02907 Dynamic Calculation Methods For Building Energy Assessment: Introduction to time series analysis

August 8, 2014

Contents

Reading instructions

The intention of the present text is to give a quick introduction, sort of a tutorial, to modeling of energy systems, in particular related to buildings, using time series analysis techniques. The aim is to enable engineers, physicists and others with a light (maybe by now forgotten) statistical background to familiarize themselves with the principles and tools needed for building models of dynamical systems (system identification) with suitable statistical techniques.

The text is arranged with examples and exercises given in the R language, with also a short introduction to R . The reader should try the examples and exercises in parallel with reading the text, such that when a new concept is presented in the text, the reader runs the R code and plays around with it to get a hands on experience. In this way hopefully the material becomes more accessible and settles better in the mind. The R code presented in the text are given in a file for each chapter, see the .zip file available with the text.

This content should all be read before the summer school, how detailed depends on the readers prior knowledge. At least go rather thorough through: introduction to R in Section 2, the most basic elements presented in Section 3 and the estimation techniques in Section 4. Finally, as preparation the exercise in Section 7.1 should be completed and a report handed in before the summer school.

Chapters

The text is divided into chapters with the following content:

- **Introduction**: An introduction to the modeling techniques and applications
- **Introduction to R** : installation tips and a brief introduction to the statistical software R
- **Probability, statistics and time series analysis**: the basic elements from statistical and probability theory needed for time series analysis
- **Estimation**: the two applied statistical estimation techniques: least squares and maximum likelihood

- **Discrete time models**: the static linear model (linear regression model) and the discrete time dynamic models
- **Continuous time models**: grey-box models based on stochastic differential equations
- **Exercises**: exercises to be carried before and during the summer school

Chapter 1

Introduction

Historically testing of eg. building components have been conducted in a steady state, where all the states (temperatures, etc.) are constant. However, such methods call for rather long testing, and they don't provide any information about the dynamical characteristics or behaviour. Dynamics testing followed by appropriate statistical methods for handling the data, provides much more accurate and detailed information about thermal performance of the building and building components.

Real life data of energy consumptions in buildings, like those obtained by using smart meters, is reflecting a dynamic situation, since the ambient air temperature, solar radiation, etc. varies, and hence such real life time series can provide rather accurate and very important information about the thermal characteristics of buildings.

The prerequisite for obtaining reliable, accurate and detailed information based on dynamic testing or time series data is that proper statistical methods are used. For most classical statistical methods used for identifying the thermal performance parameters (like the UA-value) the correlation of data in time is often disregarded. For instance in regression analysis the assumption about serial uncorrelated residuals is often violated in practice. In this text it will be demonstrated that it is crucial to take this autocorrelation into account in the modeling procedure. Also for applications such as simulations and forecasting, we will most often be able to provide much more reasonable and realistic results by taking the autocorrelation into account.

On the other hand, adequate methods and models for time series analysis can often be seen as a simple extension of linear regression analysis where previous observations of the dependent variable are included as explanatory variables in a simple linear regression type of model. This facilitates a rather easy approach for understanding many methods for time series analysis. There are a number of reasons for studying time series. These include a characterization of time series (or signals), understanding and modeling the thermal performance of a building (or a component), forecasting of future values, and optimal control of eg. the heat supply to a building.

The present text is focused on giving a ”soft introduction” to basic time series analysis concepts and techniques, which are needed for modeling of both linear and non-linear systems, and references are given to papers and books containing much more details on statistics, physics and modeling.

1.1 Selection of model and method

This text describes a number of statistical methods and models for describing the thermal characteristics of buildings using frequent readings of heat consumption, ambient air temperature, and other available climate variables. For some of the methods frequent readings of the indoor air temperature are needed or beneficial.

The ultimate list of different types of models is very long, it counts: nonlinear stochastic models, linear stochastic models, transfer function models, state space models, grey box models, frequency response function models, impulse response models and regression models, and even continues. The final choice of model depends on the purpose of the modeling, existence of prior physical knowledge, and the data. The importance of statistical model validation is discussed, and tools for that purpose are demonstrated.

The basic models are introduced, this is a natural basis for further studies of more advanced models. First the static linear models (simple linear regression models), then the discrete time dynamic model and finally the continuous time models (called grey-box models).

1.2 State space and grey box models

Some of the approaches for considering dynamical models of a system, which has also been used within the PASSYS project, can be classified as *grey box modeling*. Here the model is formulated in physical terms, partly from the physical laws determining the dynamics of the system. The final model structure is established using statistical methods, and the parameters of the embedded model is found using estimation techniques. The model is formulated as **state space model** in continuous time, i.e. the dynamics is described a set of stochastic differential equations (the **system equation**) and the observations are described by the **observation equation** which is a set of discrete time static functions describing how the observed dependent or output variables dependes on the states and the input variables.

This approach bridges the gap between physical and statistical modeling, and hence the final model is most often called a **grey box model**. This approach is attractive from a modeling viewpoint since the parameters of such a model may be directly translated to physical characteristics. Such models are most conveniently described on **state space form**.

It is useful to classify the state space model as being stochastic or deterministic, and nonlinar or linear. The stochastic model includes terms to describe the correlated noise of a system. The following table contains the considered classification of state space models.

Table 1.1: Classification of state space models

	stochastic	deterministic
nonlinear	$dX = f(X,U)dt + g(X)dw$ $Y = h(X,U) + e$	$dX = f(X,U)dt$ $U = h(X,U) + \epsilon$
linear	$dX = AXdt + BUdt + dw$ $Y = CX + DU + e$	$dX = AXdt + BUdt$ $Y = CX + DU + \epsilon$

In this table X is the state vector of the dynamical system (in our case the state vector mostly often consists of temperatures of different elements of the building), U is the input to the system, controlled or not controlled. For the stochastic models the term w is a continuous time noise process with Gaussian increments; the so-called Wiener process. Y is the discrete time measurements or output from the system, and e represents the measurement noise. The characteristics of the functions and matrices are dependent on some parameters θ. Links between the models should be noticed; like the fact that the linear models are subsets of the nonlinear models, and the deterministic models are subsets of the stochastic models.

Another approach for modeling a dynamical system, is to apply a of *black-box* description of the system. In that case we want a good model of the system, but we care less much about the physical interpretation of the parameters of the model. The model is most often described in discrete time, since measurements are given in discrete time. The discrete time model corresponding to the linear stochastic model of Table 1 is

$$X_{t+1} = X_t + \Gamma U_t + v_t \tag{1-1}$$
$$Y_t = CX_t + DU_t + e_t \tag{1-2}$$

This state space model can be rewritten in the form

$$\phi(B)Y_t = \omega(B)U_t + \theta(B)\epsilon_t \tag{1-3}$$

where B is the backshift operator. This model type are called ARMAX models. From ARMAX models we are able to get dynamical information about a system in terms of time constants and also the overall steady state characteristics, but some of the more detailed information of the building is hidden.

For the considered classes of models, there are differences in complexity. This is also reflected in the available software for estimating parameters of the different models. Whereas a number of different software packages are able to handle ARMAX models, this is more limited for the continuous time models. The linear and nonlinear

stochastic models in Table 1 can all be handled by the CTSM-R which can be downloaded from www.ctsm.info, which a R package for grey-box modeling of physical systems ([Kristensen and Madsen, 2003] and [Kristensen et al., 2004]). Another possibility is the MoCaVa software, see [Bohlin and Isaksson, 2003], which is a part of Matlab. The MoCaVa software is a grey-box modeling tool, which has a lot of built in tools for model validation. MoCaVa has a somewhat limited description of the system noise, and a comparison of MoCaVa and CTSM is found in [Kristensen et al., 2004].

1.3 History of dynamic testing

Many of the methods and models are developed and tested during a number of European Research projects focusing on outdoor testing under real weather conditions; the first being the PASSYS project ([Cools and Gicquel, 1989] and [van Dijk and Téllez, 1995]). Some of the approaches have been proposed, described and used in [Bloem, 1994], [Madsen and Holst, 1995] , [Andersen et al., 2000], [Bloem, 2007], [Jiménez and Madsen, 2008], [Jiménez et al., 2008a], and [Jiménez et al., 2008b].

1.4 Stochastic versus ordinary differential equations

Let us illustrate the difference between ODEs and SDEs, which also reflects the difference between output error methods and prediction based estimation methods.
Consider a state space model where the dynamics are described by an ordinary differential equation (ODE)

$$dT = -kTdt \ (k > 0) \tag{1-4}$$
$$Y_k = T_k + \epsilon_k \tag{1-5}$$

The temperature T is observed through the measurement Y which is encumbered with some measurement error. The solution to the ODE is an exponential decay, as illustrated by the straight line (on a log-scale) in Figure 1.4 where also the observations is depicted as a function of time. It is seen that the difference between the observations and the model predictions (called the residuals) are correlated in time. Such an autocorrelation of the residuals are often seen for output error models.
According to the ODE model the temperature T can be predicted exactly for all future time points. Also this fact illustrates a shortcoming of the ODE based modeling.
Let us now consider the corresponding stochastic state space model for which the dynamics are described by a stochastic differential equation (SDE)

$$dT = -kTdt + dw \tag{1-6}$$
$$Y_k = T_k + e_k \tag{1-7}$$

where w is assumed to be a Wiener process, ie. a stochastic process with Gaussian increments. The one-step predictions by the SDE are shown in Figure 1.4, and it is seen that the SDE adapts to the observations and gives a better description of the measurements. The prediction error is no longer autocorrelated and this fact actually provides

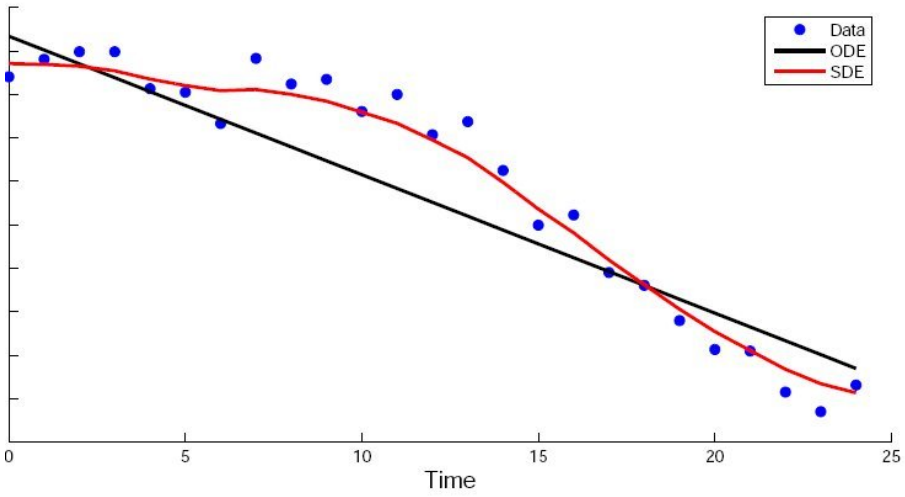


Figure 1.1: ODE (output error) and SDE (prediction error) based models

a framework for model validation. One of the most powerful methods for statistical model validation is to test whether the one-step prediction errors can be assumed to be a white noise process (see [Madsen, 2008]).
The solution to a stochastic differential equation is a stochastic process. This implies that the predictions of the future values of the temperature T are associated with some prediction error, and in fact the variance of the prediction errors depends on the prediction horizon.

1.5 Frequent readings and smart meters

In recent years conditional parametric dynamical models have been suggested for modeling the heat dynamics of buildings. In [Nielsen et al., 2010] a conditional parametric approach has been suggested for estimating the thermal characteristics of single family houses based on frequent readings of energy consumption and climate – see Table 1.2. In that table each row corresponds to a single family house.
The models used are essentially quite simple linear regression models, which are presented in detail later in the text. The time variations are estimated using locally weighted estimation of the linear model. The method gives local estimates in time of the model coefficients by only considering observations within a limited time window. This makes it possible to see if they are constant over time and is used to automatically the best period for fitting the model.

The thermal characteristics includes the response of the building to changes in temperature (heat loss transmission H_{tr} or as here UA-value), solar radiation (gA-value, here estimated for the global radiation), and wind (v-value, wind induced ventilation) – see Table 1.2. The effect of the wind is characterized both in terms of the wind speed and the wind direction, implying that v-values are estimated for different wind directions (except from north where basically no data was available). The proposted method also give a possibility for estimating the characteristic indoor air temperature as seen in Ta-

	UA	σ_{UA}	$\text{gA}^{\max}$	$v_E^{\max}$	$v_S^{\max}$	$v_W^{\max}$	T_i	σ_{T_i}
	W/°C		W	W/°C	W/°C	W/°C	°C	
4199586	169.8	8.0	779.3	15.0	10.9	11.3	25.7	1.2
4199598	204.3	8.7	606.3	58.2	57.0	54.4	19.2	0.7
4218597	180.4	6.9	358.3	2.4	2.2	11.4	24.4	0.9
4218598	217.7	8.6	481.0	10.5	7.9	6.1	21.0	0.8
4340506	187.6	8.7	621.7	3.3	5.6	6.8	22.9	1.0
4381443	211.3	9.0	356.5	1.4	13.0	16.7	22.9	0.9
4381449	220.4	10.5	799.5	31.7	35.0	34.5	17.7	0.7
4711160	153.9	5.8	472.9	16.3	5.3	11.0	20.6	0.7
4711161	180.6	5.4	548.2	9.3	8.9	13.2	22.0	0.6
4711176	173.3	4.8	611.7	5.9	1.0	12.8	23.8	0.6
4724102	161.2	7.0	257.2	11.8	17.6	22.8	21.5	0.9
4724106	127.4	4.0	230.0	11.6	11.9	8.1	18.6	0.6
4836681	156.5	8.1	578.1	41.4	29.3	22.2	23.0	1.1
4964553	114.6	5.7	607.8	6.3	12.3	8.9	21.3	0.9
4986050	154.5	10.3	664.0	11.1	8.7	5.5	19.2	1.3
5036505	222.6	8.1	373.5	3.5	12.6	14.9	18.7	0.6
5107720	186.0	14.4	685.4	42.1	30.6	16.3	20.1	1.5
5197381	185.3	9.1	870.4	38.1	27.5	27.5	21.2	1.0

Table 1.2: Estimates based on district heating consumption only. The values for the influence of solar radiation and wind speed is scaled such they are compareable with the UA-values using the 90 pct quantile of the relevant solar radiation and the wind speed.

ble 1.2.

In case only stationary characteristics are of interest it is concluded that it is sufficient to base the estimation on 24 hour averages, whereas in order to estimate the dynamic effects averages with a sample period around 4 hours are required. Also, it is beneficial to use measurements of the total energy consumption (heat and electricity). However, in most cases it results in adequate estimates if only the heat consumption is used as the response variable. Obvious exceptions are cases where eg. electrical floor heating is used for some periods. The estimated thermal characteristics have been analyzed with respect to background information regarding the households. The information is obtained via questionnaires and via the Danish Building Register (BBR).

Chapter 2

Introduction to R

In following text exercises are given in statistical programming language R . The R software is open source and in this chapter a short introduction is given. For Matlab users there is a very useful reference table, which list equivalent functions in R and Matlab. It can to be found here www.math.umaine.edu/~hiebeler/comp/matlabR.html.

2.1 Installation

Download and install R from `www.r-project.org` for your operating system (OS). This installs the software which can interpret the R language, it is not an editor for R code (though in Windows it does come with a basic editor).

A recommended and feature full editor is RStudio can be downloaded for the most common OS on rstudio.org. For Emacs users R is integrated via ESS, see ess.r-project.org.

2.2 The R language

R is a scripting language, which means that it is interpreted and evaluated one line at a time. It is really simple: An expression is written in the console, R executes it and prints the results. Just like a pocket calculator!

When using an editor the code lines are sent from the editor to the R console and executed, one at a time.

2.2.1 Working directory

The working directory of R is used every time a relative file path is given. For some exercises given later in the text it is needed to set the working directory to where the exercise scripts are located.

In RStudio use the menu:
`Session->Set Working Directory->To Source File Location` to set the working directory to the folder of the script (copy the line into the beginning of the script in order to do this step automatically next time the script is used).

