Summer School; Transfer function modelling

Exercise 3: **Application of transfer function models**
Analysis of the IEA Annex 58 testbox using ARX models

Exercise 3: Application of Transfer Function Models. Analysis of the IEA Annex 58 testbox using ARX models

June 21, 2016

Introduction

In this exercise transfer function models are applied for modelling the heat input to the IEA Annex 58 test box. The aim is to achieve a reliable estimate of the Heat Loss Coefficient (HLC) and gA-value of the box. In the preparation exercise linear regression models were applied for modelling the two-daily average heat load in a building, when the time resolution is increased it becomes necessary to include the dynamics in the model. This is carried out using ARX models (linear Auto-Regressive models with eXogenous inputs), which are called black-box models, since the parameters cannot be directly interpreted in a physical sense. However the steady-state parameters can be derived from the estimated parameters.

First, an introduction to the auto-correlation function (ACF) and the properties of white noise is given. Secondly, an AR model is introduced and finally, an ARX model is to be fitted to the test box data.

For each question accompanying R-scripts are given.

1 Question: White noise and ACF

Open the script `Q1WhitenoiseAndACF.R`. The script presents some basic properties of white noise and the ACF.

1.1 White noise

Generate a series of random i.i.d. (independent identically distributed) normal distributed numbers: white noise. Step through the script and see how the ACF is calculated and how the ACF for white noise is.

In the plot of the ACF the blue lines indicate a 95% confidence interval for the auto-correlation being significantly different from zero. In average which fraction of the bars will be out of the confidence band for white noise?

1.2 ACF of an AR(1) process

Generate a realization of an AR process

$$\phi(\mathrm{B})Y_t = \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2) \text{ and i.i.d.} \tag{1}$$

where B is the backward shift operator

$$\mathrm{B}^k Y_t = Y_{t-k} \tag{2}$$

and

$$\phi(B) = 1 + \phi_1 B + \phi_2 B^2 + \dots \tag{3}$$

An AR(1) process is

$$(1 + \phi_1 B) Y_t = \varepsilon_t \tag{4}$$

which can be rewritten into a linear regression model

$$Y_t = -\phi_1 Y_{t-1} + \varepsilon_t \tag{5}$$

Simulate with different values of ϕ_1 and see how this change the series and the ACF.

1.3 More ACF

See the ACF of a sine series and that the ACF is relative measure of linear dependence.

Then add some spikes to an AR(1) series. Try to increase the level of the "spikes", what happens to the characteristics of the ACF when the level is increased?

1.4 ACF game

Optional: Finally, a little ACF game for having times series fun! Enter a series of numbers (1 to 9) with the keyboard and plot the ACF. Try to make ACF with different patterns:

- Alternating (negative,positive,negative,...)
- Un-correlated (like white noise)
- Slowly decaying (like AR process)

2 AR as linear regression model

In the part an AR process will be simulated and the parameters will be estimated by formulating AR models as linear regression models.

2.1 AR(1) with lm

A confidence band for the parameter in the linear regression model is here

$$\hat{\omega}_1 \pm 1.96 \cdot \hat{\sigma}_{\omega_1} \tag{6}$$

where $\hat{\sigma}_{\omega_1}$ is found under `Std. Error` in the print out by `summary()` in R.

Is the true parameter value included in the 95% confidence interval?

Asymptotically in 5% of the times the true parameter value in not included in the 95% confidence band, optionally you can try to verify this.

2.2 AR(2) with lm

Simulate an AR(2) process, but fit only an AR(1) to the series. Are there any correlation left (is all information modelled) in the series? If not the model is said to be under-fitted.

What happens when an AR(3) process is fitted to the series? If the model is too complex (i.e. too many parameters) it is said to be over-fitted.

3 Modelling thermal performance of a test box with ARX models

In the part the thermal performance of the IEA Annex 58 test box is modelled with ARX models. Data from the experiment where the indoor temperature is held constant at 25 °C is used. The following series a used:

- Q_t (`Qi` in data) the heat input (W)
- T_t^{i} (`Ti` in data) the indoor air temperature (°C). The average value of the two measured indoor air temperatures.
- T_t^{e} (`Te` in data) the external temperature (°C). The average value of the two measured external air temperatures.
- G_t^{v} (`Gv` in data) the global radiation (W/m^2).

The data and the box is described in the `testboxDescription_ST_CE3b.pdf` document.

Init

Run the initialization and see plots of the series

3.1 CCF

Get familiar with the cross-correlation function (CCF). Are there are any dependencies between the series? State roughly what the patterns indicate (i.e. slow or fast dynamics)?

3.2 LM and CCF

Fit a linear model and analyze the residuals. First see if the parameters are significant?

Analyze the residuals with the ACF and CCF. Are there auto-correlation and dependencies left?

What can be seen from the time series plots in terms of the performance of the model? i.e. are any of the inputs causing troubles.

3.3 Fit an ARX model

Carry out a forward model selection. Add one lagged input at a time and in each step consider if the added parameters improves the performance of the model. In the first step extend with `Qi.l1`, then analyze the significance of the parameters, and the ACF, CCF and time series plots. In each step based on the analysis of the residuals consider which part of the model should be extended and carry on until no significant parameters can be added and the residuals are close to white noise. See Chapter 6 in the TS book for more details on model selection.

3.4 HLC, gA and step response

Calculate the step response of the identified transfer function. This is implemented in the function `stepResponse`.

Finally, calculate the estimate of the Heat Loss Coefficient HLC-value (UA-value). See that this is the steady-state response of the transfer functions. The calculations are tricky, since the two estimates of the HLC-value (one for T_i and one for T_e) is weighted to find an estimate based on those with minimum variance. The weighting is implemented in the function `HLC.Qi.ARX`, see it in `functions/HLC.Qi.ARX.R`.

4 ARMAX model (optional)

Finally, the R package marima (see article `Spliid1983.pdf`) can be used to estimate ARMAX models (even MARIMAX models (Multiple output Auto-Regressive Integrated Moving Average with eXogenous input)).

You can try to estimate an ARMAX model using the script `r/Q4ARMAXBox.R` and see the function `r/functions/estimateARMAX.R`. The method is very powerful, however the defining a the model is a little bit cryptic.