- Test set 1: Sets which overlap with the training-set
- Test set 2: Sets that does not overlap with the training set


May 2001
Power
Date
Obs
Train
Test 1
Test 2
0.0
0.2
0.4
0.6
0.8
1.0
13
14
15
16
17
18

Outline

Main idea

The main idea is to evaluate the full 48 hour predictions (possibly after some transformation)

$$\boldsymbol{Y}_i | \hat{\boldsymbol{p}}_i \sim N(\hat{\boldsymbol{p}}_i, \boldsymbol{\Sigma}_i),$$

where

- $\boldsymbol{Y}_i = [y_{1,i}, ..., y_{48,i}]$ is observed power for sequence i
- $\hat{\boldsymbol{p}}_i = [\hat{p}_{1|0,i}, ..., \hat{p}_{48|0,i}]$ is the WPPT prediction for sequence i
- $\boldsymbol{\Sigma}_i \in \mathbb{R}^{48 \times 48}$ is a covariance matrix for sequence i

The covariance structure is modelled as a function of prediction horizon and predicted power.

Technical University of Denmark DTU

Covariance structure

The covariance is modelled as

$$\mathbf{\Sigma}_i = \mathtt{diag}(\boldsymbol{\sigma_i})\boldsymbol{R}\mathtt{diag}(\boldsymbol{\sigma_i})$$

with $\boldsymbol{\sigma}_i = [\sigma(\hat{p}_{1|0,i}, 1), ..., \sigma(\hat{p}_{48|0,i}, 48)]^T$, and $\boldsymbol{R}$ a correlation matrix.

Models

The following models of increasing complexity is used

BM 0: $\sigma(\hat{p}_{t|0,i}, t) = \sigma_0$, $\quad \boldsymbol{R} = \boldsymbol{I}$

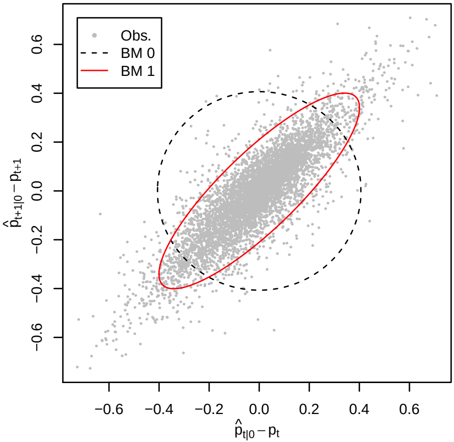
BM 1: $\sigma(\hat{p}_{t|0,i}, t) = \sigma_0$, $\quad \boldsymbol{R}_{lm} = e^{-\lambda|l-m|}$

BM 2: $\sigma(\hat{p}_{t|0,i}, t) = \sigma_0 e^{f(\hat{p}_{t|0})}$, $\quad \boldsymbol{R}_{lm} = e^{-\lambda|l-m|}$

BM 3: $\sigma(\hat{p}_{t|0,i}, t) = \sigma_0 e^{f(\hat{p}_{t|0})+g(t)}$, $\quad \boldsymbol{R}_{lm} = e^{-\lambda|l-m|}$