2.2.2 Quick introduction

Open the script `"examples/R_Intro.R"` to find the code below and run it to get some basic R practice.

|||| Example 2.1

```
## A line starting with a # is a comment and will not be executed by R

## Run the next line to set a variable x. In RStudio the line can be run by
## pressing 'Ctrl-Enter', other code runinng commands see menu 'Code->Run
## Region'
x <- 1
## See the value of x
x
## All variables i R has a class
class(x)
## To read the help of numeric
`?`(numeric)
## The R help can also be searched by e.g.
`?`(`?`(regression))
## Create a floating point of length 100
x <- numeric(100)
x
## Other basic elements, check the 'see also' section and 'examples' in
## bottom of the page
`?`(integer)
`?`(logical)
`?`(character)
## Combine elements into a vector
c(1, 4.2, -8)
c("this", "is", "a", "vector", "or", "array", "of", "characters")
## Paste them into one character
paste("this", "is", "a", "single", "character", "element")
## An integer sequence
1:10
## A sequence
seq(-2, 2, by = 0.1)
```

```r
## The usual way of handling a data set is with a data.frame, which is simply
## like a table in a spreadsheet program, i.e. a two-dimensional matrix Make
## a data.frame with three columns
X <- data.frame(t = 1:10, y1 = seq(1, 0, len = 10), y2 = seq(-2, 5, len = 10))
class(X)
## The entire data.frame
X
## The first rows
head(X)
## The last rows
tail(X)
## The columns have names
names(X)
## The columns have different classes. Column 't' is an integer and the
## others are numeric
str(X)
## Access one column
X$y1
## In RStudio try auto-completion, put the cursor in front of the dollar sign
## and press 'tab'
X$
## Take a subset of the data.frame
X[, c("t", "y2")]
## Take the two first columns
X[, c(1, 2)]
## Take all, but second and third columns (negative indexes remove)
X[, c(-2, -3)]
## Take row 4 to 8
X[4:8, ]
## Index with a TRUE-FALSE statement
X$y1 < 0.5
X[X$y1 < 0.5, ]
## Operations are element wise (opposed to Matlab)
sqrt(X[, c("y1", "y2")] * X[, c("y1", "y2")])
## Matrix operations are with '%' around, and transpose is with t()
as.matrix(X[, c("y1", "y2")]) %*% t(as.matrix(X[, c("y1", "y2")]))
```

Chapter 3

Probability, statistics and time series analysis

In this chapter the elements from probability and statistical theory, which are needed for the present "soft" introduction to time series analysis, are introduced. The probability theory provides the techniques needed to describe randomness.

3.1 Random variable

First the basic building block which is needed to describe a random outcome of an experiment is introduced, namely a *random variable*.

Definition 3.1

A *random variable* represents the outcome of an experiment, which has yet not been carried out. In this text random variables are all denoted with capital letters and are associated to a point in time t, for example

$$Y_t \tag{3-1}$$

The random variable follows a distribution, which is denoted e.g.

$$Y_t \sim N(\mu, \sigma^2) \tag{3-2}$$

indicating that Y_t follows a normal (or Gaussian) distribution with mean μ and variance σ^2.

After the experiment has been carried out N times a set of outcomes (or measurements or observations) of the random variable are available as

$$\{y_t : t = 1, \ldots, N\} \tag{3-3}$$

The outcome is called a realization of the random variable.

An experiment in this context can have different meanings, it can be a "regular" experiment which is carried out under controlled conditions e.g. in a laboratory, where some measurements are collected from a setup. However, it can just as well be in settings which are not controlled, where simply some process or system is observed. Hence an experiment can be thought of as any setting in which the outcome cannot be known exactly on beforehand, this for example also include the measurement noise, which are "errors" caused in the system used to observe with.

A random variable has a *probability density function* (pdf) which defines the probability of observing the outcome in a given range. The pdf is parameterized by a set of parameters and is further characterized by several measures, most important is the *mean* and the *variance* (the first and second central moments). Many times in statistical modeling variables are assumed independent and identically distributed, denoted by i.i.d. in the text. Finally, measures defining the dependency between random variables are important, foremost the *covariance* (and *correlation*) which defines linear dependency. If the reader doesn't recall these basic statistical terms, they should be refreshed from text on introduction to statistics, the following can be recommended [Wasserman, 2004] and [Diez et al., 2012], the latter is freely available at `www.openintro.org/stat/`.

|||| Example 3.2

One important point which often cause some confusion: The *average* (or *sample mean*) can be calculated of a realization and is denoted with a bar

$$\bar{y} = \hat{\mu}_Y = \frac{1}{N}\sum_{t=1}^{N} y_t \tag{3-4}$$

is not the same as the *mean* (or expected value) of the random variable

$$\mu_Y = E(Y) = \int_{-\infty}^{\infty} y\, f_Y(y) dy \tag{3-5}$$

where $f_Y(y)$ is the pdf of Y. The mean is kind of the "real average", the average is an estimate of it. The average asymptotically converges to the mean as more observations are available

$$\lim_{N\to\infty} \hat{\mu}_X = \mu_X \tag{3-6}$$

This is likewise for the *sample variance*

$$\hat{\sigma}_Y^2 = \frac{1}{N-1}\sum_{t=1}^{N} (y_t - \bar{y})^2 \tag{3-7}$$

which converges to the *variance*

$$\mathrm{Var}(Y_t) = \sigma_Y^2 = E[(Y-\mu_Y)^2] = \int_{-\infty}^{\infty} (y-\mu_Y)^2\, f_Y(y) dy \tag{3-8}$$

which is the squared *standard deviance* σ_Y. Sometimes these distinctions are messed up in texts and especially while speaking!

Now further confusion can occur since the normal distribution has two parameters

$$Y_t \sim N(\mu_Y, \sigma_Y^2) \tag{3-9}$$

which happens to be the mean and variance of the random variable

$$E(Y_t) = \mu_Y \tag{3-10}$$
$$\text{Var}(Y_t) = E[(Y_t - \mu_Y)^2] = \sigma_Y^2 \tag{3-11}$$

see that the μ and σ^2 symbols (which are used for denoting the mean and variance) in this case also are used to represent the parameters in the distribution.

However, this not the case for other distributions, for example the *beta distribution* has two parameters

$$X_t \sim \text{Beta}(\alpha, \beta) \tag{3-12}$$

and its mean and variance are

$$E(X_t) = \mu_X = \frac{\alpha}{\alpha + \beta} \tag{3-13}$$
$$\text{Var}(X_t) = E[(X_t - \mu_X)^2] = \sigma_X^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \tag{3-14}$$

To clarify this, try in R

```
## Generate N values from a normal distribution with mean and sd (standard
## deviance)
N <- 100
y <- rnorm(N, mean = 2, sd = 2)
## The average (or sample mean)
mean(y)
## The sample variance
var(y)
## The sample standard deviance
sd(y)
sqrt(var(y))
## The Beta distribution (shape1 is alpha and shape2 is beta)
x <- rbeta(N, shape1 = 2, shape2 = 3)
## The average (or sample mean)
mean(x)
## The sample variance
var(x)
## The sample standard deviance
sd(x)
```

3.2 Stochastic process and time series

A *stochastic process* is a series of random variables each associated with a time point. It is denoted by

$$\{Y_k\} = \{Y_{t_k}\} = \{Y_{t_k} : \ k = 1, \dots\} \tag{3-15}$$

where $t_{k-1} < t_k$ denotes the time points. A *time series* is a realization (i.e. the experiment has been carried out and data is available) of the stochastic process and is denoted by

$$\{y_{t_k}\} = \{y_{t_k} : \ k = 1, \dots, N\} \tag{3-16}$$

where N is the number of observations.
Often a system is sampled at *equi-distant* time points, such that the *sampling time* $\Delta t = t_k - t_{k-1}$ is constant, forming a *discrete time stochastic process*

$$\{Y_t\} = \{Y_t : \ t = 1, 2, \dots\} \tag{3-17}$$

here with scaled sampling time $\Delta t = 1$. A realization of discrete time stochastic process with N observations is denoted by

$$\{y_t\} = \{y_t : \ t = 1, \dots, (N-1)\} \tag{3-18}$$

In some parts of the text processes and realizations are denoted in vector format, such that a stochastic process is $\{Y_t\} = \boldsymbol{Y}$ and a realization is $\{y_t\} = \boldsymbol{y}$.

3.2.1 Stationary process

A stationary process is a process which at any time t follows the same distribution, hence its properties does not depend on the time at which it is observed.

|||| Definition 3.3 Stationary process

A stochastic process $\{Y_t\}$ is called stationary if $\{Y_t\}$ has the same distribution as $\{Y_{t+h}\}$ for any $h > 0$.

3.3 Backshift operator and ACF

The backshift operator and auto-correlation function (ACF) is presented in this section. They are very fundamental in time series analysis and are used heavily in system identification and validation. For more details, for example also the important *cross-correlation* function, see Section 5.2.2 in [Madsen, 2007].

3.3.1 Backshift operator

The backshift (or backward shift) operator B is applied to discrete time series to get the previous (or lagged or time-shifted) variables

$$B^k X_t = X_{t-k} \tag{3-19}$$

A polynomial of order p in the backshift operator can be used to condense the notation of a series of lagged variables

$$\omega(B) = \omega_0 B^0 + \omega_1 B^2 + \cdot + \omega_p B^p \tag{3-20}$$

such that

$$\omega(B) Y_t = \omega_0 Y_t + \omega_1 Y_{t-1} + \cdots + \omega_p Y_{t-p} \tag{3-21}$$

For more details see around page 82 in [Madsen, 2008].

Example 3.4

This is an example of how to lag a series in R

```r
## Example: Backshift operator and ACF

## Make a series of N values
N <- 8
## Set x to the values 1,2,...,N
x <- 1:N
## Lag (back-shift) x one step
c(NA, x[1:(N - 1)])
```

```
## [1] NA  1  2  3  4  5  6  7
```

```r
## Print to see x versus x lagged k step
k <- 2
data.frame(x, x.l2 = x[c(rep(NA, k), x[1:(N - k)])])
```

```
##   x x.l2
## 1 1   NA
## 2 2   NA
## 3 3    1
## 4 4    2
## 5 5    3
## 6 6    4
## 7 7    5
## 8 8    6
```

```r

## Define a helping function for lagging
lagVec <- function(x, k) {
    ## Shift the x vector k steps to the right
    c(rep(NA, k), x[1:(length(x) - k)])
}
## Test the function
lagVec(x, k = 1)
```

```
## [1] NA  1  2  3  4  5  6  7
```

3.3.2 ACF

The auto-correlation function (ACF) is a measure of linear lag dependency in a time series. It is simply the correlation between lagged values in a time series. It only has a meaning if the process generating the series is (weakly) stationary.

Definition 3.5

The auto-correlation function (ACF) is

$$\rho_{Y_t}(k) = \text{Corr}[Y_t, Y_{t+k}] = \frac{\text{Cov}[Y_t, Y_{t+k}]}{\sigma^2_{Y_t}} \tag{3-22}$$

where σ^2_Y is the variance of Y_t.

Example 3.6

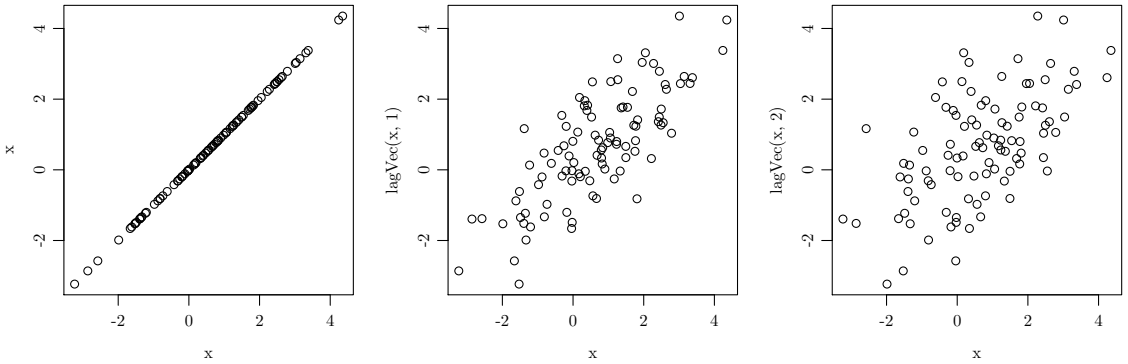
This is an example of how the ACF is calculated. A time series is generated by simulating a simple AR process (the AR part of an ARMAX model, see Section 5.2.1)

$$X_t = 0.8X_{t-1} + \varepsilon_t \tag{3-23}$$

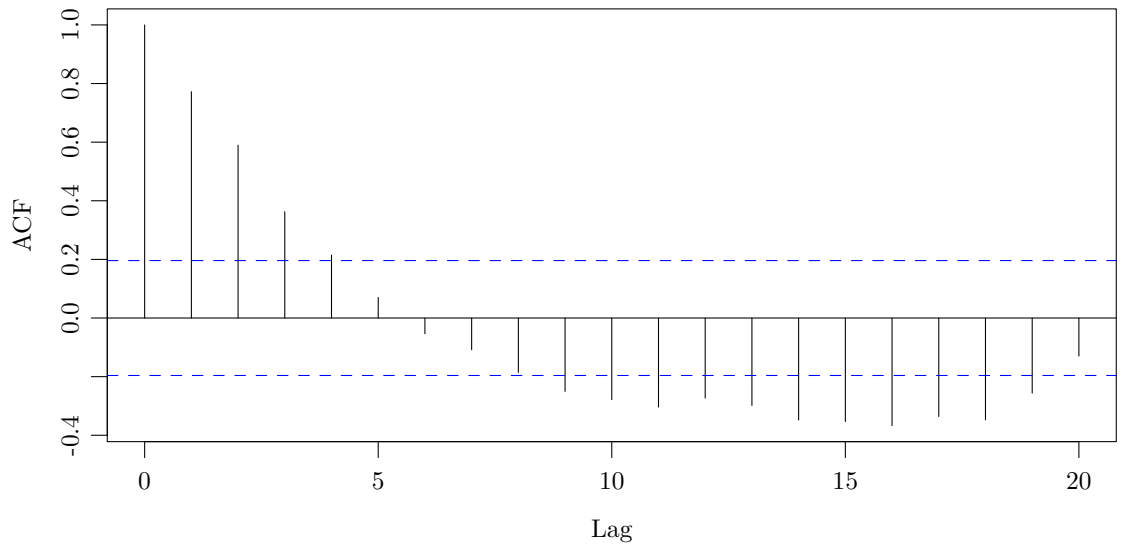
with the R code

```
## Investigate the ACF.  A numeric (vector) of length N
N <- 100
x <- numeric(N)
## For each time point (here i) take 0.8 of the last value and add a random
## number
x[1] <- rnorm(1)
for (i in 2:N) {
    x[i] <- 0.8 * x[i - 1] + rnorm(1)
}
## Plot it
plot(x, type = "l")
```

```
## Then plot it versus lagged versions to see how the spread of the points
## increase as the lag increase
par(mfrow = c(1, 3))
plot(x, x)
plot(x, lagVec(x, 1))
plot(x, lagVec(x, 2))
```




```
## Now the ACF can be calculated by
cor(x, lagVec(x, 0), use = "complete.obs")
## [1] 1

cor(x, lagVec(x, 1), use = "complete.obs")
## [1] 0.7732

cor(x, lagVec(x, 2), use = "complete.obs")
## [1] 0.5907

## go on, or simply call the function
acf(x)
```



```
## See how the values of the correlations for different lags are equal to the
## bars in the acf() plot. The blue lines of the plot indicate a 95%
## confidence band.
```

3.4 White noise

An important stochastic process is called *white noise*. This process is an i.i.d. series, hence a sequence of independent, identical distributed random variables, in this text denoted by

$$\{\varepsilon_t\} \tag{3-24}$$

A property of white noise is

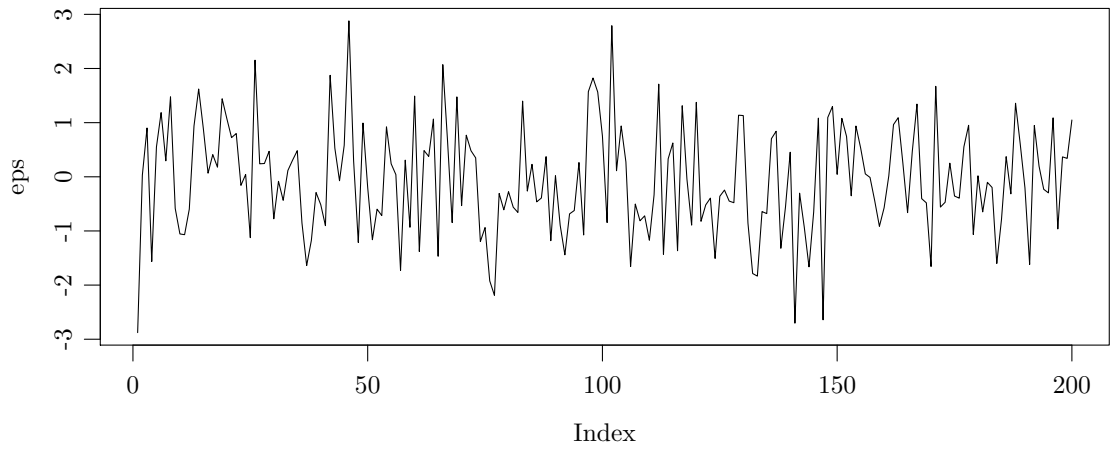
$$\rho_{\varepsilon_t}(k) = 0 \text{ for } k \neq 0 \tag{3-25}$$

It is called white noise because the frequency spectrum is flat.

