Time Series Analysis

Book · October 2008

DOI: 10.1201/9781420059687

1 author:

Henrik Madsen
Technical University of Denmark

Preface

The aim of this book is to give an introduction to time series analysis. The emphasis is on methods for modeling of linear stochastic systems. Both time domain and frequency domain descriptions will be given; however, emphasis is on the time domain description. Due to the highly different mathematical approaches needed for linear and non-linear systems, it is instructive to deal with them in seperate textbooks, which is why non-linear time series analysis is not a topic in this book—instead the reader is referred to **Madsen95b**

Theorems are used to emphasize the most important results. Proofs are given only when they clarify the results. Small problems are included at the end of most chapters, and a separate chapter with real-life problems is included as the final chapter of the book. This also serves as a demonstration of the many possible applications of time series analysis in areas such as physics, engineering, and econometrics.

During the sequence of chapters, more advanced stochastic models are gradually introduced; with this approach, the family of linear time series models and methods is put into a clear relationship. Following an initial chapter covering static models and methods such as the use of the general linear model for time series data, the rest of the book is devoted to stochastic dynamic models which are mostly formulated as difference equations, as in the famous ARMA or vector ARMA processes. It will be obvious to the reader of this book that even knowing how to solve difference equations becomes important for understanding the behavior of important aspects such as the autocovariance functions and the nature of the optimal predictions.

The important concept of time-varying systems is dealt with using a state space approach and the Kalman filter. However, the strength of also using adaptive estimation methods for on-line forecasting and control is often not adequately recognized. For instance, in finance the classical methods for forecasting are often not very useful, but, by using adaptive techniques, interesting results are often obtained.

The last chapter of this book is devoted to problems inspired by real life. Solutions to the problems are found at `http://www.imm.dtu.dk/~hm/time.series.analysis`. This home page also contains additional exercises, called assignments, intended for being solved using a computer with dedicated software for time series analysis.

I am grateful to all who have contributed with useful comments and suggestions for improvement. Especially, I would like to thank my colleagues Jan Holst, Henrik Spliid, Leif Mejlbro, Niels Kjølstad Poulsen, and Henrik Aalborg Nielsen for their valuable comments and suggestions. Furthermore, I would like to thank former students Morten Høier Olsen, Rasmus Tamstorf, and Jan Nygaard Nielsen for their great effort in proofreading and improving the first manuscript in Danish. For this 2007 edition in English, I would like to thank Devon Yates, Stig Mortensen, and Fannar Örn Thordarson for proofreading and their very useful suggestions. In particular, I am grateful to Anna Helga Jónsdóttir for her assistance with figures and examples. Finally, I would like to thank Morten Høgholm for both proofreading and for proposing and creating a new layout in LaTeX.

Lyngby, Denmark *Henrik Madsen*

Notation

All vectors are column vectors. Vectors and matrices are emphasized using a bold font. Lowercase letters are used for vectors and uppercase letters are used for matrices. Transposing is denoted with the upper index ^{T}.

Random variables are always written using uppercase letters. Thus, it is not possible to distinguish between a multivariate random variable (random vector) and a matrix. However, random variables are assigned to letters from the last part of the alphabet (X, Y, Z, U, V, ...), while deterministic terms are assigned to letters from the first part of the alphabet (a, b, c, d, ...). Thus, it should be possible to distinguish between a matrix and a random vector.

CHAPTER 1

Introduction

Time series analysis deals with statistical methods for analyzing and modeling an ordered sequence of observations. This modeling results in a stochastic process model for the system which generated the data. The ordering of observations is most often, but not always, through time, particularly in terms of equally spaced time intervals. In some applied literature, time series are often called signals. In more theoretical literature a time series is just an observed or measured realization of a stochastic process.

This book on time series analysis focuses on modeling using linear models. During the sequence of chapters more and more advanced models for dynamic systems are introduced; by this approach the family of linear time series models and methods are placed in a structured relationship. In a subsequent book, non-linear time series models will be considered.