the functions f, g are approximated by natural cubic B-splines

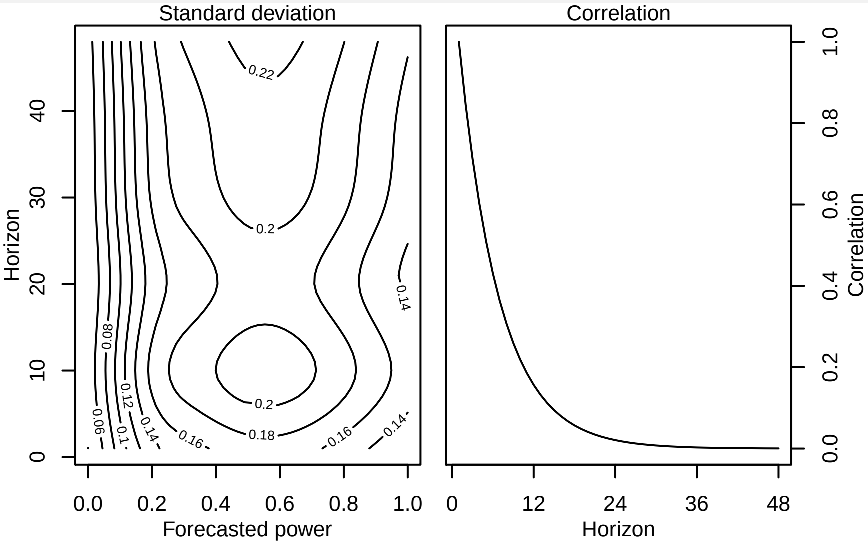
Correlation



Likelihood table (BM models)

	df	l(train)	p-value	l(test1)	l(test2)	l(test)
BM 0	1	2709		35050	27933	62983
BM 1	2	7361	< 0.0001	81312	56303	137615
BM 2	7	8393	< 0.0001	88690	67777	156467
BM 3	12	**8419**	< 0.0001	**88692**	**67853**	**156546**

BM 3 - standard deviation and correlation



Outline

Stochastic differential equations

The SDE model consist of a deterministic drift term and a stochastic diffusion term. These terms determine the dynamics of the system. Furthermore we include an observation equation, which determines what from the system is observed. In general the setup looks as follows:

$$
\begin{aligned}
dx_t &= f(x_t, \boldsymbol{u}_t, \boldsymbol{\theta})dt + \sigma(x_t, \boldsymbol{u}_t, \boldsymbol{\theta})dw_t \\
y_k &= h(x_{t_k}, \boldsymbol{u}_{t_k}, \boldsymbol{\theta}, e_k).
\end{aligned}
$$

Scope

The scope of the SDE-modelling is to generate covariance structures based on SDE formulations.
Given the SDE-formulation

$$dx_t = f(x_t, \boldsymbol{u}_t, \boldsymbol{\theta})dt + \sigma(\boldsymbol{u}_t, \boldsymbol{\theta})dw_t$$

the expected value ($\hat{x}_t$), and variance (P_t) of x_t (for given initial value x_0) is given (approximated) by

$$\frac{d\hat{x}_t}{dt} = f(\hat{x}_t, \boldsymbol{u}_t, \boldsymbol{\theta}); \qquad \frac{dP_t}{dt} = 2A_t + \sigma(\boldsymbol{u}_t, \boldsymbol{\theta})^2 \tag{1}$$

with

$$A_t = \left.\frac{\partial f}{\partial x}\right|_{x=\hat{x}_t}$$

A SDE-formulation for error propagation

The starting point (the Pearson diffusion)

$$dx_{t,i} = -\,\theta \cdot (x_{t,i} - \hat{p}_{t|0,i})dt + \sqrt{2\theta a_{t,i} x_{t,i} \cdot (1 - x_{t,i})}dw_{t,i},$$

with $\hat{p}_{t|0,i} \in (0,1)$, $a_{t,i} < min(\hat{p}_{t|0,i}, 1 - \hat{p}_{t|0,i})$.
The SDE-formulation is reformulated (restricted) to

$$dx_{t,i} = -\,\theta \cdot (x_{t,i} - \hat{p}_{t|0,i})dt + \sqrt{2\theta \tilde{a} \hat{p}_{t|0,i}(1 - \hat{p}_{t|0,i}) x_{t,i} \cdot (1 - x_{t,i})}dw_{t,i},$$

with $\tilde{a} \in (0,1)$ and constant, and state dependent diffusion.