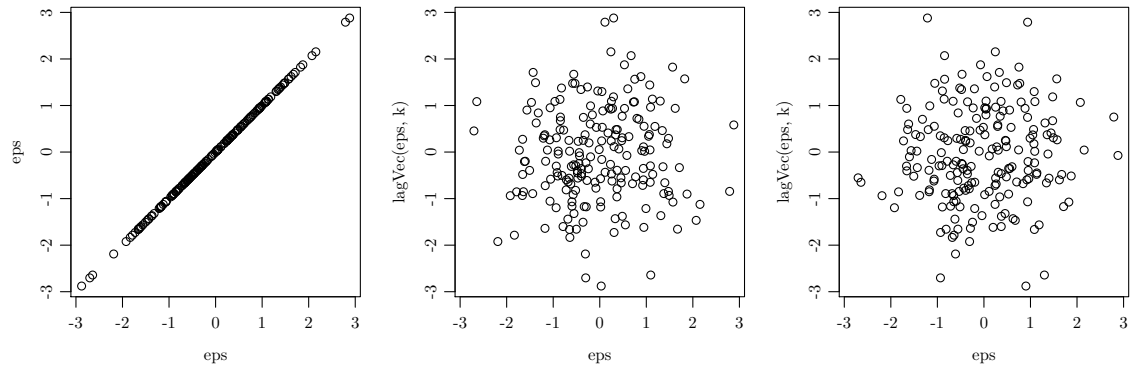
|||| Example 3.7

This is a simple where the properties of white noise are investigated.

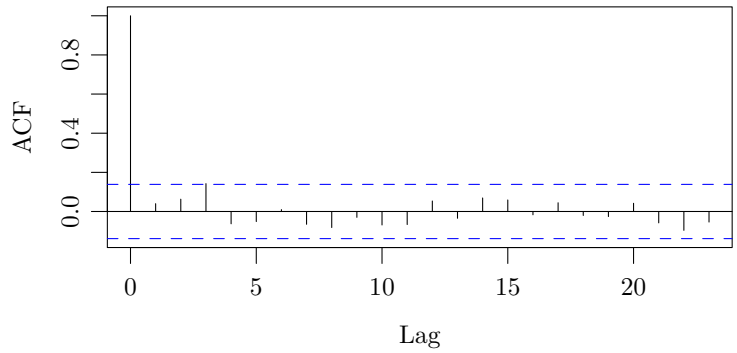
```
## First a white noise time series is generated
N <- 200
eps <- rnorm(N)
plot(eps, type = "l")
```



```
## Then the sequence is plotted versus the sequence lagged zero, one and two
## steps
par(mfrow = c(1, 3))
plot(eps, eps)
k <- 1
plot(eps, lagVec(eps, k))
k <- 2
plot(eps, lagVec(eps, k))
```




```
## Calculate the ACF for the generated white noise
acf(eps)
```



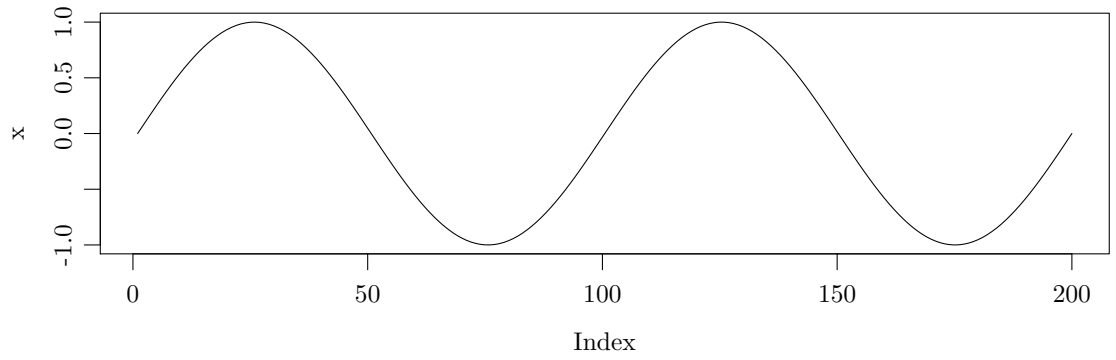
See how approximately every 20 bar is outside the confidence band.

One important point is that the ACF can be non-significant for other processes than white noise. Hence a non-significant ACF does not directly indicate white noise, a few extreme values can make the ACF look like it is from white noise, when it is not, as demonstrated in the following example.

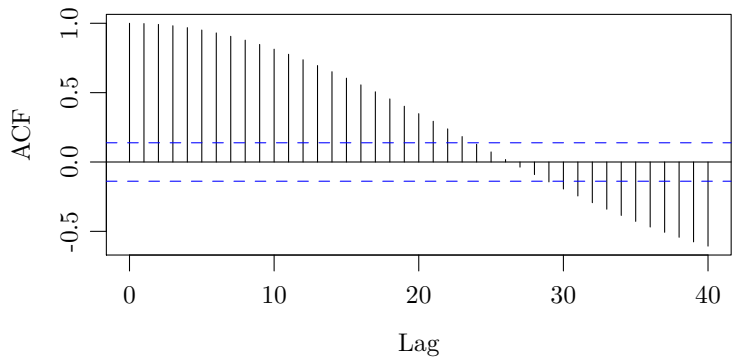
Example 3.8

In this example it is shown how a non-white noise sequence can have the same non-significant ACF as white noise.

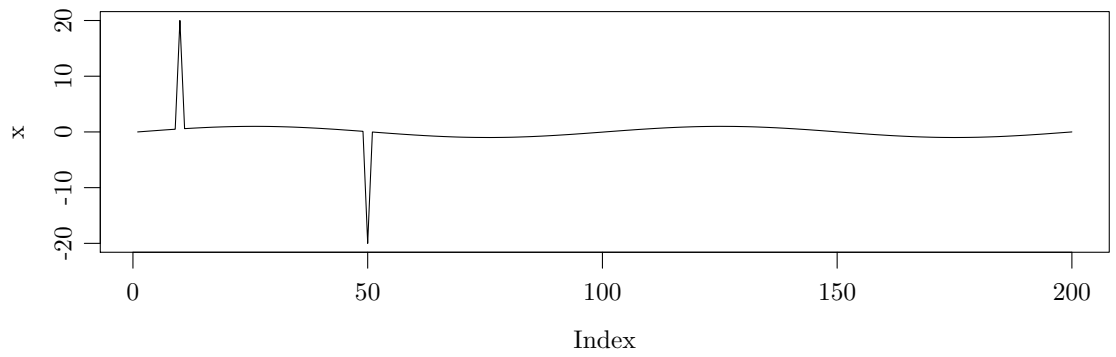
```
## Set a vector of time
time <- seq(0, 4 * pi, len = 200)
## Make a sine
x <- sin(time)
## Plots it
plot(x, type = "l")
```



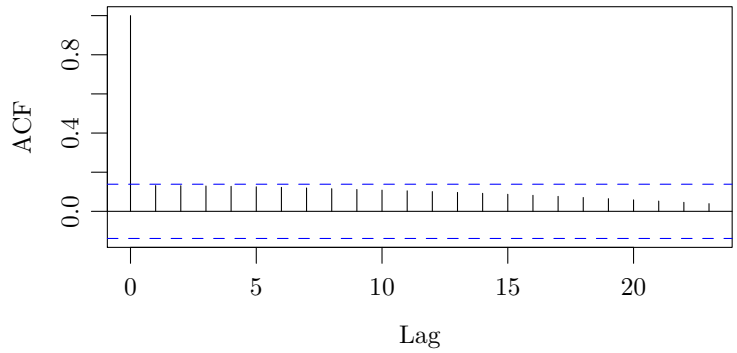

```
## The ACF of a sine is a sine
acf(x, lag.max = 40)
```




```
## Add two outliers
x[10] <- 20
x[50] <- -20
## Plots it
plot(x, type = "l")
```




```
## The ACF
acf(x)
```




```
## see that the ACF is non-significant as for white noise, however it
## certainly is not white noise
```

More ways to test for white noise exists: runtest etc.

Finally, a short example of how the ACF reveals the process, i.e. it can be used for system identification, and stationarity. For a much elaborated system identification introduction, see Chapter 6 in [Madsen, 2007].

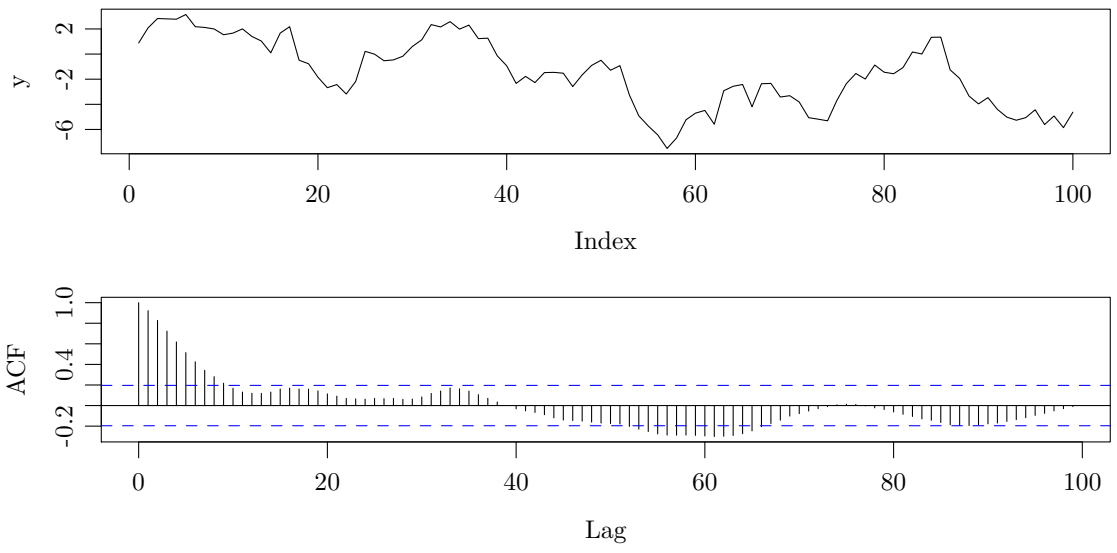
|||| Example 3.9

Here an example where it is shown how a pattern in the ACF can be used to reveal the properties of the process, including stationarity.
Consider a simple auto-regressive (AR) process (see Section 5.2.1 for definition of AR processes)

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t \tag{3-26}$$

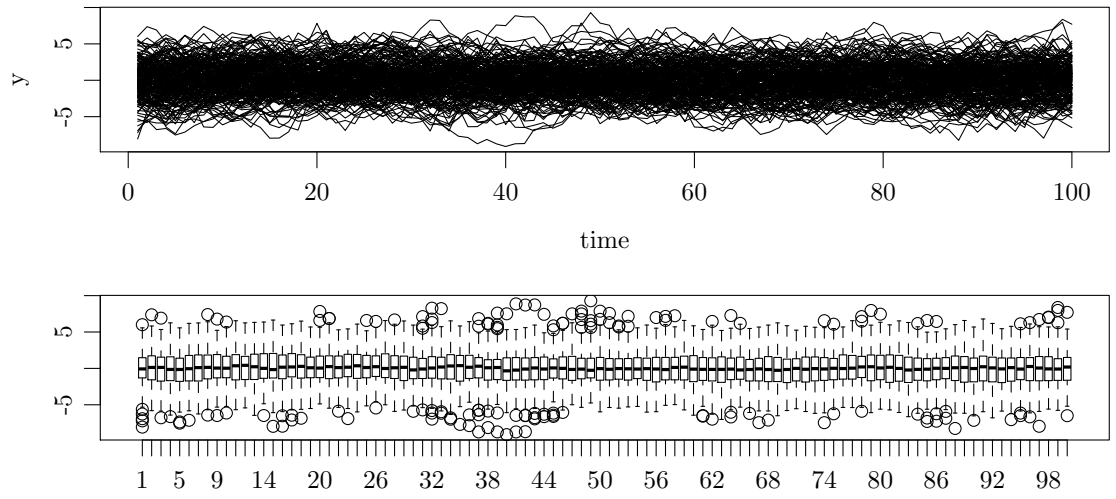
If $\phi_1 < 1$ the process is stationary, if $\phi_1 \geq 1$ then the process is not stationary. Try with simulation to see the properties of the process

```r
## Number of samples
N <- 100
## A little function for simulation of an AR(1) process
simAR1 <- function(phi1, N, Nburnin = 30) {
    ## Do Nburnin points as a burn-in period
    Nsim <- N + Nburnin
    ## White noise
    eps <- rnorm(Nsim)
    ## Output vector
    y <- numeric(Nsim)
    ## Simulate
    for (i in 2:Nsim) {
        y[i] <- phi1 * y[i - 1] + eps[i]
    }
    ## Remove the burn-in period
    y <- y[-1:-Nburnin]
    return(y)
}
## Simulate the process, when it is stationary
y <- simAR1(phi1 = 0.95, N = 100)
## An exponential decaying ACF indicates an AR process
par(mfrow = c(2, 1))
plot(y, type = "l")
acf(y, lag.max = 100)
```




```r
## Question: Try changing the phi1 coefficient. How does this change the ACF?
```

```r
## Now repeat the simulations many times see the distribution of the process
## Number of repetitions
Nrep <- 200
## Length of simulation
N <- 100
## Keep in a matrix
yMat <- matrix(NA, nrow = N, ncol = Nrep)
for (i in 1:Nrep) {
    yMat[, i] <- simAR1(phi1 = 0.9, N = 100)
}
## Make a plot of all the simulations
par(mfrow = c(2, 1))
plot(yMat[, 1], type = "n", ylim = range(yMat), xlab = "time", ylab = "y")
for (i in 1:Nrep) {
    lines(yMat[, i])
}
## A box plot for each time point
boxplot(t(yMat))
```




```r
## Question: Do the repeated simulations with a non-stationary AR process
## (phi1 >= 1, how does this change the distribution depending on the time?
```

3.5 Notation and terms: statistical model, prediction, estimate, error, residuals, hats, ...

This section gives a little overview of how statistical models are usually notated and what the different terms are called.

A *statistical model* has the form

$$\text{output} = \text{function(inputs)} + \text{error} \tag{3-27}$$

where the random (stochastic) part is the *error*. The statistical model of the *output* is the entire right side.