At the same time the book intends to provide the reader with an understanding of the mathematical and statistical background for time series analysis and modeling. In general the theory in this book is kept in a second order theory framework, focussing on the second order characteristics of the persistence in time as measured by the autocovariance and autocorrelation functions.

The separation of linear and non-linear time series analysis into two books facilitates a clear demonstration of the highly different mathematical approaches that are needed in each of these two cases. In linear time series analysis some of the most important approaches are linked to the fact that superposition is valid, and that classical frequency domain approaches are directly usable. For non-linear time series superposition is not valid and frequency domain approaches are in general not very useful.

The book can be seen as a text for graduates in engineering or science departments, but also for statisticians who want to understand the link between models and methods for linear dynamical systems and linear stochastic processes. The intention of the approach taken in this book is to bridge the gap between scientists or engineers, who often have a good understanding of methods for describing dynamical systems, and statisticians, who have a good understanding of statistical theory such as likelihood-based approaches.

In classical statistical analysis 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 book 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, as demonstrated in various chapters of this book.

There are a number of reasons for studying time series. These include a characterization of time series (or signals), understanding and modeling the data generating system, forecasting of future values, and optimal control of a system.

In the rest of this chapter we will first consider some typical time series and briefly mention the reasons for studying them and the methods to use in each case. Then some of the important methodologies and models are introduced with the help of an example where we wish to predict the monthly wheat prices. Finally the contents of the book is outlined while focusing on the model structures and their basic relations.

1.1 Examples of time series

In this section we will show examples of time series, and at the same time indicate possible applications of time series analysis. The examples contain both typical examples from economic studies and more technical applications.

1.1.1 Dollar to Euro exchange rate

The first example is the daily US dollar to Euro interbank exchange rate shown in Figure 1.1. This is a typical economic time series where time series analysis could be used to formulate a model for forecasting future values of the exchange rate. The analysis of such a problem relates to the models and methods described in Chapters 3, 5, and 6.

1.1.2 Number of monthly airline passengers

Next we consider the number of monthly airline passengers in the US shown in Figure 1.2. For this series a clear annual variation is seen. Again it might be useful to construct a model for making forecasts of the future number of airline passengers. Models and methods for analyzing time series with seasonal variation are described in Chapters 3, 5, and 6.

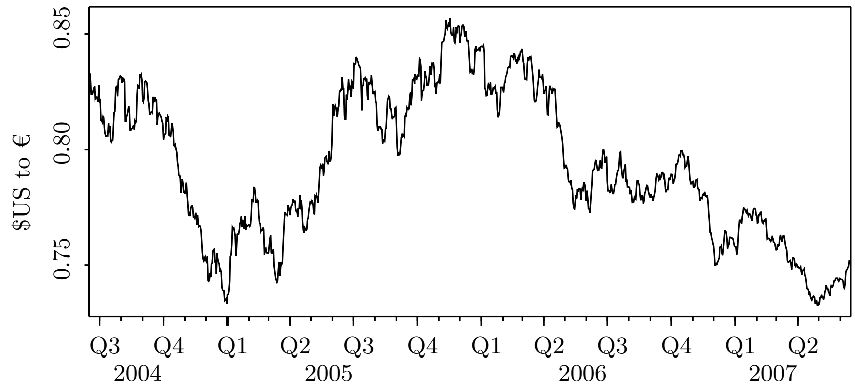

Figure 1.1: *Daily US dollar to Euro interbank exchange rate.*

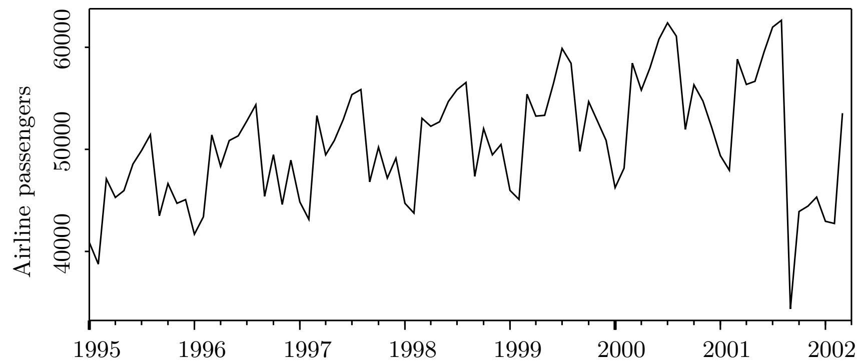

Figure 1.2: *Number of monthly airline passengers in the US. A clear annual variation can be seen in the series.*