Lamperti transform, state independent diffusion

Consider the transformation

$$z_t = \psi(x_t) = \int \frac{1}{\sqrt{\xi(1-\xi)}} d\xi \bigg|_{\xi = x_t} = \arcsin(2x_t - 1)$$

then the SDE for z_t is given by (Itô's lemma)

$$\begin{aligned} dz_t = & - \frac{\theta \left(x_t - \hat{p}_{t|0} + \frac{1}{2}(1-2x_t) \cdot \tilde{a} \cdot \hat{p}_{t|0}(1-\hat{p}_{t|0}) \right)}{\sqrt{x_t(1-x_t)}} dt \\ & + \sqrt{2 \cdot \theta \tilde{a} \hat{p}_{t|0}(1-\hat{p}_{t|0})} dw_t \end{aligned}$$

with

$$x_t = \frac{1}{2}(1 + \sin(z_t))$$

Results

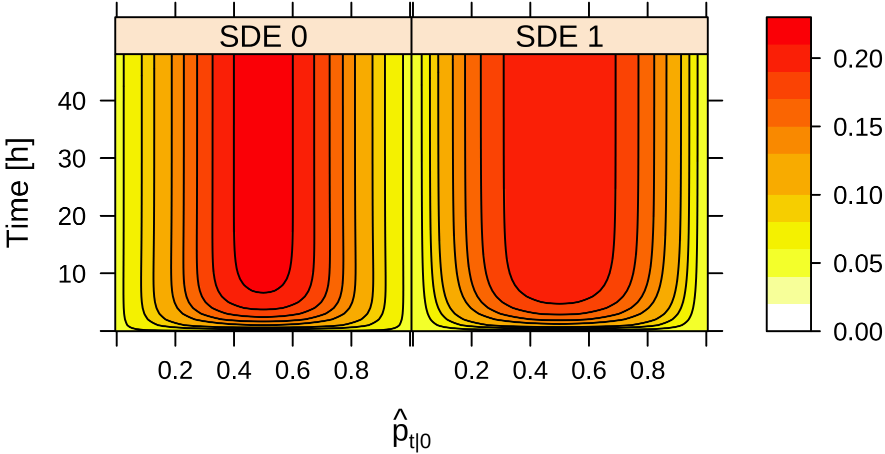
	df	l(train)	p-value	l(test1)	l(test2)	l(test)
BM 2	7	8393		88690	67777	156467
BM 3	12	**8419**	< 0.0001	**88692**	**67853**	**156546**
SDE 0	3	8408.4		92376	69222	161598
SDE 1	4	**8532.4**	< 0.0001	**94292**	**70719**	**165011**

$$SDE0: \quad dx_{t,i} = -\theta \cdot (x_{t,i} - \hat{p}_{t|0,i})dt + 2\sqrt{\theta\alpha\hat{p}_{t|0,i}(1-\hat{p}_{t|0,i})x_{t,i} \cdot (1-x_{t,i})}dw_{t,i}$$

$$SDE1: \quad dx_{t,i} = -\theta \cdot (x_{t,i} - \hat{p}_{t|0,i} - c\hat{p}_{t|0,i}(1-\hat{p}_{t|0,i})(1-2x_{t,i}))dt + 2\sqrt{\theta\alpha\hat{p}_{t|0,i}(1-\hat{p}_{t|0,i})x_{t,i} \cdot (1-x_{t,i})}dw_{t,i}$$

Technical University of Denmark

Standard deviation



Outline

Models

Model of increasing complexity are analysed, s is (an estimated constant) in all models. Non-linear relations are explored by

$$\alpha(\hat{p}_{t|0}, t) = \frac{1}{1 + exp(-\alpha_0 - f_\alpha(\hat{p}_{t|0}) - g_\alpha(t))}$$
$$\theta(\hat{p}_{t|0}, t) = \frac{K_\theta}{1 + exp(-\theta_0 - f_\theta(\hat{p}_{t|0}) - g_\theta(t))}$$
$$c(\hat{p}_{t|0}, t) = \frac{K_c}{1 + exp(-c_0 - f_c(\hat{p}_{t|0}) - g_c(t))}$$

the non-linear functions in f and g are modelled by natural cubic splines. K_c and K_θ are used to control the range of c and θ, we use $(K_c = K_\theta = 20)$.