In usual notation Y_t is the output, X_t is the input and ε_t is the error

$$Y_t = f(X_t) + \varepsilon_t \tag{3-28}$$

Note that in state-space models X_t is usually the states and the inputs are U_t. What is sought is the *estimate* of the function relating the inputs to the outputs. An estimate has been found when a model has been *fitted* to some data. The estimate is indicated by a hat

$$y_t = \hat{f}(x_t) + \hat{\varepsilon}_t \tag{3-29}$$

The parts are now usually called

$$\text{observations} = \text{predictions} + \text{residuals} \tag{3-30}$$

|||| Example 3.10

If the function $f(X_t)$ is linear, then it is a *linear model*, for example the most simple with one input and one output is

$$Y_t = \mu + \omega X_t + \varepsilon_t \tag{3-31}$$

where μ is called the *intercept* and usually $\varepsilon_t \sim N(0, \sigma^2)$. The *model parameters* are

$$\theta = (\mu, \omega, \sigma) \tag{3-32}$$

where μ and ω are also called the *coefficients*. The model is now fitted using an estimation method (as described in the following chapter) and estimates of the parameters are found

$$\hat{\theta} = (\hat{\mu}, \hat{\omega}, \hat{\sigma}) \tag{3-33}$$

and the residuals are found by subtracting the predictions from the observations

$$\hat{\varepsilon}_t = y_t - \hat{f}(x_t) = y_t - (\hat{\mu} + \hat{\omega} x_t) \tag{3-34}$$

Note that the *error* and the *residual* are often confused. The error is a random variable formed by the difference between the output and the ”real” model function of the inputs. The residuals are the realization (observation) of the error, hence when the model is fitted to data the error are observed as the residual.

3.6 Model validation and selection

One very important part of statistical modeling is the *model validation* (or *model checking*). The starting point is the distributional assumptions of the model error, which

usually is that the error is white noise. Hence after fitting the model it can be validated by checking if the residuals are indeed i.i.d. and how they are distributed. The model validation is also a very important ”detectives tool” in the process of finding a suitable (the best) model. The following validation steps should be carried out before the model is accepted

- The ACF should be insignificant indicating that the residuals are not significantly different from white noise residuals
- Scatter plots of the residuals vs. the inputs should show no clear patterns. Plots of residual vs. other available input variables not used in the model can be used to see if other inputs should be added to the model. So if the residuals are higher in some period, it can seen if one of the inputs a large fluctuation in the similar period
- Time series plots of the residuals and the inputs to see if any patterns in the residuals occur which can be linked to the inputs
- Histogram and QQ-plots of the residuals should be used to verify the distribution of the residuals

3.7 Model selection

Statistical model selection is concerned with techniques for selection of a suitable model for a given set of data.

If a model is too simple it is called *under-fitted* or *under-parameterized*, indicating that is doesn't contain enough elements or parameters to describe the data well enough. If a model is under-fitted then it will be *biased*, meaning that it is too “rigid” and not able to “follow” the data.

If a model is too complex it is called *over-fitted* or *over-parameterized*, indicating that it contains too many and thereby redundant elements or parameters. In this case the model will be too *varying*, meaning that it follows the data too much and when presented with new data it will not be able to predict well and the parameter estimates will not be asymptotically correct. The most over-fitted model will match the data perfect by just looking up the output using the data (it has a parameter for each point in the data), however when presented with new data it will not predict very well.

Hence the *suitable* model, which is neither under- nor over-fitted, must be found. This is however not straightforward, since regular measures of how well the model fit data (R^2, likelihood, etc.) increase as the complexity of the model increases. Therefore the following three approaches can be used

- Adjusted measures taking the model complexity (number of parameters) into account, such as the Akaike information criteria (AIC), Baysian information criteria (BIC) or adjusted R^2.

- Statistical tests for model extended models. It can be tested whether an extended model increase the description of the data significantly, for example with an *F test* or *likelihood-ratio test*.
- Dividing the data into a *learning* (or *training*) and a *test set*. Byfitting the model to the learning set and then evaluating on the test set, the problem of overfitting the model is avoided. This is also the principle used in *leave one out* techniques.

3.7.1 Selection procedure

A model selection procedure can be applied to systematically find a suitable model. Most applied is a *forward selection* approach: starting from a simple model the most significant model parts are added iteratively until no significant improvement of the model fit is found. A *backward selection* procedure starts with the most complex model and from which the non-significant parts are iteratively removed.

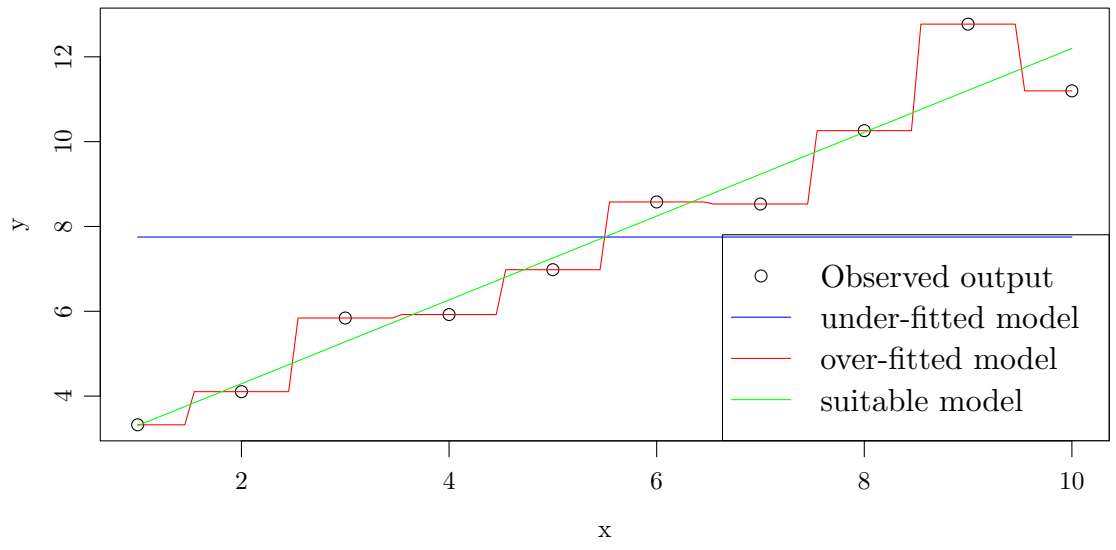
The subject of model selection is really important and should have been covered more in detail here. However, this is left for a more advanced course on statistical modeling and the reader is referred to the following material:

- Introduction to basic principles of model selection can be found in [Madsen and Thyregod, 2010], [Hastie et al., 2009], and many other books on statistics
- Model selection for time series models, see [Madsen, 2007]
- A procedure for selection of a suitable grey-box model are presented in [Bacher and Madsen, 2011]

Example 3.11

A short example of an under-fitted and over-fitted model.

```
## Generate data from a very simple process
N <- 10
x <- 1:N
y <- x + rnorm(N, mean = 2, sd = 1)
## An under-fitted model would be a constant, which we can fit with LS
## (described later in text)
underfitted <- lm(y ~ 1)
## An over-fitted model could for example be the model which simply looked up
## the data we have
overfitted <- function(xnew, x, y) {
    ## Find the y value closest to the data in x
    return(y[which.min(abs(xnew - x))])
}
## The suitable model, is in this case not difficult to find since we knOw
## the process which generated the data
suitable <- lm(y ~ x)

## Plot the three models
xseq <- seq(min(x), max(x), len = N * 10)
plot(x, y)
lines(xseq, sapply(xseq, overfitted, x = x, y = y), col = "red")
lines(xseq, predict(underfitted, newdata = data.frame(x = xseq)), col = "blue")
lines(xseq, predict(suitable, newdata = data.frame(x = xseq)), col = "green")
legend("bottomright", c("Observed output", "under-fitted model", "over-fitted model",
    "suitable model"), lty = c(0, 1, 1, 1), pch = c(1, -1, -1, -1), col = c("black",
    "blue", "red", "green"))
```



```r
## The important point is that if we try to measure how well they fit our
## data, for examplewith the R^2 R^2 indicate how well the model fits data
## and is between zero (a model as poor as the constant model) and one (a
## perfect model)
rSquared <- function(ypred, x, y) {
    1 - sum((y - ypred)^2)/sum((y - mean(y))^2)
}

## The under-fitted model
rSquared(predict(underfitted, newdata = data.frame(x = x)), x, y)
```

```
## [1] 0
```

```r
## The over-fitted has a perfect prediction of the data
rSquared(sapply(x, overfitted, x = x, y = y), x, y)
```

```
## [1] 1
```

```r
## Hence if we just looked at this we would conclude that the over-fitted
## model was very good!  However when we try to predict new data from the
## same process
ynew <- x + rnorm(N, mean = 2, sd = 1)
## Then neither the under-fitted nor over-fitted are the best
rSquared(predict(underfitted, newdata = data.frame(x = x)), x, y)
```

```
## [1] 0
```

```r
rSquared(sapply(x, overfitted, x = x, y = ynew), x, y)
```

```
## [1] 0.7907
```

```r
## Then the suitable model is the best to predict the new data
rSquared(predict(suitable, newdata = data.frame(x = x)), x, y)
```

```
## [1] 0.9461
```

Chapter 4

Estimation

Different estimation techniques exists. In two fundamental techniques are presented: Least squares (LS) and maximum likelihood (ML). Actually, the LS estimates are equivalent to the ML estimates if the residuals are white noise (some slight differences, which are only important if only few sample are available).

One important point is to notice is:

- *Closed form*. The estimates can be calculated with a closed form expression, this makes the estimation technique very robust and fast. The calculations are simple matrix calculations. LS estimates for linear regression models can be calculated with closed form, usually this is simply referred to as LS estimation. For example this is what the R function `lm()` does. Linear static models and the simple dynamic models (ARX) can be estimated in closed form.
- *Non-closed form*. This is when the parameters are estimated by optimizing an objective function with a recursive optimization algorithm. This is necessary for dynamic models (ARMAX) and grey-box models (since they can only be fitted recursively). It can be LS estimation, however this usually what is meant with referring to ML estimation.

4.1 Estimation using closed form Least Squares Methods

The LS-estimator is given as the solution to the problem of finding the best estimates defined as the parameter values that minimize the mean square error between the observations and the predictions given by the model. In this section the closed form estimation algorithm is presented.

4.1.1 LS-estimates for linear regression models

This is in fact a projection in ^{N}.

Observations:

$$(y_1, x_1), (y_2, x_2), \cdots, (y_N, x_N). \tag{4-1}$$

Model:

$$Y_t = f(\boldsymbol{x}_t; \boldsymbol{\theta}) + \varepsilon_t \tag{4-2}$$

where $\boldsymbol{x}_t$ is a vector of known explanatory variables, and $\boldsymbol{\theta}$ is a vector of the unknown parameters.
LS-estimate (unweighted):

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\theta} S(\boldsymbol{\theta}) \tag{4-3}$$

$$S(\boldsymbol{\theta}) = \sum_{t=1}^{N} [y_t - f(\boldsymbol{x}_t; \boldsymbol{\theta})]^2 = \sum_{t=1}^{N} \epsilon_t^2(\boldsymbol{\theta}) \tag{4-4}$$

According to [Ljung, 1987] this is a Prediction Error Method (PEM).
An estimate of the variance of the estimator:

$$\text{Var}[\widehat{\boldsymbol{\theta}}] = 2\hat{\sigma}^2 \left[\frac{\partial^2}{\partial^2 \boldsymbol{\theta}} S(\boldsymbol{\theta}) \right]^{-1}_{|\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \tag{4-5}$$

where $\hat{\sigma}^2 = S(\widehat{\boldsymbol{\theta}})/(N-p)$.

4.1.2 The General Linear Model (GLM)

An important LS estimator occurs for the *General Linear Model* (also often referred to as simple or multiple linear regression):

$$Y_t = \boldsymbol{x}_t^T \boldsymbol{\theta} + \epsilon_t \tag{4-6}$$

For all observations the model is written:

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix} = \begin{pmatrix} x_1^T \\ \vdots \\ x_N^T \end{pmatrix} \boldsymbol{\theta} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_N \end{pmatrix} \tag{4-7}$$

or

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\epsilon} \tag{4-8}$$

We assume that $V[\boldsymbol{\epsilon}] = \sigma^2$.
The Ordinary Least Squares (OLS) (LS) estimator is found by minimizing:

$$S(\boldsymbol{\theta}) = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\theta})^T (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\theta}) \tag{4-9}$$

The estimate is found as the solution to the *normal equations*:

$$(\boldsymbol{X}^T \boldsymbol{X})\widehat{\boldsymbol{\theta}} = \boldsymbol{X}^T \boldsymbol{Y} \tag{4-10}$$

If $\boldsymbol{X}^T\boldsymbol{X}$ has full rang:

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \tag{4-11}$$

Proof: The normal equations follows by putting $\nabla S(\boldsymbol{\theta}) = -2\boldsymbol{X}^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\theta}) = 0$
The properties of the LS-estimator:

1. The estimator is linear in the observations.
2. The estimator is unbiased.
3. The variance is

$$\text{Var}[\widehat{\boldsymbol{\theta}}] = \text{E}\left[(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\right] = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}. \tag{4-12}$$

4. The estimator is BLUE (Best Linear Unbiased Estimator).

An unbiased estimator for σ^2 is

$$\hat{\sigma}^2 = S(\widehat{\boldsymbol{\theta}})/(N - p) \tag{4-13}$$

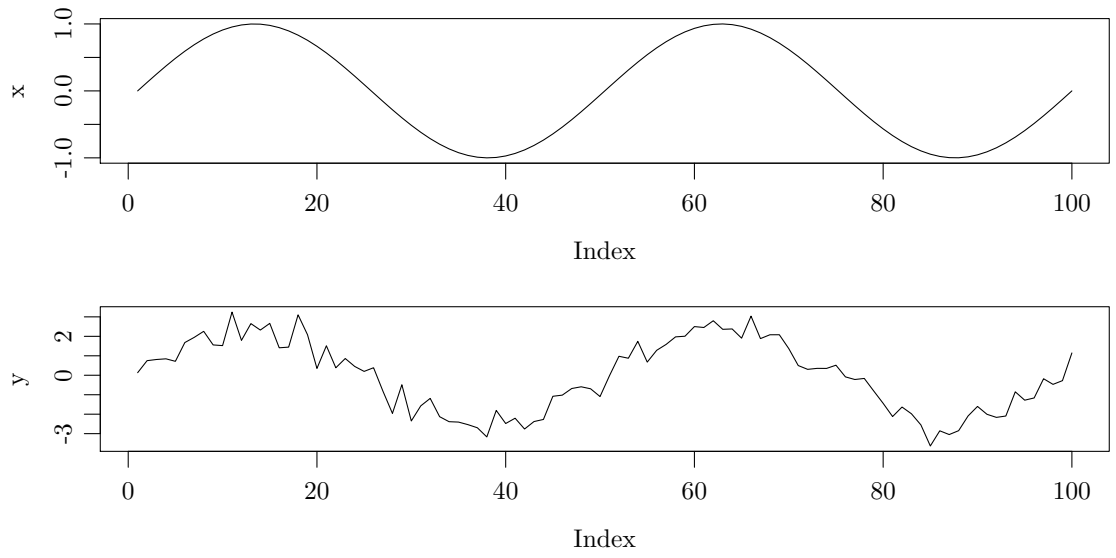
|||| Example 4.1 Example of closed form least squares estimation

First simulate a very simple linear regression model

$$Y_t = \alpha X_t + \varepsilon_t \tag{4-14}$$

where ε_t is white noise, i.e. $\varepsilon_t \sim N(0, \sigma^2)$.

```
## Set the parameters for simulation The coefficient for the input
alpha <- 2.4
## The standard deviance of the white noise (square root of the variance)
sigma <- 0.5
## Number of points
N <- 100
## Generate the input as a sine
x <- sin(seq(0, 4 * pi, len = N))
## Simulate the model
y <- alpha * x + rnorm(N, 0, sigma)
## Plot
par(mfrow = c(2, 1))
plot(x, type = "l")
plot(y, type = "l")
```




```
## Question: Plot a scatter plot of x vs. y to see their functional relation.
```

