A Unified Method for Probabilistic Forecast of Wind Power Generation

Jan Kloppenborg Møller
Emil Banning Iversen
Henrik Madsen

DTU-Compute - Technical University of Denmark
jkmo@dtu.dk

Outline

Outline

Motivation

An increasing part of electricity supply is generated by wind

- Wind power cover about 29% of total system load
- 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 timescales

- 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 in use

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

Most methods are implemented in Wind Power Prediction Tool (WPPT).

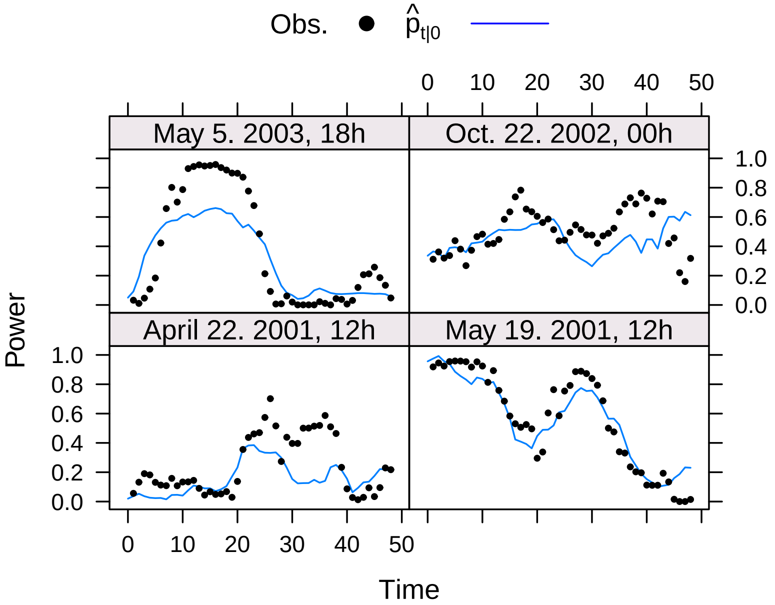
WPPT point-forecast

WPPT provide a point forecast

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

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



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

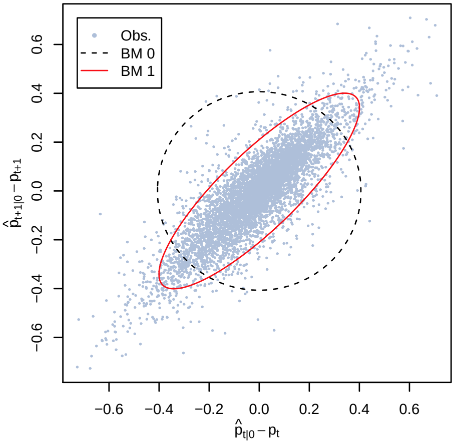
Data

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

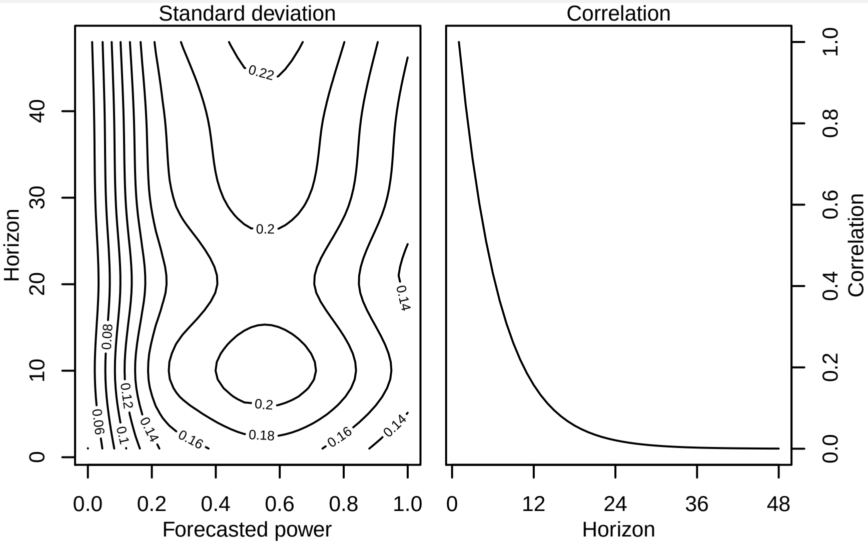
- Hourly measurements of actual power production
- 48 hour point-forecast of power production (issued at 00h, 06h, 12h, 18h)
- Only sets where *all point forecasts and all measurements* are non-missing are used in this analysis (2593 complete sets in total)
- 150 sets are used to train the SDE models