1.1.3 Heat dynamics of a building

Now let us consider a more technical example. Figure 1.3 on the following page shows measurements from an unoccupied test building. The data on the lower plot show the indoor air temperature, while on the upper plot the ambient air temperature, the heat supply, and the solar radiation are shown.

For this example it might be interesting to characterize the thermal behavior of the building. As a part of that the so-called resistance against heat flux from inside to outside can be estimated. The resistance characterizes the insulation of the building. It might also be useful to establish a dynamic model for the building and to estimate the time constants. Knowledge of the time constants can be used for designing optimal controllers for the heat supply.

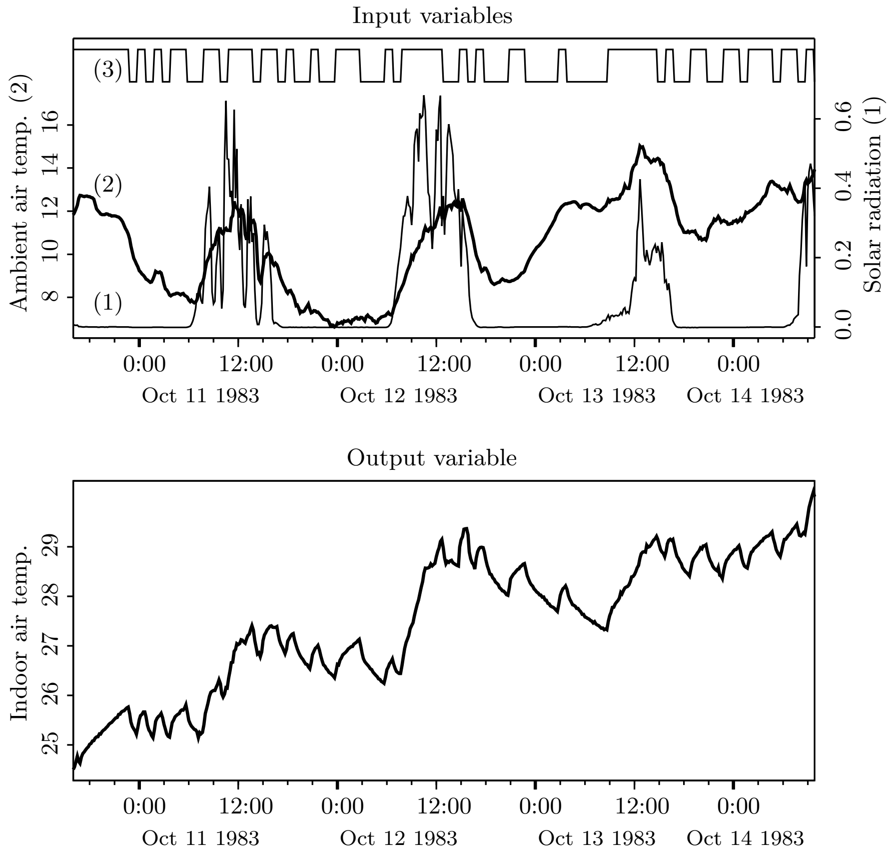

Figure 1.3: *Measurements from an unoccupied test building. The input variables are (1) solar radiation, (2) ambient air temperature, and (3) heat input. The output variable is the indoor air temperature.*

For this case methods for transfer function modeling as described in Chapter 8 can be used, where the input (explanatory) variables are the solar radiation, heat input, and outdoor air temperature, while the output (dependent) variable is the indoor air temperature. For the methods in Chapter 8 it is crucial that all the signals can be classified as either input or output series related to the system considered.