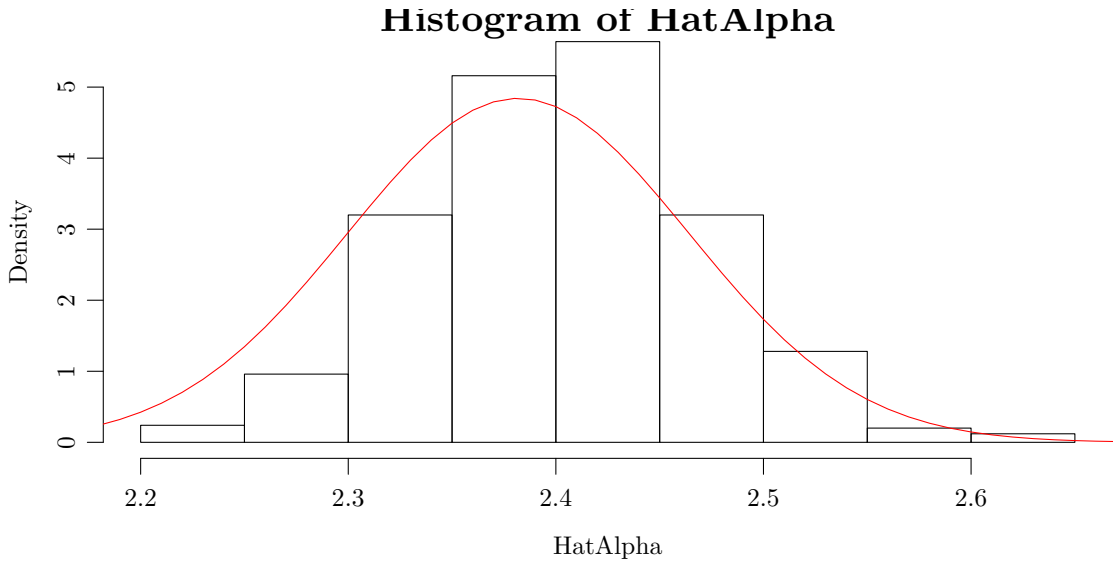
Now simulated data from the model is available, so time for estimation of the parameters. First use Eq. (4-11) the solution to the normal equations.

```
## Estimate the parameters with matrix calculations Set up the X matrix: a
## column of
X <- matrix(x, ncol = 1)
## And the Y matrix
Y <- matrix(y, ncol = 1)
## Calculate the parameters theta, in this case only consisting of alpha.
## The function solve() is used for matrix division
solve(t(X) %*% X, t(X) %*% Y)

## Or use the lm() function for the same Set the data into a data.frame
D <- data.frame(x = x, y = y)
## First argument is the model formula and second is the data, see ?lm and
## ?formula for further info
fit <- lm(y ~ 0 + x, data = D)
## Write out the result
summary(fit)
```

Now turn the attention to the uncertainty of the parameter estimates. First use Eq. (4-13) and (4-12)

```r
## Investigate the uncertainty of parameter estimates

## Calculate an estimate of the variance (sigma^2) of the error noise process
## The sum of squared residuals
S <- t(Y - X %*% Theta) %*% (Y - X %*% Theta)
hatVarEps <- S/(N - 1)
## Then use this to calculate the standard error (standard deviance of the
## parameter estimates Theta)
sqrt(hatVarEps %*% solve(t(X) %*% X, 1))

## Or see it from the result returned by the lm() function
fit <- lm(y ~ 0 + x, data = D)
## As the 'Std. Error'
summary(fit)
```

```r
## This is one of the fantastic features of statistics, not only is the value
## of the model parameters estimated, but also the parameter uncertainty. But
## what is the meaning of the 'Std. Error'?

## Under the assumed conditions (i.i.d. residuals) the parameter estimates
## are approximately normal distributed, which enable us to construct
## confidence bands.

## Try to ''see'' that the parameter estimates are normal distributed by
## repeating the simulation and estimation.  Repeat Nrep times
Nrep <- 500
## Number of samples
N <- 100
## For keeping the estimates
HatAlpha <- numeric(Nrep)
## Repeat
for (i in 1:Nrep) {
    ## Simulate the model
    y <- alpha * x + rnorm(N, 0, sigma)
    ## Estimate the parameters
    fit <- lm(y ~ 0 + x, data = data.frame(x, y))
    ## Keep the estimate
    HatAlpha[i] <- fit$coefficients["x"]
}
## Now see the distribution of the estimates
hist(HatAlpha, probability = TRUE)
## Draw the estimated distribution of Theta from the last fit
tmp <- seq(2, 2.8, by = 0.01)
xStdError <- summary(fit)$coefficients["x", "Std. Error"]
lines(tmp, dnorm(tmp, mean = fit$coefficients["x"], sd = xStdError), col = 2)
```




```
## A 95% confidence band is
fit$coefficients["x"] + 1.97 * c(-xStdError, xStdError)
```

```
## Question: Try to vary the number of generated samples (i.e. N).
## How does this affect the uncertainty of the estimates?

## Question: Is it true that the real parameter value falls inside the 95%
## confidence band in average 95% of the time? Verify by repeating the
## simulations.

## Question: Extend the simulation model by adding an intercept
## Y_t = \mu + \alpha X_t + \varepsilon_t
## and generate the data again.
## Now estimate the parameters either with matrix calculations (a column of
## ones must be added to the X matrix) or with the lm() function (the 0
## must be changed to a 1 in the formula or leaved out).

## Question: Try to investigate the impact of not using the right (suitable)
## model for estimation, are the parameter estimates be un-biased?
```

4.2 Estimation using Maximum Likelihood Methods

In this section a brief introduction to the maximum likelihood (ML) basis and how it is applied for estimation using a non-closed expressions. For more in-depth descriptions the reader is referred to [Pawitan, 2001].

4.2.1 Maximum likelihood

Maximum likelihood forms a basis for statistics very elegantly based on probability theory. Two components are needed: a set of data and a model describing the distribution of the data. With these the *likelihood* can be calculated, it is a measure of how likely it is that the data originates from that particular model.

|||| Definition 4.2 Likelihood

A stochastic process is characterized fully by its multi-dimensional (or joint) probability density function (pdf)

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) \tag{4-15}$$

The pdf can be used to calculate the probability of events when evaluated before the experiment is carried out, i.e. data is yet not available. After carrying out the experiment and data is available (as a realization of the random variable) the value of the pdf is called *likelihood*.
With a set of data and a model the likelihood is a function of the model parameters

$$L(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{x}) = f_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta}) \tag{4-16}$$

Hence the likelihood (i.e. not the "probability") of observing the output series $\boldsymbol{y}$ given the input series $\boldsymbol{x}$ and the model parameters $\boldsymbol{\theta}$. The model is given as the multi-dimensional (joined) pdf $f_{\boldsymbol{Y}}(\cdot)$.

Typically the model is formulated as a function of the inputs plus white noise (i.e. i.i.d. random variables)

$$Y_t = f(X_t,\boldsymbol{\theta}) + \varepsilon_t \tag{4-17}$$

where $\varepsilon_t \sim N(0,\sigma^2)$. If so then the *residual*

$$\varepsilon_t = y_t - f(x_t,\boldsymbol{\theta}) \tag{4-18}$$

is used for calculating the likelihood

$$L(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{x}) = f_\varepsilon(\varepsilon_1)\cdot f_\varepsilon(\varepsilon_2)\cdot \ldots \cdot f_\varepsilon(\varepsilon_N) \tag{4-19}$$

Note here that due to the i.i.d. assumption of the residuals, then the joined pdf is simplified to a product of the pdf for each of the observation. This is a very important point, since this simplification is almost always done and it allows for *model validation* (or *model checking*) where the residuals are analyzed to verify that they are white noise.

Definition 4.3 Maximum likelihood estimates

The maximum likelihood estimates are the parameter values which maximize the likelihood function

$$\hat{\boldsymbol{\theta}}_{\mathrm{ML}} = \max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{x}) \tag{4-20}$$

Under the assumption of white noise residuals and the model specified as above the calculation of the ML estimates are carried out by using the log-likelihood function

$$l(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{x}) = \log L(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{x}) = \log f_\varepsilon(\varepsilon_1) + \log f_\varepsilon(\varepsilon_2) + \cdots + \log f_\varepsilon(\varepsilon_N) \tag{4-21}$$

and negating and minimizing it

$$\hat{\boldsymbol{\theta}}_{\mathrm{ML}} = \min_{\boldsymbol{\theta}} -l(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{x}) \tag{4-22}$$

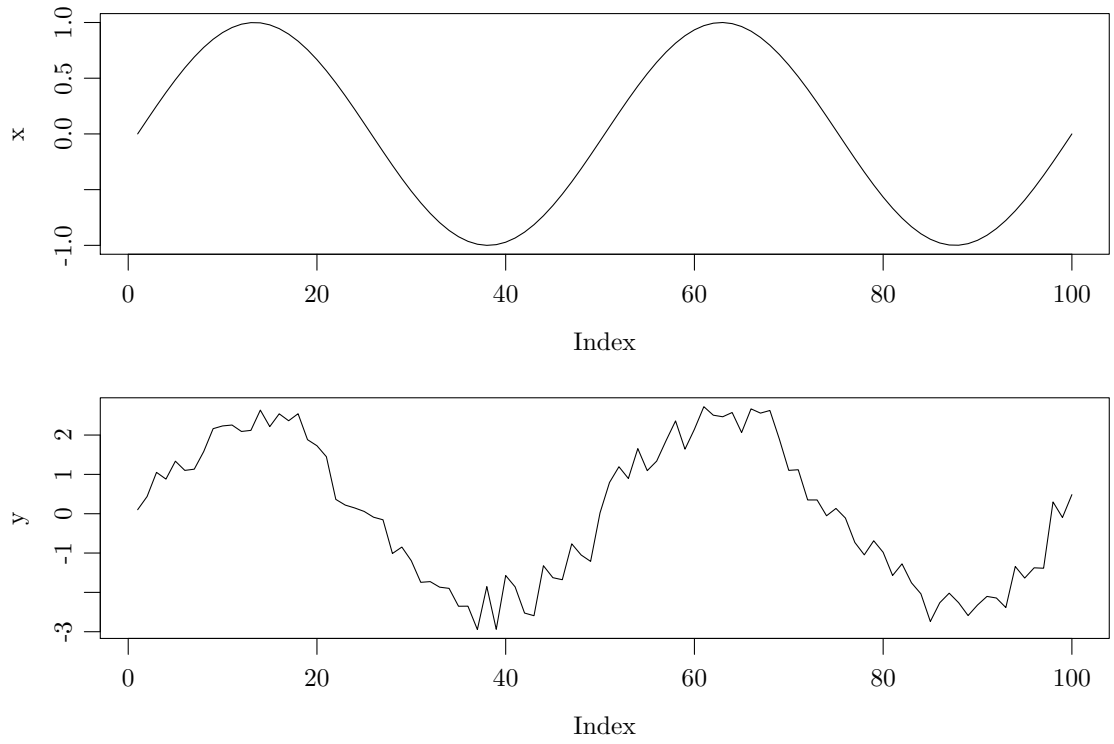
Example 4.4 Maximum likelihood estimation

In this example a simple model is simulated and the parameters are estimated. The same process is simulated as for the LS example (Example 4.1), namely

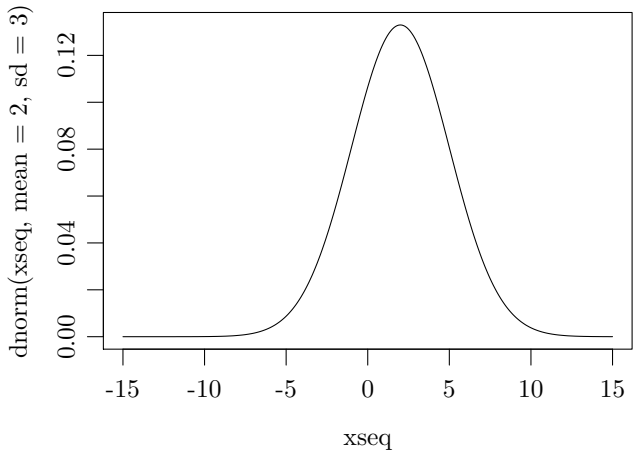
$$Y_t = \alpha X_t + \varepsilon_t \tag{4-23}$$

where ε_t is white noise, i.e. $\varepsilon_t \sim N(0, \sigma^2)$.

```
## Set the parameters for simulation The coefficient for the input
alpha <- 2.4
## The standard deviance of the white noise (square root of the variance)
sigma <- 0.4
## Number of points
N <- 100
## Generate the input as a sine
x <- sin(seq(0, 4 * pi, len = N))
## Simulate the model
y <- alpha * x + rnorm(N, 0, sigma)
## Plot
par(mfrow = c(2, 1))
plot(x, type = "l")
plot(y, type = "l")
```




```
## Now data is available, hence the likelihood can be calculated. For the
## this the pdf is needed. In R the pdf (and cdf,quantiles and random
## numbers) from many distributions can easily be obtained. In this case we
## need the normal pdf.

## As an example plot the pdf of the normal distribution with mean=2 (mu) and
## standard deviation=3 (sigma) See ?dnorm for more info
xseq <- seq(-15, 15, by = 0.1)
plot(xseq, dnorm(xseq, mean = 2, sd = 3), type = "l")
```



```
## Now carry out maximum likelihood estimation of the two parameters

## Define a function which return the nogative log-likelihood for the model.
## Note that the first argument to the function must be the vector of
## parameters
negloglik <- function(theta, x, y) {
    ## theta are the two parameters
    alpha <- theta[1]
    sigma <- theta[2]
    ## Calculate the residuals
    res <- y - alpha * x
    ## With the normal pdf calculate the negative loglikelihood
    -sum(log(dnorm(res, sd = sigma)))
}
## Use an optimizer to minimize the function.  First argument is the initial
## values of the parameters and then the function plus additional arguments
## to the function
fit <- optim(par = c(1, 1), negloglik, x = x, y = y)
## See that fit is a list and what is in it
str(fit)
## The parameter estimates are
fit$par


## Actually when the error is a white noise process, then the ML estimates
## and the LS estimates of the parameters are equal, except for the estimate
## of the standard deviance (sigma) of the error process
fitlm <- lm(y ~ 0 + x)
## See that the slope parameter (alpha, here named x) estimates are equal
summary(fitlm)
## The standard deviance are found by
sd(fitlm$residuals)
## Which is slightly different from the ML estimate, however this a marginal
## difference and can only play a role if only few samples are available
## (N<30)
```

Features of ML estimation

- Can be used to estimate parameters in any model, i.e. other distributions than normal distribution and certainly non-linear models
- Non-closed form (i.e. can be both LS and ML, but normally referred to as ML) are needed to calculate estimation for dynamic models (ARMAX and grey-box), which cannot be fitted with the closed form LS. They are recursive in nature and therefore a non-closed form must be applied.
- Due to the assymptotic properties of the curvature of the loglikelihood at the

maximum likelihood the uncertainty of the parameter estimates can be calculated.

Theorem 4.5

For ML-estimators it holds asymptotically that

$$\mathrm{Var}[\widehat{\boldsymbol{\theta}}] = \mathbf{I}^{-1}(\boldsymbol{\theta}) \tag{4-24}$$

where $\mathbf{I}(\boldsymbol{\theta})$ is the *Fisher's information matrix*

$$\mathbf{I}(\boldsymbol{\theta}) = -\mathrm{E}[\frac{\partial^2 \log L}{\partial \boldsymbol{\theta}^2}] \tag{4-25}$$

- Parameter estimates are the same as LS for general linear model and other imporant models, with only minor differences unless very little data is available (less than 30 samples).
- Many R packages exists for different types of ML estimation
- R package `MaxLik` can be used for maximum likelihood to simplify the coding

Chapter 5

Discrete time models

In this section discrete time models are presented. First the linear static models (no dynamics modelled) and then linear dynamic models.

5.1 Linear static models

The use of General Linear Models (also referred to as multiple linear regression models) are very widespread in all fields of engineering. They can be applied for estimation of the steady-state thermal performance of buildings and are here referred to as linear static models. They do not model the dynamics of the system, hence they are only valid for data measured under steady-state conditions or for data where the dynamics has been filtered out, most often done by using a sample period which is significant longer than the dynamics of the system. For most buildings this means at least one or two-day average values should be used.