Outline

Correlation



BM 3 - standard deviation and correlation



Outline

Scope

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

Given the (continuous time) SDE-formulation

$$dx_t = f(x_t, u_t, \boldsymbol{\theta})dt + \sigma(x_t, u_t, \boldsymbol{\theta})dw_t; \quad x_0 \text{ given}$$

and the (discrete time) observation equation

$$\boldsymbol{p} = \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\theta}) + \boldsymbol{e}; \qquad \boldsymbol{e} \sim N(\boldsymbol{0}, \boldsymbol{S})$$

where

- $\boldsymbol{p} \in [0, 1]^{48}$ is observed power
- $\boldsymbol{x} = \{x_1, ..., x_{48} | x_0\} \in [0, 1]^{48}$, is the state vector
- u_t is some input (here predicted power)

A SDE-formulation for error propagation

The starting point (the Pearson/logistic diffusion)

$$dx_t = -\,\theta \cdot (x_t - \mu)dt + \sqrt{2\theta a x_t \cdot (1 - x_t)}dw_t,$$

with $\mu \in (0, 1)$, $a < min(\mu, 1 - \mu)$.

- $x_t \in (0, 1)$
- Diffusion (variance) small when x_t is close to 0 or 1.
- Long term average equal μ
- Stationary distribution is a beta distribution (with parameters $\frac{\mu}{a}$ and $\frac{1-\mu}{a}$)
- Interdependence structure controlled by θ

Basic models

The second order moment representation can be solved for $a \in (0, 2)$, and we choose

$$dx_t = -\theta \cdot (x_t - \mu - \gamma(1 - 2x_t))dt + 2\sqrt{\theta a x_t \cdot (1 - x_t)}dw_t,$$

with the observation equation given by

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{e}; \quad \boldsymbol{e} \sim N(\boldsymbol{0}, \boldsymbol{S}),$$

where

- $\mu = \hat{p}_{t|0}$ (mean)
- $a = \alpha \hat{p}_{t|0,i}(1 - \hat{p}_{t|0,i})$ with $\alpha \in (0, 1)$ (variance)
- $\gamma = c\hat{p}_{t|0,i}(1 - \hat{p}_{t|0,i})$ with $c \in \mathbb{R}_+$ (bias)

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**

$$SDE0 : \text{No bias}$$

$$SDE1 : \text{Including bias}$$

Outline

Models

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

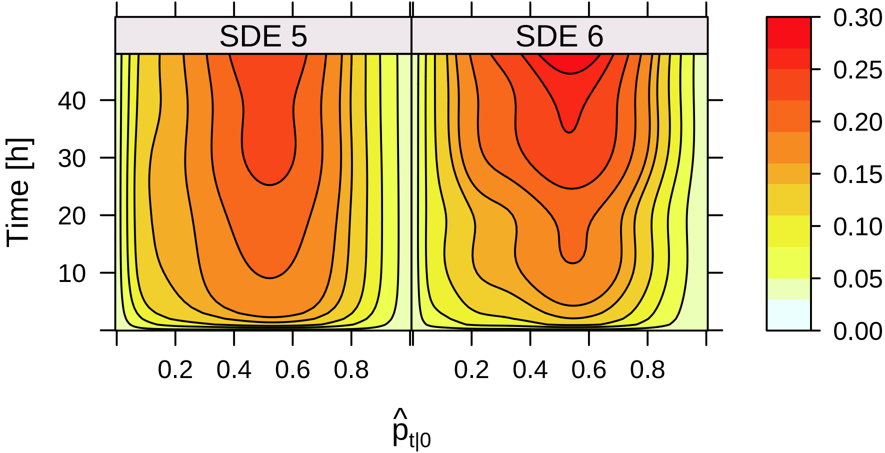
$$\alpha(\hat{p}_{t|0}, t) = \frac{1}{1 + exp(-\alpha_0 - f_\alpha(\hat{p}_{t|0}) - g_\alpha(t))}$$

