02417: Time Series Analysis

Lasse Engbo Christiansen
DTU Compute
Technical University of Denmark

Fall 2016

Assignment 3: Estimating ARMA Processes and Seasonal Processes

Exhaust gasses from combustion engines contain NO_x which is the sum of NO and NO_2 (Sunlight and ozone affects the balance between the two). As part of a national surveillance program the NO_x concentration is measured every hour at Jagtvej in Copenhagen. The sensor is located between the road and the bikelane.
The data file `A3_jagt_NOx.csv` is made using "," as column separator. The file contains three columns: The date, the hour within day where the measurement is taken and the concentration of NO_x in $\mu g(NO_2 equiv.)/m^3$.
You should not use the last 48 hours when estimating your model - as they should be used for testing.
Data originates from: `http://www.data.kk.dk/dataset/luftforurening/resource/98275e9c-22da-4158-b44b-ac4844df8ab4`

Question 3.1: Presenting the data Plot the NO_x concentration. Consider plotting for subsets of the data to show the structure. Comment on the behaviour including considerations on stationarity and transformations.

Question 3.2: ACF and PACF Estimate the autocorrelation function and the partial autocorrelation function of the NO_x concentration and if relevant also for series derived from the concentration, e.g. transformations.

Question 3.3: Model selection Select an initial model structure. Estimate the parameters. Validate the model. Consider tests for lower model order. Consider updating the model structure.

Argue for the choices you make. Remember that the model building process is an iterative process and you should always consider stepping back and reconsider your choices.

Question 3.4: Predictions Use the model you have developed for predicting the NO_x concentration 48 hours ahead and include prediction limits. Compare with the data that was left out. Include a table with the 1h, 24h and 48h predictions.

HINT:
If you want to convert the first two columns of the data to a time stap then the following functions may be useful: `substr`, `strsplit` and `as.POSIXct`.
Some functions behave nicer per default if data are treated as time series objects with the natural seasonality. E.g. using the following `xts <- ts(x, frequency = 24)`.