Definition 5.1

The single input single output linear static (SISO) model is

$$Y_t = \omega X_t + \varepsilon_t \tag{5-1}$$

where Y_t is the output, X_t is the input and the error $\varepsilon_t \sim N(0, \sigma^2)$ is white noise.

The SISO can be extended to the multiple input single output (MISO) linear static model

$$Y_t = \omega_X X_t + \omega_Z Z_t + \ldots + \varepsilon_t \tag{5-2}$$

where as many inputs $X_t, Z_t, \ldots$ as needed can be added.

The parameters in this model can be estimated with the LS and ML techniques presented in Section 4.1 and 4.2.

5.1.1 Linear static model for estimation of building thermal performance

Based on the energy balance in steady-state conditions a linear static model is formulated, which can be applied to estimate the thermal performance from simple data consisting of only the heat load and ambient climate observations, which is the typical data obtained from smart meters.

Starting point for the model is the steady-state energy balance

$$\Phi = H_{\mathrm{tr}}(T_{\mathrm{i}} - T_{\mathrm{a}}) + gA_{\mathrm{sol}}I_{\mathrm{sol}} + H_{\mathrm{ve}}W_{\mathrm{s}}(T_{\mathrm{i}} - T_{\mathrm{a}}) \tag{5-3}$$

where the output and inputs of the model are:

- Φ the heat flow rate from the heating system (plus other sources: electrical appliances, humans, etc.) inside the building (W)
- T_{i} the indoor temperature (K)
- T_{a} the ambient temperature (K)
- I_{sol} the incoming solar radiation (W/m^2)
- W_{s} the wind speed (m/s)

the parameters the model are

- H_{tr} the transmission heat loss coefficient, which can also be considered as the UA-value (W/K)
- g the solar transmittance
- A_{sol} the effective collecting area, solar aperture (m^2)
- H_{ve} the wind induced ventilation heat loss coefficient (W/K)

The symbols and definitions are taken from the ISO 13790 standard.

The observations available from smart meters in buildings and climate stations are typically

- Φ: the heat load as flow rate from the heating system in the building (W)
- T_{a}: the ambient temperature (K)
- G: the global solar radiation (W/m^2)
- W_{s}: the wind speed (m/s)

Thus in order to estimate the parameters in the energy balance above it is rewritten into the linear static model

$$\Phi_t = \mu + \omega_{\mathrm{T_a}}\, T_t^{\mathrm{a}} + \omega_{\mathrm{G}}\, G_t + \omega_{\mathrm{H_{ve}T_i}}\, W_t^{\mathrm{s}} + \omega_{\mathrm{H_{ve}}}\, W_t^{\mathrm{s}} T_t^{\mathrm{a}} + \varepsilon_t \tag{5-4}$$

in which the coefficients which can be estimated represents

- μ: is the so called *intercept* of the linear model, it is simply a constant. It represents in this case the following constants: $H_{\text{tr}}T_t^{\text{i}}$, hence the indoor temperature is assumed to be constant. This coefficient also include all non climate dependent heat losses, e.g. heat loss to the ground, leaks, etc.
- $-\omega_{\text{T}_\text{a}}$: the transmission heat loss coefficient H_{tr}, including non-wind induced ventilation.
- ω_{G}: this is not exactly the same as the gA_{sol} value, since the coefficient is for the global radiation G, which is the horizontal solar radiation, not the incoming solar radiation to the building. Hence the coefficient represents another measure of solar absorption of the buildings
- $\omega_{\text{H}_\text{ve}\text{T}_\text{i}}$: the wind induced ventilation for $T_{\text{a}} = 0$.
- $\omega_{\text{H}_\text{ve}}$: the wind induced ventilation H_{ve}. This can also be interpreted as a modification of the transmission heat loss coefficient H_{tr} caused by the wind.

Hence it is seen that it is possible to estimate the building thermal performance parameters using simple smart meter data with no observations of the indoor temperature. However the interpretation of the coefficients, which can be estimated with a linear static model, needs some considerations.

5.2 Linear dynamic models

In this chapter linear input-output time series models are presented.

|||| Example 5.2

A simple example is given to elucidate the relation between a continuous time description of the dynamics derived from first-principles physics to a discrete time linear model.
Consider the simplest linear system formed by a single resistor and a single capacitor. The dynamics can be described by first order ordinary differential equation (ODE)

$$\frac{dT_{\text{i}}}{dt} = \frac{1}{RC}(T_{\text{a}} - T_{\text{i}}) \tag{5-5}$$

where T_t^{a} is the ambient temperature at time t, T_t^{i} is the indoor temperature, R is the thermal resistance and C is the heat capacity.
In order to form a difference equation equivalent to the differential equation the ODE is solved. This gives the indoor temperature Δt time ahead

$$T_{\text{i}}(t + \Delta t) = T_{\text{a}}(t) + e^{-\frac{\Delta t}{RC}}\left(T_{\text{i}}(t) - T_{\text{a}}(t)\right) \tag{5-6}$$

this is implies keeping the model input $T_{\text{a}}(t)$ constant in the interval between t and $t + \Delta t$.
Now the if the time-points t are equidistant, i.e. Δt is constant, then a discrete

$$T_{\text{i},t+\Delta t} = e^{-\frac{\Delta t}{RC}}T_{\text{i},t} + (1 - e^{-\frac{\Delta t}{RC}})T_{\text{a},t} \tag{5-7}$$

since $e^{-\frac{\Delta t}{RC}}$ is between 0 and 1, it can be

$$T_{\mathrm{i},t+\Delta t} = \phi_1 T_{\mathrm{i},t} + \omega_1 T_{\mathrm{a},t} \tag{5-8}$$

where ϕ_1 and ω_1 are between 0 and 1. To simplify notation set $\Delta t = 1$ and shift the time index to get

$$T_{\mathrm{i},t} = \phi_1 T_{\mathrm{i},t-1} + \omega_1 T_{\mathrm{a},t-1} \tag{5-9}$$

Finally, add a noise term an we have the ARX model

$$T_{\mathrm{i},t} = \phi_1 T_{\mathrm{i},t-1} + \omega_1 T_{\mathrm{a},t-1} + \epsilon_t \tag{5-10}$$

5.2.1 The ARMAX model

The discrete time linear dynamic models presented in this text are all of the AutoRegressive-MovingAverage with eXogenous input (ARMAX) type.

Definition 5.3 ARMAX model

The ARMAX(p, q, r) model is

$$\begin{aligned} Y_t + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} = \omega_0 X_t + \omega_1 X_{t-1} + \cdots + \omega_r X_{t-r} \\ + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \end{aligned} \tag{5-11}$$

where $\{Y_t\}$ is the output, $\{X_t\}$ is the input and $\{\varepsilon_t\}$ is white noise.

By introducing the following polynomials in B

$$\varphi(B) = 1 + \varphi_1 B + \cdots + \varphi_p B^p \tag{5-12}$$
$$\omega(B) = \omega_0 + \omega_1 B + \cdots + \omega_r B^r \tag{5-13}$$
$$\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q \tag{5-14}$$

the ARMAX model can be written as

$$\varphi(B) Y_t = \omega(B) X_t + \theta(B) \varepsilon_t \tag{5-15}$$

where

$\varphi(B) Y_t$ is the Auto-Regressive (AR) part
$\theta(B) \varepsilon_t$ is the Moving-Average (AR) part
$\varphi(B) X_t$ is the eXogenous input (X) part

5.2.2 ARX model

Thus a model

$$\varphi(B)Y_t = C(B)X_t + \varepsilon_t \tag{5-16}$$

which without the MA part is called an ARX model. This model can be formulated as a linear regression model and therefore estimated with LS techniques described in Section 4.1.

5.2.3 Transfer function

The ARMAX model can written as transfer function model

$$Y_t = \frac{\omega(B)}{\varphi(B)}X_t + \frac{\theta(B)}{\varphi(B)}\varepsilon_t \tag{5-17}$$

where the transfer function from X_t to Y_t is

$$H(B) = \frac{\omega(B)}{\varphi(B)} \tag{5-18}$$

(Note that here this is written with the backshift operator B, which implies a discretization of the system and therefore a dependence of the sampling time Δt. In continuous-time the transfer function is normally written with ω denoting the frequency).

The transfer function describes how the system "react" to the input. When the system is linear the transfer function is uniquely described by the impulse response

|||| Definition 5.4

Impulse functions (Continuous time) *Dirac's delta function* (or defined by

$$\int_{-\infty}^{\infty} f(t)\delta(t - t_0)\, dt = f(t_0) \tag{5-19}$$

(Discrete time) *Kronecker's delta sequence* (or *impulse function*) is

$$\delta_k = \begin{cases} 1 & \text{for } k = 0 \\ 0 & \text{for } k = \pm 1, \pm 2, \cdots \end{cases} \tag{5-20}$$

The following theorem is fundamental for the theory of linear dynamic systems.

Theorem 5.5

Existence of impulse response functions For a linear, time-invariant system there exists a function h such that the output is obtained as the *convolution integral*

$$y(t) = \int_{-\infty}^{\infty} h(u)x(t-u)\,du \tag{5-21}$$

in continuous time, or the *convolution sum*

$$y_t = \sum_{k=-\infty}^{\infty} h_k x_{t-k} \tag{5-22}$$

in discrete time. The weight function, h, is called the *impulse response function*, since the output of the system is $y = h$ if the input is the impulse function. Sometimes the weight function is called the *filter weights*.

Example 5.6

Deriving the UA-value as the steady-state gain from an AR(MA)X model.

Chapter 6

Continuous time models

6.1 Introduction

Often non-linearities and non-stationarities are most conveniently modelled by considering continuous time stochastic models.
By that approach information from physics is more easily used. The modelling technique described in this chapter is called *grey box modelling*, since it provides good tools for combining information from physics with information from the data.
This chapter describes in detail a *maximum likelihood method* for estimation of linear stochastic differential equations in state space form. Also the estimation of a class of non-linear differential equations is considered. The state space formulation is adequate for modelling some non-stationary time series.
Furthermore the *maximum aposteriori estimation technique* (MAP), the *prediction error method* (PEM) (or *non-linear least squares*), and the *indirect prediction error method* (IPEM) are considered. Finally, some methods for filtering the residuals along with the estimation and some methods for making the estimation robust are described.
At the end of this chapter several cases of modelling physical systems using stochastic differential equatins are presented.

6.2 The continous time linear stochastic state space model

The deterministic part of a model considered is frequently simply a system of ordinary differential equations.
In matrix notation the equations can be concatenated in *the deterministic linear state space model in continuous time*

$$\frac{d\boldsymbol{X}}{dt} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{B}\boldsymbol{U} \tag{6-1}$$

where $\boldsymbol{X}$ is the state vector and $\boldsymbol{U}$ is the input vector.
The dynamical behavior of the system is characterized by the matrix $\boldsymbol{A}$, and $\boldsymbol{B}$ is a matrix, which specify how the input signals enter the system.

To describe the deviation between (6-1) and the true variation of the states an additive noise term is introduced.
Then we have the stochastic differential equation

$$d\boldsymbol{X} = \boldsymbol{A}\boldsymbol{X}dt + \boldsymbol{B}\boldsymbol{U}dt + d\boldsymbol{w}(t) \tag{6-2}$$

where the stochastic process $\boldsymbol{w}(t)$ is assumed to be a process with independent increments. (6-2) is *the stochastic linear state space model in continuous time*.
There are many reasons for introducing a noise term:

- *Modeling approximations*. For instance the dynamics, as described by the matrix $\boldsymbol{A}$ in (6-2) might be an approximation to the true system.
- *Unrecognized and unmodeled inputs*. Some variables, which are not considered, such as e.g. wind speed, may affect the system.
- *Measurements of the input are noise corrupted*. In this cases the measured input is regarded as the actual input to the system, and the deviation from the true input is described by $\boldsymbol{w}(t)$.

In the general case we assume that only a linear combination of the states is measured, and if we introduce $\boldsymbol{Y}$ to denote the measured or recorded variables we can write

$$\boldsymbol{Y}(t) = \boldsymbol{C}\boldsymbol{X}(t) + \boldsymbol{D}\boldsymbol{U}(t) + \boldsymbol{e}(t) \tag{6-3}$$

where $\boldsymbol{C}$ is a constant matrix, which specifies which linear combination of the states that actually are measured, and $\boldsymbol{D}$ is a constant matrix, which accounts for input variables which directly affect the output.
The equation is for obvious reasons called *the measurement equation*.
The term $\boldsymbol{e}(t)$ is the

- *Measurement error*. The sensors that measure the output signals are affected by noise and drift.

In the following it is assumed that $w(t)$ and $e(t)$ are mutually uncorrelated, which seems to be quite reasonable.
As an example of a model belonging to the class (6-2) consider the following model describing the heat dynamics for a building

$$\begin{aligned}
\begin{bmatrix} dT_m \\ dT_i \end{bmatrix} &= \begin{bmatrix} \frac{-1}{r_i c_m} & \frac{1}{r_i c_m} \\ \frac{1}{r_i c_i} & -\left(\frac{1}{r_a c_i} + \frac{1}{r_i c_i}\right) \end{bmatrix} \begin{bmatrix} T_m \\ T_i \end{bmatrix} dt + \\
&\quad \begin{bmatrix} 0 & 0 & A_w p/c_m \\ 1/(r_a c_i) & 1/c_i & A_w(1-p)/c_i \end{bmatrix} \begin{bmatrix} T_a \\ \phi_h \\ \phi_s \end{bmatrix} dt + \begin{bmatrix} dw_m(t) \\ dw_i(t) \end{bmatrix} \qquad (6\text{-}4) \\
\boldsymbol{T}_r(t) &= (0\ 1)\boldsymbol{T}(t) + \boldsymbol{e}(t) \qquad (6\text{-}5)
\end{aligned}$$

The states of the model are given by the temperature T_m of a large heat accumulating medium with the heat capacity c_m, and by the temperature T_i of the room air and

possibly the inner part of the walls with the capacity c_i. The term r_i is the resistance against heat transfer between the room air and the large heat accumulating medium, while r_a is the resistance against heat transfer from the room air to the ambient air with the temperature T_a.
The input energy is supplied by the electrical heaters ϕ_h and the solar radiation which penetrates through the windows facing south $A_w\phi_s$, where A_w is the effective window area. The effective window area is the window area corrected for shade effects, and absorption and reflection by the triple glazed windows. Note that only the indoor air temperature is measured.

6.3 A maximum likelihood method for parameter estimation in continous time stochastic state space models

6.3.1 The Model

It is assumed that the model of the dynamics is described by the stochastic differential equation

$$d\boldsymbol{X} = \boldsymbol{A}\boldsymbol{X}dt + \boldsymbol{B}\boldsymbol{U}dt + d\boldsymbol{w}(t) \tag{6-6}$$

where $\boldsymbol{X}$ is m dimensional, and the m dimensional stochastic process $\boldsymbol{w}(t)$ is assumed to be a process with independent increments.
With the purpose of calculating the likelihood function, $\boldsymbol{w}(t)$ is restricted to be a Wiener-process with the incremental covariance $\boldsymbol{R}_1^c(t)$.
The measured variables are written

$$\boldsymbol{Y}(t) = \boldsymbol{C}\boldsymbol{X}(t) + \boldsymbol{D}\boldsymbol{U}(t) + \boldsymbol{e}(t) \tag{6-7}$$

The term $\boldsymbol{e}(t)$ is the measurement error. It is assumed that $\boldsymbol{e}(t)$ is normally distributed white noise with zero mean and variance $\boldsymbol{R}_2$. Furthermore, it is assumed that $\boldsymbol{w}(t)$ and $\boldsymbol{e}(t)$ are mutually independent.