$$\theta(\hat{p}_{t|0}, t) = \frac{K_\theta}{1 + exp(-\theta_0 - f_\theta(\hat{p}_{t|0}) - g_\theta(t))}$$

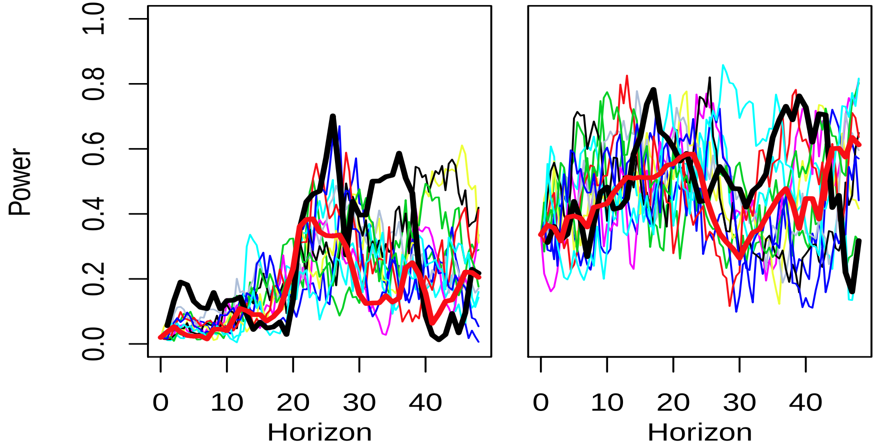
$$c(\hat{p}_{t|0}, t) = \frac{K_c}{1 + exp(-c_0 - f_c(\hat{p}_{t|0}) - g_c(t))}$$

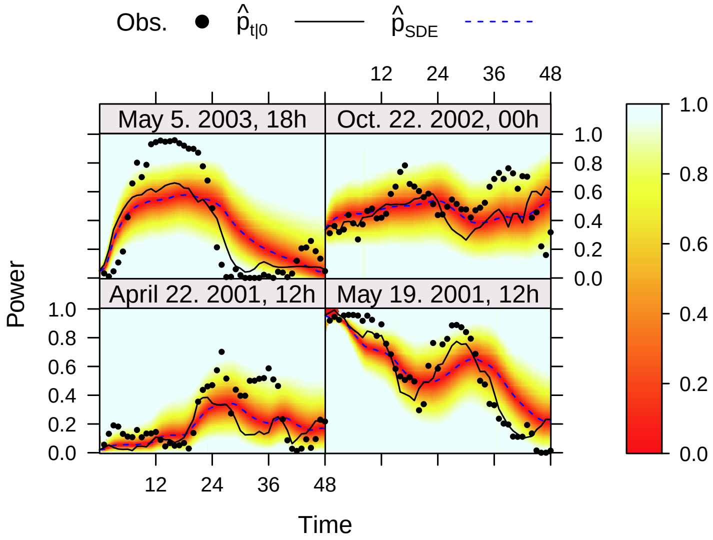
the non-linear functions in f and g are modelled by natural cubic splines. K_θ and K_c are used to control the range of c and θ, we use $(K_\theta = K_c = 20)$.