1.1.4 Predator-prey relationship

This example illustrates a typical multivariate time series, since it is not possible to classify one of the series as input and the other series as output. Figure 1.4 shows a widely studied predator-prey case, namely the series of annually traded skins of muskrat and mink by the Hudson's Bay Company

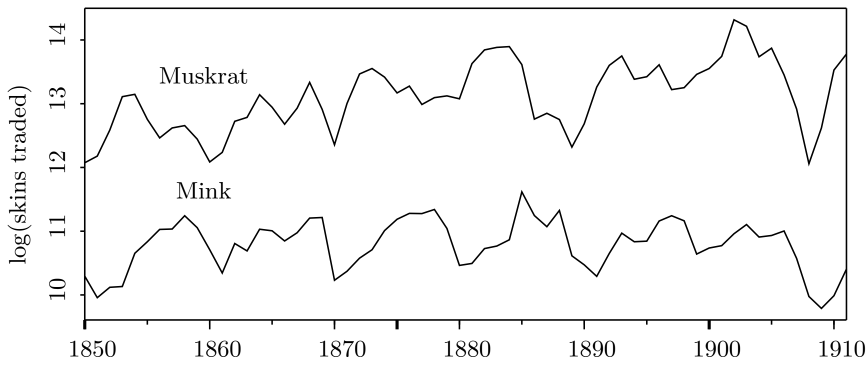

Figure 1.4: *Annually traded skins of muskrat and mink by the Hudson's Bay Company after logarithmic transformation. It is not possible to classify one of the series as input and the other series as output.*

during the 62 year period 1850–1911. In fact the population of muskrats depends on the population of mink, and the population of mink depends on the number of muskrats. In such cases both series must be included in a multivariate time series. This series has been considered in many texts on time series analysis, and the purpose is to describe in general the relation between populations of muskrat and mink. Methods for analyzing such multivariate series are considered in Chapter 9.

1.2 A first crash course

Let us introduce some of the most important concepts of time series analysis by considering an example where we look for simple models for predicting the monthly prices of wheat.

In the following, let P_t denote the price of wheat at time (month) t. The first naive guess would be to say that the price next month is the same as in this month. Hence, the *predictor* is

$$\widehat{P}_{t+1|t} = P_t \,. \tag{1.1}$$

This predictor is called the *naive predictor* or the *persistent predictor*. The syntax used is short for a prediction (or estimate) of P_{t+1} given the observations $P_t, P_{t-1}, \ldots$.

Next month, i.e., at time $t+1$, the actual price is P_{t+1}. This means that the *prediction error* or *innovation* may be computed as

$$\varepsilon_{t+1} = P_{t+1} - \widehat{P}_{t+1|t} \,. \tag{1.2}$$

By combining Equations (1.1) and (1.2) we obtain the *stochastic model* for the wheat price

$$P_t = P_{t-1} + \varepsilon_t \tag{1.3}$$

If $\{\varepsilon_t\}$ is a sequence of uncorrelated zero mean random variables (*white noise*), the process (1.3) is called a *random walk*. The random walk model is very often seen in finance and econometrics. For this model the optimal predictor is the naive predictor (1.1).

The random walk can be rewritten as

$$P_t = \varepsilon_t + \varepsilon_{t-1} + \cdots \tag{1.4}$$

which shows that the random walk is an integration of the noise, and that the variance of P_t is unbounded; therefore, no stationary distribution exists. This is an example of a *non-stationary process*.

However, it is obvious to try to consider the more general model

$$P_t = \varphi P_{t-1} + \varepsilon_t \tag{1.5}$$

called the *AR*(1) *model* (the autoregressive first order model). For this process a stationary distribution exists for $|\varphi| < 1$. Notice that the random walk is obtained for $\varphi = 1$.

Another candidate for a model for wheat prices is

$$P_t = \psi P_{t-12} + \varepsilon_t \tag{1.6}$$

which assumes that the price this month is explained by the price in the same month last year. This seems to be a reasonable guess for a simple model, since it is well known that wheat price exhibits a *seasonal variation*. (The noise processes in (1.5) and (1.6) are, despite the notation used, of course, not the same).