6.3.2 From Continuous to Discrete Time

Since it is assumed that the system is described by the stochastic differential equation (6-6), it is possible analytically to perform an integration, which under some assumptions exactly specifies the system evolution between discrete time instants.
The discrete time model corresponding to the continuous time model (6-6) is obtained by integrating the differential equation through the sample interval $[t, t+\tau)$.
Thus the sampled version of (6-6) can be written as

$$\boldsymbol{X}(t+\tau) = e^{\boldsymbol{A}(t+\tau-t)}\boldsymbol{X}(t) + \int_t^{t+\tau} e^{\boldsymbol{A}(t+\tau-s)}\boldsymbol{B}\boldsymbol{U}(s)ds + \int_t^{t+\tau} e^{\boldsymbol{A}(t+\tau-s)}d\boldsymbol{w}(s) \tag{6-8}$$

Under the assumption that $\boldsymbol{U}(t)$ is constant in the sample interval the sampled version can be written as the following discrete time model in state space form

$$\boldsymbol{X}(t+\tau) = \boldsymbol{\phi}(\tau)\boldsymbol{X}(t) + \boldsymbol{\Gamma}(\tau)\boldsymbol{U}(t) + \boldsymbol{v}(t;\tau) \tag{6-9}$$

where

$$\boldsymbol{\phi}(\tau) = e^{\boldsymbol{A}\tau}\ ; \quad \boldsymbol{\Gamma}(\tau) = \int_0^\tau e^{\boldsymbol{A}s}\boldsymbol{B}\,ds\ ; \quad \boldsymbol{v}(t;\tau) = \int_t^{t+\tau} e^{\boldsymbol{A}(t+\tau-s)}\,d\boldsymbol{w}(s) \tag{6-10}$$

On the assumption that $\boldsymbol{w}(t)$ is a Wiener process, $\boldsymbol{v}(t;\tau)$ becomes normally distributed white noise with zero mean and covariance

$$\boldsymbol{R}_1(\tau) = E\left[\boldsymbol{v}(t;\tau)\boldsymbol{v}(t;\tau)^T\right] = \int_0^\tau \boldsymbol{\phi}(s)\boldsymbol{R}_1^c\boldsymbol{\phi}(s)^T ds \tag{6-11}$$

If the sampling time is constant (equally spaced observations), the stochastic difference equation can be written

$$\boldsymbol{X}(t+1) = \boldsymbol{\phi}\boldsymbol{X}(t) + \boldsymbol{\Gamma}\boldsymbol{U}(t) + \boldsymbol{v}(t) \tag{6-12}$$

where the time scale now is transformed such that the sampling time becomes equal to one time unit.
The problem of calculating the exponential of a matrix is a major task - see [Madsen and Melgaard, 19

6.3.3 Maximum Likelihood Estimates

In the following it is assumed that the observations are obtained at regularly spaced time intervals, and hence that the time index t belongs to the set $\{0, 1, 2, ..., N\}$. N is the number of observations.
We further introduce

$$\mathcal{Y}(t) = [\boldsymbol{Y}(t), \boldsymbol{Y}(t-1), \ldots, \boldsymbol{Y}(1), \boldsymbol{Y}(0)]^T \tag{6-13}$$

i.e. $\mathcal{Y}(t)$ is a matrix containing all the observations up to and including time t. Finally, let $\boldsymbol{\theta}$ denote a vector of all the unknown parameters of the *embedded continous time model* in (6-6) – (6-7) - including the unknown variance and covariance parameters in $\boldsymbol{R}_1^c$ and $\boldsymbol{R}_2$.
The likelihood function is the joint probability density of all the observations assuming that the parameters are known, i.e.

$$\begin{aligned} L^T(\boldsymbol{\theta};\mathcal{Y}(N)) &= p(\mathcal{Y}(N)|\boldsymbol{\theta}) \\ &= p(\boldsymbol{Y}(N)|\mathcal{Y}(N-1),\boldsymbol{\theta})p(\mathcal{Y}(N-1)|\boldsymbol{\theta}) \\ &= \left(\prod_{t=1}^{N} p(\boldsymbol{Y}(t)|\mathcal{Y}(t-1),\boldsymbol{\theta})\right)p(\boldsymbol{Y}(0)|\boldsymbol{\theta}) \end{aligned} \tag{6-14}$$

where successive applications of the rule $P(A \cap B) = P(A|B)P(B)$ is used to express the likelihood function as a product of conditional densities.

Since both $\boldsymbol{v}(t)$ and $\boldsymbol{e}(t)$ are normally distributed the conditional density is also normal. The normal distribution is completely characterized by the mean and the variance. Hence, in order to parameterize the conditional distribution, we introduce the conditional mean and the conditional variance as

$$\hat{\boldsymbol{Y}}(t|t-1) = E[\boldsymbol{Y}(t)|\mathcal{Y}(t-1),\boldsymbol{\theta}] \quad \text{and} \tag{6-15}$$

$$\boldsymbol{R}(t|t-1) = V[\boldsymbol{Y}(t)|\mathcal{Y}(t-1),\boldsymbol{\theta}] \tag{6-16}$$

respectively. It is noticed that (6-15) is the one-step prediction and (6-16) the associated variance. Furthermore, it is convenient to introduce the one-step prediction error (or innovation)

$$\boldsymbol{\epsilon}(t) = \boldsymbol{Y}(t) - \hat{\boldsymbol{Y}}(t|t-1) \tag{6-17}$$

Using (6-14)–(6-17) the conditional likelihood function (conditioned on $\boldsymbol{Y}(0)$) becomes

$$L(\boldsymbol{\theta};\mathcal{Y}(N)) = \prod_{t=1}^{N}\left((2\pi)^{-m/2}\det\boldsymbol{R}(t|t-1)^{-1/2}\exp(-\tfrac{1}{2}\boldsymbol{\epsilon}(t)^T\boldsymbol{R}(t|t-1)^{-1}\boldsymbol{\epsilon}(t))\right) \tag{6-18}$$

where m is the dimension of the $\boldsymbol{Y}$ vector. Most frequently the logarithm of the conditional likelihood function is considered. It is written

$$\log L(\boldsymbol{\theta};\mathcal{Y}(N)) = -\tfrac{1}{2}\sum_{t=1}^{N}\left(\log\ \det\boldsymbol{R}(t|t-1) + \boldsymbol{\epsilon}(t)^T\boldsymbol{R}(t|t-1)^{-1}\boldsymbol{\epsilon}(t)\right) + \text{const} \tag{6-19}$$

The conditional mean $\hat{\boldsymbol{Y}}(t|t-1)$ and the conditional variance $\boldsymbol{R}(t|t-1)$ can be calculated recursively by using *a Kalman filter*.
The maximum-likelihood estimate (ML-estimate) is the set $\hat{\boldsymbol{\theta}}$, which maximizes the likelihood function. Since it is not possible analytically to optimize the likelihood function, a numerical method has to be used. A reasonable method is the quasi-Newton method.
An estimate of the uncertainty of the parameters is obtained by the fact that the ML-estimator is asymptotically normally distributed with mean $\boldsymbol{\theta}$ and variance

$$\boldsymbol{D} = \boldsymbol{H}^{-1} \tag{6-20}$$

where the matrix $\boldsymbol{H}$ is given by

$$\{h_{lk}\} = -E\left[\frac{\partial^2}{\partial\theta_l\partial\theta_k}\log L(\boldsymbol{\theta};\mathcal{Y}(N))\right] \tag{6-21}$$

An estimate of $\boldsymbol{D}$ is obtained by equating the observed value with its expectation and applying

$$\{h_{lk}\} \approx -\left(\frac{\partial^2}{\partial\theta_l\partial\theta_k}\log L(\boldsymbol{\theta};\mathcal{Y}(N))\right)\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \tag{6-22}$$

The above equation is thus used for estimating the variance of the parameter estimates. The variances serves as a basis for calculating t-test values for test under the hypotesis that the parameter is equal to zero. The correlation between the estimates is readily found based on the variance matrix.

6.4 A linear continuous time model for heat dynamics of a building

The thermal characteristic of buildings is frequently approximated by a simple network with resistors and capacitances, see for instance [Hammersten et al., 1988] or [Madsen and Holst, 1995]. In this section, such a (simplified) lumped parameter model for the heat dynamics of a simple building, called a test cell is presented. The dominating heat capacity of the test cell is located in the outer wall. For such buildings, the model with two time constants shown in Figure 6.1 is frequently found adequate.

The states of the model are given by the temperature, T_i, of the indoor air and possibly inner part of the walls with heat capacity C_i, and by the temperature, T_m, of the heat accumulating medium, with the heat capacity C_m. H_i is the transmittance of the heat transfer between the room air and the walls, while H_m is the heat transmittance between the inner part of the walls and the external surface of the walls. The input to the system is the heat supply, Q_h, and the outdoor surface temperature, T_e. By considering the outdoor *surface* temperature instead of the outdoor *air* temperature as the input, the effect of solar radiation is automatically taken into account.
In state space form the model is written,

$$\begin{bmatrix} dT_i \\ dT_m \end{bmatrix} = \begin{bmatrix} -H_i/C_i & H_i/C_i \\ H_i/C_m & -(H_i+H_m)/C_m \end{bmatrix} \begin{bmatrix} T_i \\ T_m \end{bmatrix} dt + \begin{bmatrix} 0 & 1/C_i \\ H_m/C_m & 0 \end{bmatrix} \begin{bmatrix} T_e \\ Q_h \end{bmatrix} dt + \begin{bmatrix} dw_i(t) \\ dw_m(t) \end{bmatrix} \tag{6-23}$$

An additive noise term is introduced to describe deviations between the model and the true system. The term can also be considered as noise on the input signals.

Hence, the model of the heat dynamics is given by the (matrix) stochastic differential equation

$$d\boldsymbol{T} = \boldsymbol{A}\boldsymbol{T}dt + \boldsymbol{B}\boldsymbol{U}dt + d\boldsymbol{w}(t)\ , \tag{6-24}$$

where $\boldsymbol{w}(t)$ is assumed to be a Wiener process with incremental covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{1,i}^2 & 0 \\ 0 & \sigma_{1,m}^2 \end{bmatrix}. \tag{6-25}$$

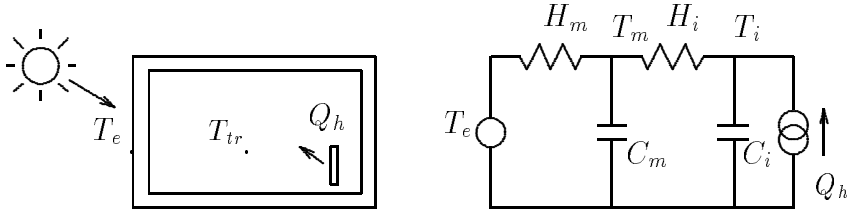


Figure 6.1: A model with two time constants of the test building and the equivalent electrical network.

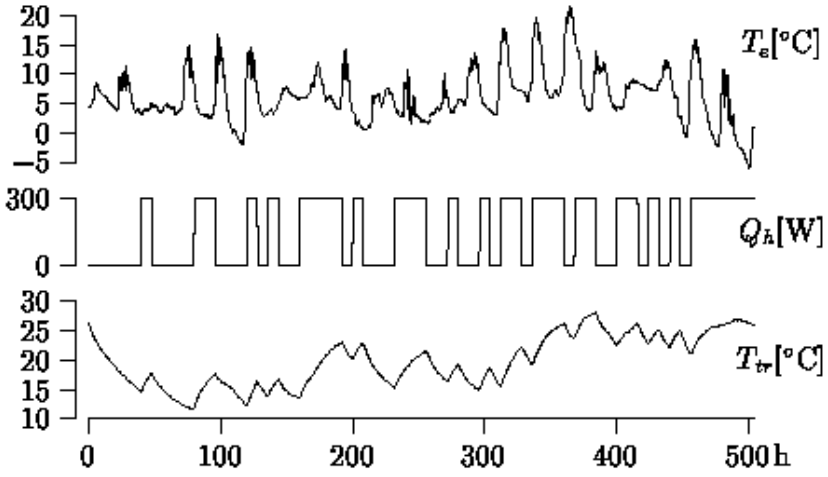


Figure 6.2: The signals in the simulations: the outdoor surface temperature, the controlled heat power and the indoor air temperature.

The measured air temperature is naturally encumbered with some measurement errors, and hence the measurement equation is written

$$T_{tr}(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} T_i \\ T_m \end{bmatrix} + e(t), \tag{6-26}$$

where $e(t)$ is the measurement error, assumed to be normally distributed with zero mean and variance σ_2^2.
The following values of the parameters have been estimated in an earlier experiment on a test cell: $H_i = 55.29\,\mathrm{W/K}$, $H_m = 13.86\,\mathrm{W/K}$, $C_i = 325.0\,\mathrm{Wh/K}$, $C_m = 387.8\,\mathrm{Wh/K}$, $\sigma_{1,i}^2 = 0.00167\,\mathrm{K}^2$, $\sigma_{1,m}^2 = 0.00978\,\mathrm{K}^2$, and $\sigma_2^2 = 0.00019\,\mathrm{K}^2$. Corresponding to these parameters, the time constants of the system are $\tau_1 = 3.03$ hours and $\tau_2 = 54.28$ hours.
Using these parameters for the model, we are able to simulate the sample paths of the system. By simulating several sample paths from the system and estimating the parameters of the model from each sample path one can validate the estimation procedure. By considering several simulated sequences this investigation considers both the mean values and the variances of the estimated parameters. For the simulation study the following input and output signals of the model have been used: T_e is measured surface temperature, from a Danish test building, Q_h the heat supply, is a PRBS (pseudo-random binary sequence) with the number of stages, $n = 6$ and smallest switching interval T_{prbs} = 8 hours (see [Godfrey, 1980]), switching between 0 W and 300 W, and T_{tr} is the indoor room temperature, simulated from the specified model.
In this study the sampling time is fixed to $T_{sampl} = 20$ minutes, and the length of each experiment is 21 days, which is equal to 1512 observations per simulated series. We have simulated 50 equal series, but with different realizations of the noise sequences.
In table 6.1 and Table 6.2 the results are summarized using CTSM [Melgaard and Madsen, 1993],

Parameter	x_{sim}	$\bar{x}$	s_x^2	$\bar{s^2}$	F-stat.	$\lvert t\rvert$-stat.
H_i [W/K]	55.290	55.279	2.1645	2.3772	0.9105	0.053
H_m [W/K]	13.860	13.867	0.0360	0.0390	0.9244	0.261
C_i [Wh/K]	325.00	325.0	9.5667	10.746	0.8902	0.011
C_m [Wh/K]	387.78	387.33	162.26	159.89	1.0148	0.250
$\sigma_{1,i}^2$ [K^2]	1.670e-3	1.650e-3	2.185e-8	2.475e-8	0.8825	0.975
$\sigma_{1,m}^2$ [K^2]	9.780e-3	9.554e-3	8.650e-7	1.109e-6	0.7803	1.716
σ_2^2 [K^2]	1.900e-4	1.912e-4	7.956e-10	7.446e-10	1.0686	0.308
UA [W/K]	11.082	11.084	0.0112	0.0111	1.0114	0.125
CI [MJ/K]	2.2862	2.2846	1.4553e-3	1.3815e-3	1.0535	0.304

Table 6.1: Results from estimation of $n_e = 50$ series, using 1-step predictions in the criterion. The mean autocorrelation of residuals is, $\bar{\rho}(1) = -0.001$.

which is a tool for maximum likelihood estimation of multivariate stochastic differential equations with a linear or non-linear state space formulation.
In the tables results are summarized for the parameters of the model and for two physical characteristics, which can be calculated as functions of the model parameters. The UA-value is the overall heat transmittance and CI is the internal heat capacity of the building. For all parameters, the mean of the estimated values are given, which can be compared to the simulated values. Also the empirical variance and the mean of the estimated variances of the parameters are given. A comparison of these values will indicate if the method is able to estimate the right uncertainty of the parameters.
In table 6.1 the results are obtained using 1-step predictions in the criterion for maximum likelihood. When estimating the true model, which is the case here, using 1-step predictions is the optimal choice [Stoica and Nehorai, 1989].
On the other hand if the true model is not contained in the model set, it can be advantageous to filter the residuals before estimating the parameters. One way of doing this is to use a k-step prediction-horizon in the criterion for some $k > 1$. This kind of filtering works like a low-pass filter and will put more weight on the low frequency part of the dynamics. In the considered case the physical characteristic parameters UA- and CI-values represent low frequency dynamics. The columns of the tables are: x_{sim}, the simulated values, $\bar{x}$, the mean of the estimated values, s_x^2 is the empirical variance of the estimated parameters, $\bar{s^2}$ is the mean of the estimated variance of the parameters, F-stat. is an F statistic given by $Z_F = s_x^2/\bar{s^2}$ and $|t|$-stat. is a t statistic given by $Z_t = |\bar{x} - x_{sim}|/(s_x\sqrt{n_e})$.
In order to verify if the variance of the parameters provided by the estimation tool is equal to the empirical variance, one wish to test the hypotheses

$$H_0 : ,$$
$$H_1 : ,$$

Under H_0 we have in this case that $Z_F \sim F(49, \infty)$. The critical set for this test is $\{z < F(49,\infty)_{\alpha/2} \vee z > F(49,\infty)_{1-\alpha/2}\}$ on level α. By choosing $\alpha = 0.1$ we obtain the critical set $\{z < 0.74 \vee z > 1.35\}$. It is seen from Table 6.1 that we cannot reject H_0 for