Technical University of Denmark DTU

Parameter Estimates

	SDE0	SDE1	SDE2	SDE3	SDE4	SDE5	SDE6	SDE7
s	0.058	0.048	0.045	0.057	0.048	0.041	0.037	0.037
α	0.307	0.472	0.568	0.433				
θ	0.206	0.129	0.109					
c	0	2.879						
f_c		0	0	0	0	0	0	NS
g_c		0	NS	NS	NS	NS	NS	NS
f_θ	0	0	0	NS	NS	NS	NS	NS
g_θ	0	0	0	0	0	NS	NS	NS
f_α	0	0	0	0	NS	NS	NS	NS
g_α	0	0	0	0	0	0	NS	NS

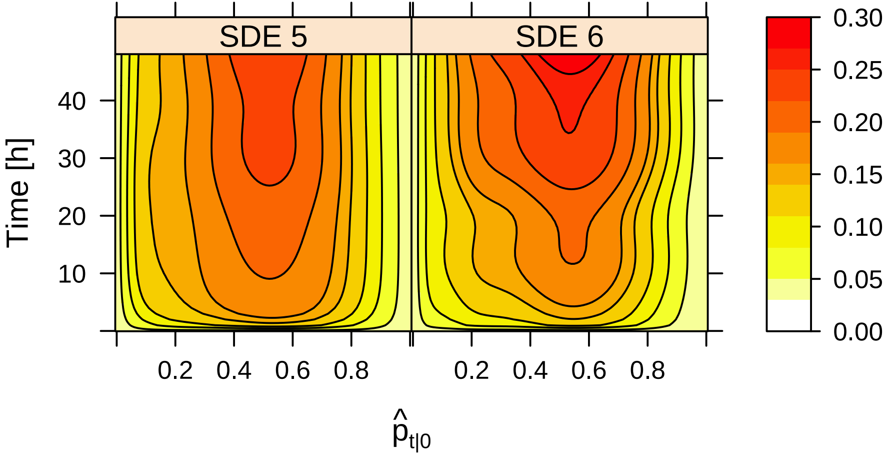
Technical University of Denmark

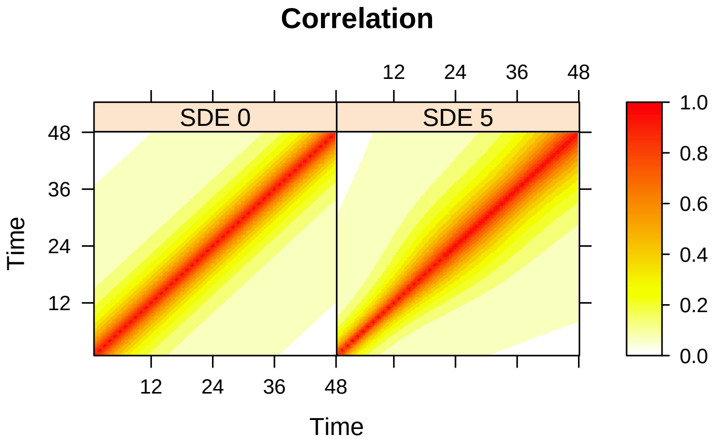
Likelihood

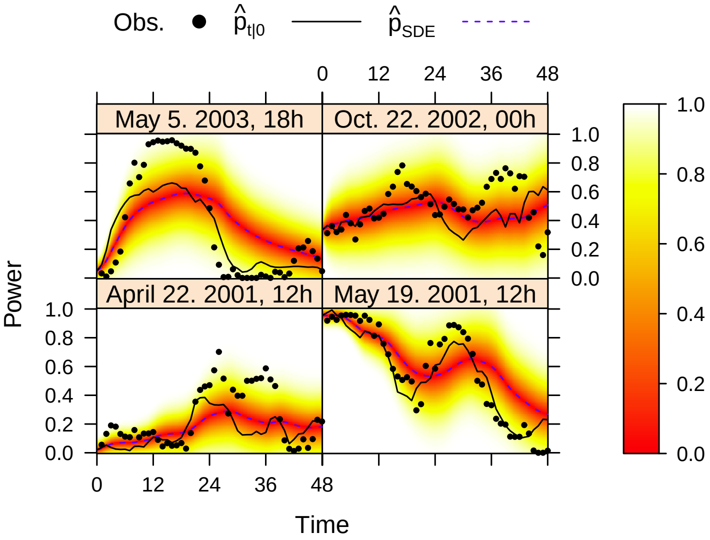
Table: Value of the log-likelihood function for the training set and the test sets.

	df	l(train)	p-value	l(test1)	l(test2)	l(test)
SDE 0	3	8408.4		92376	69222	161598
SDE 1	4	8532.4	< 0.0001	94292	70719	165011
SDE 2	9	8645.3	< 0.0001	94396	70930	165326