For wheat prices it is obvious that both the actual price and the price in the same month in the previous year might be used in a description of the expected price next month. Such a model is obtained if we assume that the innovation ε_t in model (1.5) shows an annual variation, i.e., the combined model is

$$(P_t - \varphi P_{t-1}) - \psi(P_{t-12} - \varphi P_{t-13}) = \varepsilon_t\,. \tag{1.7}$$

Models such as (1.6) and (1.7) are called *seasonal models*, and they are used very often in econometrics.

Notice, that for $\psi = 0$ we obtain the AR(1) model (1.5), while for $\varphi = 0$ the most simple seasonal model in (1.6) is obtained.

By introducing the *backward shift operator* B by

$$\mathrm{B}^k P_t = P_{t-k} \tag{1.8}$$

the models can be written in a more compact form. The AR(1) model can be written as $(1 - \varphi\,\mathrm{B})P_t = \varepsilon_t$, and the seasonal model in (1.7) as

$$(1 - \varphi\,\mathrm{B})(1 - \psi\,\mathrm{B}^{12})P_t = \varepsilon_t \tag{1.9}$$

If we furthermore introduce the *difference operator*

$$\nabla = (1 - \mathrm{B}) \tag{1.10}$$

then the random walk can be written $\nabla P_t = \varepsilon_t$ using a very compact notation. In this book these kinds of notations will be widely used in order to obtain compact equations.

Given a *time series* of observed monthly wheat prices, $P_1, P_2, \dots, P_N$, the *model structure* can be identified, and, for a given model, the time series can be used for *parameter estimation*.

The *model identification* is most often based on the estimated autocorrelation function, since, as it will be shown in Chapter 6, the autocorrelation function fulfils the same difference equation as the model. The autocorrelation function shows how the price is correlated to previous prices; more specifically the autocorrelation in lag k, called $\rho(k)$, is simply the correlation between P_t and P_{t-k} for stationary processes. For the monthly values of the wheat price we might expect a dominant annual variation and, hence, that the autocorrelation in lag 12, i.e., $\rho(12)$ is high.

The models above will, of course, be generalized in the book. It is important to notice that these processes all belong to the more general class of linear processes, which again is strongly related to the theory of linear systems as demonstrated in the book.

1.3 Contents and scope of the book

As mentioned previously, this book will concentrate on analyzing and modeling dynamical systems using statistical methods. The approach taken will focus on the formulation of appropriate models, their theoretical characteristics, and on links between the members of the class of stochastic dynamic models considered. In general, the models considered are all linear and formulated in discrete time. However, some results related to continuous time models are provided.

This section describes the contents of the subsequent chapters. In order to illustrate the relation between various models, some fundamental examples of the considered models are outlined in the following section. However, for more rigorous descriptions of the details related to the models we refer to the following chapters.

In Chapter 2 the concept of multivariate random variables is introduced. This chapter also introduces necessary fundamental concepts such as the conditional mean and the linear projection. In general, the chapter provides the formulas and methods for adapting a second order approach for characterising random variables. The second order approach limits the attention to first and second order central moments of the density related to the random variable. This approach links closely to the very important second order characterisation of stochastic processes by the autocovariance function in subsequent chapters.

Although time series are realizations of dynamical phenomena, non-dynamical methods are often used. Chapter 3 is devoted to describing *static models* applied for time series analysis. However, in the rest of the book dynamical models will be considered. The methods introduced in Chapter 3 are all linked to the class of regression models, of which the general linear model is the most important member. A brief description of the general linear model follows here.

In the following, let Y_t denote the dependent variable and $\boldsymbol{x}_t = (x_{1t}, x_{2t}, \ldots, x_{pt})^T$ a known vector of p explanatory (or independent) variables indexed by the time t. The *general linear model (GLM)* is a linear relation between the variables which can be written

$$Y_t = \sum_{k=1}^{p} x_{kt}\theta_k + \varepsilon_t \tag{1.11}$$

where ε_t is a zero mean random variable, and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^T$ is a vector of the p parameters of the model. Notice that the model (1.11) is a static model since all the variables refer to the same point in time.