Parameter	x_{sim}	$\bar{x}$	s_x^2	$\bar{s^2}$	F-stat.	$\lvert t\rvert$-stat.
H_i [W/K]	55.290	55.339	2.2259	0.6154	3.617	0.232
H_m [W/K]	13.860	13.863	0.0383	0.0099	3.877	0.109
C_i [Wh/K]	325.00	324.8	9.486	3.3338	2.842	0.481
C_m [Wh/K]	387.78	387.20	163.74	47.685	3.4338	0.3210
$\sigma_{1,i}^2$ [K^2]	1.670e-3	1.616e-3	4.158e-8	1.620e-8	2.5665	1.880
$\sigma_{1,m}^2$ [K^2]	9.780e-3	9.562e-3	8.680e-7	4.156e-7	2.0884	1.656
σ_2^2 [K^2]	1.900e-4	2.072e-4	3.700e-9	4.145e-9	0.8926	1.997
UA [W/K]	11.082	11.084	0.0115	0.0028	4.1290	0.107
CI [MJ/K]	2.2862	2.2837	1.469e-3	3.954e-4	3.7159	0.460

Table 6.2: Results from estimation of $n_e = 50$ series, using 4-step predictions in the criterion. The mean autocorrelation of residuals is, $\bar{\rho}(1) = 0.712$.

any parameter on the chosen level.
Another test is performed in order to verify that the estimated parameters are not biased. The following hypotheses are tested

$$H_0 : ,$$
$$H_1 : ,$$

Under H_0 the distribution of the test statistic is $Z_t \sim t(49)$. The critical set for this test is $\{z > t(49)_{1-\alpha/2}\}$ on level α. For $\alpha = 0.1$, the critical set is $\{z > 2.0\}$, thus from Table 6.1 we cannot reject H_0 for any of the parameters on the chosen level.
In table 6.2 the results from estimation using a 4-step prediction-horizon in the criterion are shown. In this ideal case where the true model is contained by the model set there is no difference in the estimated parameters which is also seen from the tables, they are still unbiased. A problem, though, when "filtering" the residuals in this way is that the residuals are no longer white noise. From the tables we have that the mean autocorrelation of residuals are $\bar{\rho}(1) = 0.712$ when using the criterion based on 4-step predictions against $\bar{\rho}(1) = -0.001$ for 1-step predictions. When the residuals are autocorrelated a number of statistical tests for model validation are no longer valid. Another problem in table 6.2 is that the uncertainties of the parameters are underestimated when the autocorrelation of residuals is not taken into account in the calculation of the uncertainties. Wahlberg ([Wahlberg and Ljung, 1986]) has shown, that asymptotically, it is only the prediction horizon itself, that affects the weighting in the frequency domain, and not how it is split up into sampling interval times number of predicted sampling instants. This means that, asymptotically, we could obtain the same results as using a k-step prediction, by using a proper anti-aliasing filter followed by a new sampling of the data and then estimate with a one-step prediction criterion. In this way one could avoid the autocorrelated residuals.

Chapter 7

Exercises

In this chapter contains the exercises which are to be completed before and during the summer school.

7.1 Linear regression models for the heat load of buildings

The chapter contains an exercise which deals with linear regression models for the heat load of buildings applied for estimating building thermal performance parameters.

This exercise is a part of the preparation to the summer school and it is should be completed before the beginning of the summer school and a short report handed in through CampusNet `www.campusnet.dtu.dk`.

The modeling is based on two days averaged values, such that the heat dynamics of the building can be disregarded and a linear static model can be applied. Data from three houses is available. The task is to make a model for the heat load of each house with climate inputs. The buildings are typical single-family houses constructed in bricks in the fifties and sixties. The ground floor area of the buildings is 140, 122, and 127 m^2, respectively. The houses are located in the area around Borgmester Andersens vej, Sønderborg (search Google maps to have a look). The climate measurements are recorded at a weather station located at the local district heating plant.

Unzip the `exLinearRegressionFiles.zip`. The available data for the three houses is in the file `soenderborg1hour.csv` and consists of hourly values of the inputs and output. The indoor temperature was not been measured. The provided script in `exLinearRegression.R` can be used for the modeling. See the introduction to R in Chapter 2, it is *important to set the working directory* (see page 11).

Questions

First read Section 5.1 on static linear models. The most simple feasible static linear model, in which only the heat transfer due to the temperature difference is included, is

$$\Phi_t = \omega_{\Delta t} \left(T_t^{\mathrm{i}} - T_t^{\mathrm{a}} \right) + \varepsilon_t \tag{7-1}$$

where the error is normal distributed white noise $\varepsilon_t \sim N(0, \sigma^2)$, hence i.i.d. (independent and identically normal distributed).

Since the indoor temperature is not available the model needs to be reformulated. By assuming a constant indoor temperature a model with which the thermal performance parameters can be estimated (an identifiable model) is

$$\Phi_t = \mu - \omega_{\mathrm{T_a}} \, T_t^{\mathrm{a}} + \varepsilon_t \tag{7-2}$$

where μ is called the intercept.
The parameters in the model are

$$\theta = (\mu, \omega_{\mathrm{T_a}}, \sigma) \tag{7-3}$$

and the predicted heat load is

$$\widehat{\Phi}_t = \hat{\mu} - \hat{\omega}_{\mathrm{T_a}} \, T_t^{\mathrm{a}} \tag{7-4}$$

hence the residuals are

$$\hat{\varepsilon}_t = \Phi_t - \widehat{\Phi}_t \tag{7-5}$$

and the least squares estimates of the parameters are

$$\hat{\theta} = \min_{\theta} \sum_{t=1}^{N} \varepsilon_t^2 \tag{7-6}$$

The task is now to build a suitable linear regression model and carry out an analysis separate for each house by considering the questions in the following. Use the provided R script to carry out the analysis for House 1 and then copy the script for the analysis for the other houses.

Question 1

Consider the following

- Estimate the transmission heat loss coefficient with the simple model and the indoor temperature for House 1 (You have to derive the indoor temperature yourself)
- Consider the plot of the estimated function between the ambient temperature and heat load and discuss how the estimates can be directly seen from the plot
- Shortly discuss the estimates values. Are they realistic?

Question 2

Select a suitable model for the house. This should be carried out both from a physical and a statistical perspective. Basically two approaches can be taken here: either forward or backward model selection. In a forward model selection approach the simple model is expanded by adding one input at a time. In a backward selection approach the full model with all inputs are fitted and the least significant input is removed until only significant inputs are remaining.
Try a forward selection by adding inputs one by one to the simple model and keep only the significant inputs.
In the `summary(fit)` the coefficient of performance R^2 can be seen. It is a fundamental statistical measure of the performance of a model calculated by

$$R^2 = 1 - \frac{\sum_{t=1}^{N} \hat{\varepsilon}_t^2}{\sum_{t=1}^{N} (\Phi_t - \bar{\Phi})^2} \tag{7-7}$$

where $\bar{\Phi}$ is the average heat load, i.e.

$$\bar{\Phi} = \frac{1}{N} \sum_{t=1}^{N} \Phi_t \tag{7-8}$$

So the R^2 is (roughly said): how much better is the considered model compared to a model using only the average, or how much of the variation in the data is explained by the model. Clearly, R^2 have to increase as more inputs are added to the model. Try to watch this as inputs are added to the model. Furthermore the `"Adjusted R-squared value"` should be considered, it is the R^2 value adjusted for the number of parameters in the model, see `?summary.lm` for details.

The model can also be improved by considering prior physical knowledge. One example, which might yield an improvement, is to approximate the global radiation on a vertical surface by the simple projection

$$G_t^{\text{ver}} \approx \frac{G_t}{\sin(\theta_t^{\text{s}})} \tag{7-9}$$

where θ_{s} is the solar elevation angle. Now this can be used as an input to the model instead of G. This is just a suggestion for further improved based on prior physical knowledge.

Discuss the findings, as for example:

- The uncertainty of the parameter estimates, considering that a 95% confidence band on the parameter estimates are approximately plus-minus two times the standard error of the estimates (i.e. ”Std.Error” in summary(fit)).
- Give some considerations when the different parts are added to the model. Do the added terms improve the model fit?

- Are they significant and are they in accordance with the assumed physics? (Influence of changes to the building, heating system, correlation of the inputs...)
- What happens with the addition of the wind speed (and $W_s T_a$ interaction) to the model? What influence does this have on the interpretation of the estimated heat loss transmission coefficient ω_{T_a}?
- Discuss the suggested transformation to vertical solar radiation, does is provide an improvement? Optionally, suggestions for other model improvements can be tried and discussed.

Question 3

Validate the model by checking the assumption a normal white noise error process, this should be carried out with statistical tests and plots of the residuals, see Section 3.6 for an short overview of model validation.
Check the following:

- Plot of the residuals vs. inputs, here T_t^a. The residuals should show any remaining dependence on the inputs, which potentially could indicate non-linearities.
- Time series plot of the residuals should show identical distributed residuals independent of the time
- If there is a significant auto-correlation, then the residuals are not white noise and therefore not i.i.d. distributed. The blue lines on the ACF plot indicate a 95% significance band, such that in case the residuals are white noise then 5% of the bars should be (slightly) above these lines.
- The histogram of the residuals should look like the bell formed normal distribution and should not be too skewed to one side.
- A Q-Q plot of the residuals. This is a plot of the quantiles of two distributions versus each other. It is used to check if the distributions are very different from each other. The quantile for each residual is estimated and plotted versus the value of the same quantile for the normal distribution. If the distributions are equivalent the points are on a straight line.

Question 4

Use now a linear static model on hourly time series. Fit a model the selected model for the hourly data. Compare with the parameter estimates with the ones for the two-days values and comment on:

- The values of the parameter estimates. Are they reasonable considering the parameter estimates (also the uncertainty) based on two-days values?
- The estimated parameter uncertainties. How do they compare to the estimated parameter uncertainties based on two-days values?

Finally, carry out a model validation by analyzing the residuals for the assumptions of white noise. Can the parameter estimates and uncertainties based on hourly values be trusted?

There can be several reasons why the residuals are auto-correlated and why there is a need for applying more advanced models. In particularly, the heat dynamics of the building are very important to take into account in the models. The focus of the summer school will be on using dynamic models. See you there for much more modeling :-)

Questions and comments: `pbac@dtu.dk`

Bibliography

[Andersen et al., 2000] Andersen, K., Madsen, H., and Hansen, L. (2000). Modelling the heat dynamics of a building using stochastic differential equations. *Energy and Buildings*, 31:13–34.

[Bacher and Madsen, 2011] Bacher, P. and Madsen, H. (2011). Identifying suitable models for the heat dynamics of buildings. *Energy & Buildings*, 43(7):1511–1522.

[Bloem, 2007] Bloem, J. (2007). Dynamic analysis methods and modelling. application to energy performance assessment. In *2nd PALENC Conference*, pages 132–136, Grete Island.

[Bloem, 1994] Bloem, J. J., editor (1994). *System Identification Applied to Building Performance Data*. CEC-EUR 15885 EN.

[Bohlin and Isaksson, 2003] Bohlin, T. and Isaksson, A. (2003). Grey-box model calibrator and validator. In Publications, I., editor, *Proceedings of the IFAC System Identification*, pages 1477–1482, Rotterdam. IFAC.

[Cools and Gicquel, 1989] Cools, C. and Gicquel, R. (1989). The passys project, final reports phase 1 1986-89. Technical report, Office of the Official Publications of EU, Luxembourg.

[Diez et al., 2012] Diez, D. M., Barr, C. D., and Cetinkaya-Rundel, M. (2012). *OpenIntro statistics*. CreateSpace independent publishing platform.

[Godfrey, 1980] Godfrey, K. (1980). Correlation methods. *Automatica*, 16:527–534.

[Hammersten et al., 1988] Hammersten, S., van Hattem, D., Bloem, H., and Colombo, R. (1988). Passive solar component testing with identification methods. *Solar Energy*, 41(1):5–13.

[Hastie et al., 2009] Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., and Tibshirani, R. (2009). *The elements of statistical learning*, volume 2. Springer.

[Jiménez and Madsen, 2008] Jiménez, M. and Madsen, H. (2008). Models for describing the thermal characteristics of building components. *Building and Environment*, 43:152–162.

[Jiménez et al., 2008a] Jiménez, M., Madsen, H., and Andersen, K. (2008a). Identification of the main thermal characteristics of building components using matlab. *Building and Energy*, 43:170–180.

[Jiménez et al., 2008b] Jiménez, M., Madsen, H., Bloem, H., and Dammann, B. (2008b). Estimation of nonlinear continuous time models for the heat dynamics of building integrated photovoltaic modules. *Energy and Buildings*, 40:157–167.

[Kristensen and Madsen, 2003] Kristensen, N. R. and Madsen, H. (2003). Continuous time stochastic modelling, CTSM 2.3 - mathematics guide. Technical report, DTU.

[Kristensen et al., 2004] Kristensen, N. R., Madsen, H., and Jørgensen, S. B. (2004). Parameter estimation in stochastic grey-box models. *Automatica*, 40(2):225–237.

[Ljung, 1987] Ljung, L. (1987). *System Identification: Theory for the User*. Prentice-Hall, New York.

[Madsen, 2007] Madsen, H. (2007). *Time Series Analysis*. Chapman & Hall/CRC.

[Madsen, 2008] Madsen, H. (2008). *Time Series Analysis*. Chapman and Hall.

[Madsen and Holst, 1995] Madsen, H. and Holst, J. (1995). Estimation of continuous-time models for the heat dynamics of a building. *Energy and Buildings*, 22(1):67–79.

[Madsen and Melgaard, 1991] Madsen, H. and Melgaard, H. (1991). The mathematical and numerical methods used in ctlsm - a program for ml-estimation in stochastic, continuous time dynamical models. Technical Report 7, IMSOR, DTU, 2800 Lyngby.

[Madsen and Thyregod, 2010] Madsen, H. and Thyregod, P. (2010). *Introduction to General and Generalized Linear Models*. CRC Press.

[Melgaard and Madsen, 1993] Melgaard, H. and Madsen, H. (1993). Ctlsm – continuous time linear stochastic modeling (version 2.6). Technical Report 1, IMSOR, DTU, 2800 Lyngby.

[Nielsen et al., 2010] Nielsen, H. A., Mortensen, S., Bacher, P., and Madsen, H. (2010). Analysis of energy consumption in single family houses. In *Dynamic Methods for Building Energy Assessment*. DYNASTEE, Brussels.

[Pawitan, 2001] Pawitan, Y. (2001). *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford University Press.

[Stoica and Nehorai, 1989] Stoica, P. and Nehorai, A. (1989). On multistep prediction error methods for time series models. *Journal of Forecasting*, 8:357–368.

[van Dijk and Téllez, 1995] van Dijk, H. and Téllez, F. (1995). Measurement and data analysis procedures, compass final report. Technical report, EU Brussels.

[Wahlberg and Ljung, 1986] Wahlberg, B. and Ljung, L. (1986). Design variables for bias distribution in transfer function estimation. *IEEE Transactions on Automatic Control*, 31(2):134–144.

[Wasserman, 2004] Wasserman, L. (2004). *All of statistics: a concise course in statistical inference*. Springer.