02417: Time Series Analysis
Lasse Engbo Christiansen & Sebastian Wolf
Spring 2019

DTU Compute
Technical University of Denmark

Assignment 1: Steady state building model

In this assignment we consider a linear steady state model that describes a simplified energy balance for a building.

$$\Phi_h = H_{tot}(T_i - T_e) + gA_{sol}I_{sol} + \varepsilon, \tag{1}$$

where the quantities are

- Φ_h heating power inside the building [W]
- T_i indoor temperature (controlled at a constant setpoint of 25 $^\circ C$)
- T_e outdoor temperature [$\deg C$]
- I_{sol} solar irradiation [Wm^{-2}]
- H_{tot} sum of total transmission and ventilation loss coefficient [W/K]
- gA_{sol} parameter which is the product of: g solar transmittance of the transparent facade elements and their total area A_{sol} [m^2]
- ε noise

Interpretation: If the indoor temperature is kept constant, then the heating power must compensate transmission and ventilation losses and solar gains.

Note that this model is a simplification as it does not consider any temporal dynamics. In reality, changes of temperature and solar radiation will affect the indoor temperature also in future time steps, due to heat storage in the walls.

The data is provided in `house_data_6h.csv` (6 hours averages) and `house_data_3h.csv` (3 hours averages) and includes four columns:

t: Date and time of the observation

Ph: The heating [W]

Te: The outdoor temperature [$degC$]

Isol: The solar radiation [W/m^2]

You should not use the observations for the last 24 hours for estimations - only for comparisons.

Question 1.1: *Read the 3h and the 6h data and plot the given quantities as a function of time. Do indicate which data is used for training and testing. Comment on the evolution of the values over time. It is OK if the plot is combined with results from the following questions.*

Question 1.2: *The indoor temperature is controlled by the spatial heating to a fixed setpoint of 25 °C.*

1. *Formulate a linear regression model to estimate H_{tot}, gA_{sol} and the setpoint of T_i, in which it is assumed that the residuals have a constant variance and are independent (i.e. $\Sigma = I$). Estimate the parameters using the 6h data - do include a measure of uncertainty for each of the estimates. Plot the residuals for this model.*
2. *Now, we assume that the correlation structure of the residuials is an exponential decaying function of the time distance between two observations, i.e.*

$$Cor(\varepsilon_{t_i}, \varepsilon_{t_j}) = \rho^{|t_i - t_j|} \tag{2}$$

for some $0 < \rho < 1$, which results in the following correlation matrix:

$$\Sigma = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots \\ \rho & 1 & \rho & \cdots \\ \rho^2 & \rho & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{3}$$

Estimate ρ using five iterations of the relaxation algorithm. Then estimate H_{tot}, gA_{sol} and the setpoint of T_i. Provide a measure of uncertainty for each of the estimates. Again plot the residuals. Comment on the results.

3. *Switch to the dataset sampled every 3 hours (**house_data_3h.csv**) and repeat what you did with the 6h data (The two steps above). Comment on the effect of the sampling periods.*

Question 1.3: *Use a local linear trend model on the outdoor temperature in the 6h training data using $\lambda = 0.8$.*

1. *Plot the training data and the corresponding one step predictions for all observations in the training data.*
2. *Plot the one step prediction errors.*
3. *Predict the four observations in the test data (Based solely on the training data) Make a plot and a table with the predictions - including a 95% prediction interval. Compare with the test data.*
4. *Comment on the results.*

Question 1.4: Predict observations for the last day (test data) of the spatial heating Ph based on the 6h data. For this, use the estimated model in Q1.2 and the outdoor temperature from Q1.3. You may use future values of I_{sol}. Present the predictions in a plot and in a table.

Comment on the results.

Hints: When providing a measure of uncertainty then you can choose variance, standard deviation, and confidence/prediction intervals. For the latter do use a 95% level if nothing else is given. Often it is more informative to provide an interval as it includes information on the number of observations used in the estimation.

Comment on the results.