02427 Advanced Time Series Analysis

Computer exercise 4a

For the fourth computer exercise you may choose among several different challenges. This document is a description of the exercise concerning forecasting of wind power based on weather forecasts.

Forecasting of wind power

The power production from wind turbiness is very dependent on the wind. So weather forecasts for the wind direction and speed is of great interest when trying to control the power production. In several countries the production of wind power accounts for an increasing part of the total power production and in order to control the conventional power plants and to trade power on NordPool, the nordic power exchange, a prediction of the power production from wind turbines is needed.

The aim of this exercise is to develop adaptive models for predictions of the power production at different time horisonts given the available weather forecasts.

Description of data

The data in this exercise is from Klim, a wind farm located near Fjerritslev in the Northwest of Jutland. A new weather forecast is made every 6 hours and it consists of estimates of the wind speed, direction, and temperature for every hour in the following 48 hours. Every hour the power production from the wind farm is recorded.

The data in the file `cex4WindDataInterpolated.csv` has been processed and interpolated to make you spend the time on modelling, rather than cleaning and processing data. The file contains both hourly averages of wind power measurements and weather forecasts. The signals are aligned such that the power and the forecast for a given hour is in the same row, i.e. the power p_t and the 1-hour ahead forecast of wind speed $\hat{W}_{t|t-1}$ are in the same row (t is here in hours).

The following signals are in the file:

`t`	The time in UTC. This is the end point of the sample period, i.e. the values in the row of t_i is the average between t_i and t_{i-1}.
`toy`	The time of year in days.
`p`	The measured average wind power.
`Ws1`	The 1-hour ahead forecasted wind speed.
`Wd1`	The 1-hour ahead forecasted wind direction.
`T1`	The 1-hour ahead forecasted temperature.
`Ws2`	The 2-hour ahead forecasted wind speed.
`Wd2`	The 2-hour ahead forecasted wind direction.
`T2`	The 2-hour ahead forecasted temperature.
`Ws3`	The 3-hour ahead forecasted wind speed.
`Wd3`	The 3-hour ahead forecasted wind direction.
`T3`	The 3-hour ahead forecasted temperature.

For the wind power measurements linear interpolation of gaps shorter than or equal to 28 hours is carried out. For the weather forecast missing values are replaced with older forecasts, and afterwards linear interpolation of gaps shorter than or equal to 26 hours is carried out. If you want to experiment with another interpolation scheme, the signals in `cex4WindDataNoInterpolation.csv` are uninterpolated.

The raw data is also available and scripts both for R and Matlab to process and interpolate the data, see the last section.

R hint
To read the data in R use

```
X <- read.table("../cex4WindDataInterpolated.csv", sep=",",
header=TRUE, stringsAsFactors=FALSE)
X$t <- as.POSIXct(X$t, tz="UTC")
```

Exercise

A nonlinear dependence on wind speed and direction is expected. You are requested to develop adaptive models for the prediction of wind power 1, 2, and 3 hours ahead. You have to specify your own model. An example could be a model with diurnal variations, an autoregressive part, and a dependence on changes in wind speed.

Look at the parameter traces and see what you can identify.

Be aware that the roughness of the surroundings has a seasonal variation, this is partly due to leaves on trees.

Hint: Start with a simple model, e.g. pure autoregressive, and see how improvements of the model results in better predictions.

Suggestion: As a beginning it may be an idea to estimate the power curve, the dependence of the power production on the wind speed, for different wind directions.

Raw data

The raw data is also available, it is in the files `wind_nwp.dep` and `wind_pow.dep`, both having a header with information on the content. The time format in both files is `yyyymmddhhmm` and missing data is labeled as 'NaN'. The recorded productions are in the file `wind_pow.deb` (1Mb) which contains three columns:

- TimeMeas - Time of the measurement
- ToyMeas - Time of the year (days since new year)
- PowerMeas - Measured power production (kW)

The weather forecasts are in the file `wind_nwp.deb` (20Mb) which contains six columns:

- TimeNWP - Time of predictions
- ToyNWP - Time of the year
- PTimeNWP - Time when forecast was calculated
- WSpdNWP - Predicted wind speed (m/s)
- WDirNWP - Predicted wind direction (deg)
- TempNWP - Predicted temperature (K)

The wind speed and direction and temperature are in HIRLAM level 31 which corresponds to a height of about 70m. It is important to understand the combination of the first and third columns. As mentioned there is a new forecast every 6 hours and every forecast covers $[0; 48]$h with a resolution of one hour. Hence, every forecast produces 49 lines in the file with the time of the calculation time in the third column and the time of the predictions in the first column.

In Matlab `wind_read.m` can be used to read and process the data, and in R the script `readDataCex4WindPower.R` does the same. Note that paths to the data files have to modified.