SDE 3	14	8673.9	< 0.0001	94391	70982	165373
SDE 4	19	8703.3	< 0.0001	94771	71227	165998
SDE 5	24	8728.9	< 0.0001	**94857**	71439	**166296**
SDE 6	29	8760.5	< 0.0001	94781	**71480**	166262
SDE 7	34	**8770.6**	0.0012	94715	71409	166124

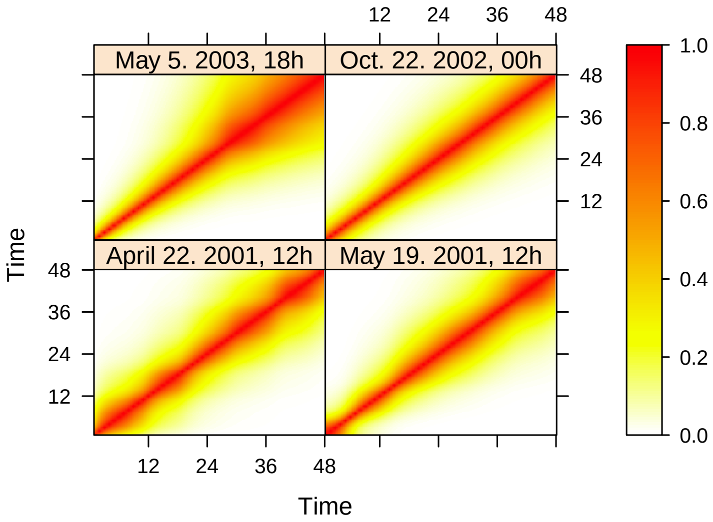
Standard deviation




Correlation
SDE 0
SDE 5
Time
Time
12
24
36
48
1.0
0.8
0.6
0.4
0.2
0.0


Obs.
$\hat{p}_{t|0}$
$\hat{p}_{SDE}$
May 5. 2003, 18h
Oct. 22. 2002, 00h
April 22. 2001, 12h
May 19. 2001, 12h
Power
Time
0
12
24
36
48
0.0
0.2
0.4
0.6
0.8
1.0

Predicted Correlation



Outline

A SDE model for wind speed variability.

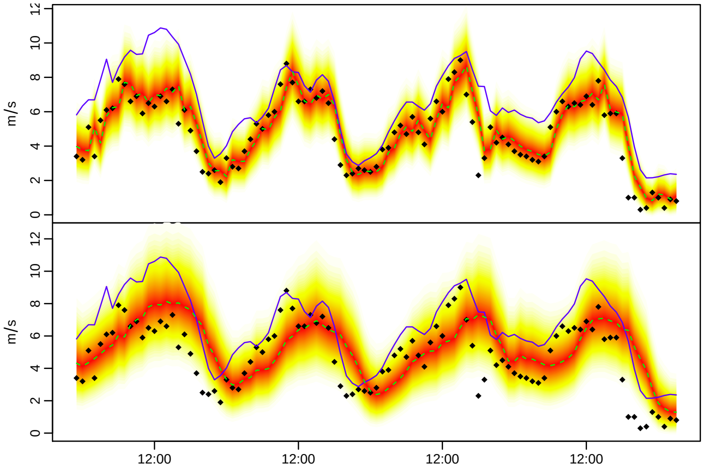
We start out with a simple SDE model. Instead of predicted power output, the numerical weather prediction (NWP) is used:

$$dx_t = -\theta \cdot (x_t - \mathrm{NWP}_t)dt + \sigma \cdot x_t^{\gamma} \cdot dw_t.$$

As there is not physical upper bound for the wind speed as opposed to wind power, we choose the diffusion term $g(x,t) = \sigma \cdot x^{\gamma}$.

We see, however, as for the wind power a lagging behind of the NWP in the SDE model. This is especially clear for longer horizons where we see shifted peaks between the NWP and the point prediction of the SDE model, as the following slide illustrates:

Predictive Density, 1h & 5h



Introducing the derivative of the prediction

We now introduce the derivative of the NWP with respect to time, $\dot{\text{NWP}}$, to the model:

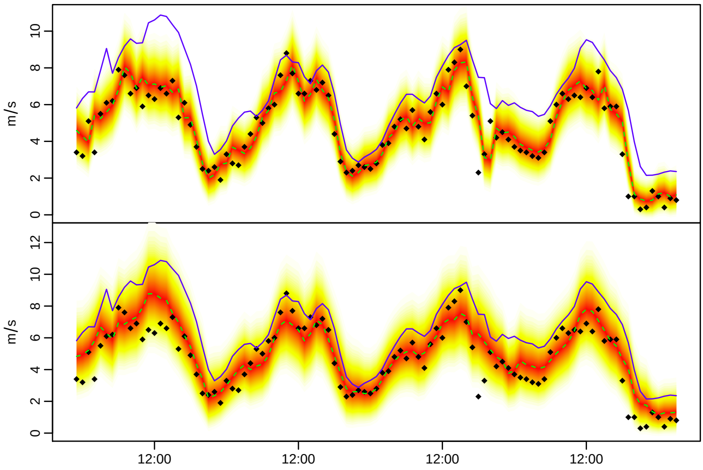
$$dx_t = (-\theta_1 \cdot (x_t - \text{NWP}_t) + \theta_2 \cdot (1 - e^{-x_t}) \cdot \dot{\text{NWP}})dt + \sigma \cdot x_t^{\gamma} \cdot dw_t.$$

Here we have introduced the term $(1 - e^{-x_t})$ for the model to remain feasible, as it makes sure that the process remains in $\mathbb{R}^+$, as the influence of $\dot{\text{NWP}}$ drops to zero when x_t approaches zero.

The prediction and predictive densities of this model is illustrated on the following slide:

Technical University of Denmark DTU

Predictive Density, 1h & 5h



Introducing the derivative of the prediction

Some of the benefit from introducing $\dot{\mathrm{NWP}}$:

- The derivative of NWP w.r.t. time, $\dot{\mathrm{NWP}}$, is implemented and it can be observed that the SDE predictions no longer lag behind the numerical weather predictions.
- The predictive density becomes narrower, as to more accurately model the dynamics.
- The point forecast is also notably improved over the one provided by the numerical weather prediction.
- The long term behaviour of the model and multi-step predictions are accurate.

Technical University of Denmark DTU

Outline

Summary and conclusion

- We model wind power production as multivariate Gaussian (possibly after some transformation), and focus on the covariance structure.
- One sequence of 48hour power observations are considered as one multivariate random variable.
- SDEs are used to generate very flexible covariance structures.
- The SDE formulation allows us to control variance and correlation.
- The SDE models perform well in terms of log-likelihood, when compared to the bench mark models.
- The SDEs model quite complex covariance structures.
- The SDE model provides accurate point predictions as well as predictive densities when the time derivative of the attractor is included.