Standard deviation

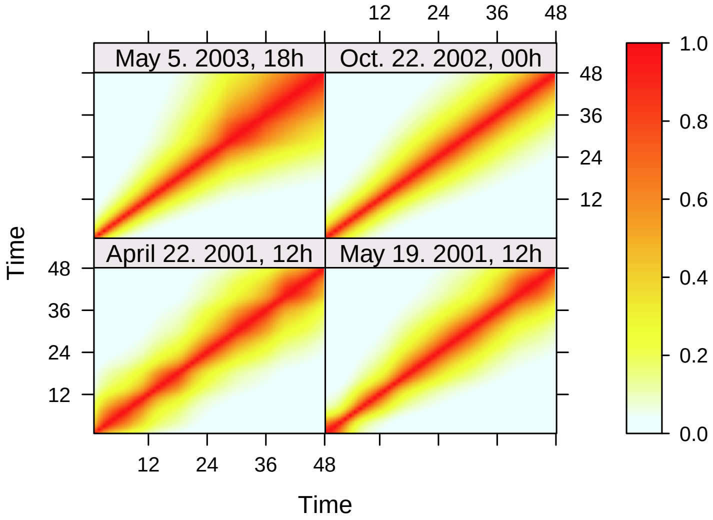


Ensemble forecasts





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
12
24
36
48
0.0
0.2
0.4
0.6
0.8
1.0


May 5. 2003, 18h
Oct. 22. 2002, 00h
April 22. 2001, 12h
May 19. 2001, 12h
12
24
36
48
Time
1.0
0.8
0.6
0.4
0.2
0.0

Outline

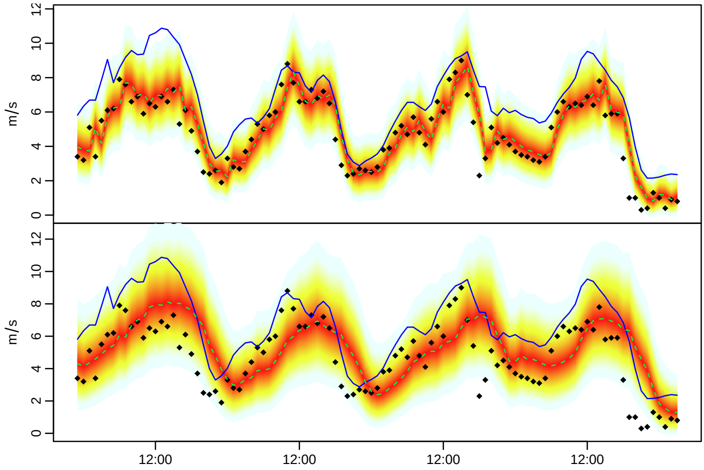
A SDE model for wind speed variability.

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 $g(x,t) = \sigma \cdot x^{\gamma}$.

Predictive Density, 1h & 5h



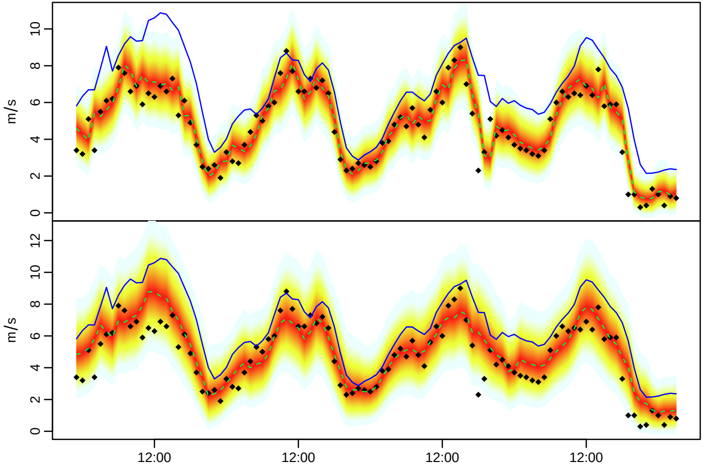
Introducing the derivative of the prediction

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

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

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 $\dot{\mathrm{NWP}}$ drops to zero when x_t approaches zero.

Predictive Density, 1h & 5h



Outline

Summary and conclusion

- We model wind power production as multivariate Gaussian (possibly after some transformation), and focus on the covariance structure
- SDEs are used to generate very flexible covariance structures
- The SDEs model quite complex covariance structures
- Ensemble forecast and prediction densities (and intervals) are easily obtained from SDE's
- Introducing the derivative of NWP w.r.t. time, gave large improvements in terms of predictive densities and time lags.