On-line and recursive methods are very important for time series analysis. These methods provide us with the possibility of always using the most recent data, e.g., for on-line predictions. Furthermore, changes in time of the considered phenomena calls for adaptive models, where the parameters typically are allowed to vary slowly in time. For on-line predictions and control, adaptive estimation of parameters in relatively simple models is often to be preferred, since the alternative is a rather complicated model with explicit time-varying parameters. Adaptive methods for estimating parameters in the general linear model are considered in Chapter 3. This approach introduces *exponential smoothing*, the *Holt-Winter procedure*, and *trend models* as important special cases.

The remaining chapters of the book consider linear systems and appropriate related *dynamical models*. A linear system converts an input series to an output series as illustrated in Figure 1.5.

In Chapter 4 we introduce linear dynamic deterministic systems. In this chapter, one should note that for random variables capital letters are used whereas for deterministic variables we use lower case letters.

As a background for Chapter 4, one should be aware that for linear and time-invariant systems the fundamental relation between the deterministic

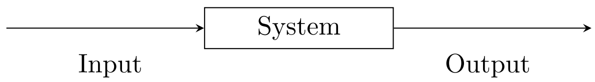

Figure 1.5: *Schematic representation of a linear system.*

input x_t and the corresponding output y_t is the *convolution*

$$y_t = \sum_{k=-\infty}^{\infty} h_k x_{t-k} \tag{1.12}$$

The sequence $\{h_k\}$ is called the *impulse response function* for the linear dynamic system. For physical systems where the output does not depend on future values of the input, the sum in (1.12) is from $k = 0$. Based on the impulse response function we will obtain the *frequency response function* by a Fourier transformation, and the *transfer function* by using the z transformation.

A very important model belonging to the model class described by (1.12) is the *linear difference equation*

$$y_t + \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} = \omega_0 x_t + \omega_1 x_{t-1} + \cdots + \omega_q x_{t-q} \tag{1.13}$$

Chapter 5 considers stochastic processes, and the focus is on the linear stochastic process $\{Y_t\}$ which is defined by the convolution

$$Y_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} \tag{1.14}$$

where $\{\varepsilon_t\}$ is the so-called white noise process, i.e., a sequence of mutually uncorrelated identically distributed zero mean random variables. Equation (1.14) defines a zero mean process, however, if the mean is not zero the mean μ_Y is just added on the right hand side of (1.14).

Notice the similarity between (1.14) and (1.12). This implies that a transfer function can be defined for the linear process as for the deterministic linear systems. Stochastic processes with a rational transfer function are the ARMA(p, q) process, the ARIMA(p, d, q) process, and the multiplicative seasonal processes. These important processes are considered in detail. As an example the process $\{Y_t\}$ given by

$$Y_t + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \tag{1.15}$$

where $\{\varepsilon_t\}$ is white noise, is known as the *ARMA(p, q) process*. Notice the similarity between (1.14) and (1.15). Such a process is useful for describing the data related to the dollar to Euro exchange rate in Section 1.1.1, and the seasonal process models are useful for modeling the monthly number of airline passengers in Section 1.1.2.

Given a time series of observations $Y_1, Y_2, \ldots, Y_N$, Chapter 6 deals with identification, estimation, and model checking for finding an appropriate model for the underlying stochastic process. This chapter focuses on time domain methods where the autocorrelation function is the key to an identification. Frequency domain methods are typically linked to the spectral analysis which is the subject of Chapter 7.

The so-called *transfer function models* are considered in Chapter 8. This class of models describes the relation between a stochastic input process $\{X_t\}$ and the output process $\{Y_t\}$. Basically the models can be written

$$Y_t = \sum_{k=0}^{\infty} h_k X_{t-k} + N_t \tag{1.16}$$

where $\{N_t\}$ is a correlated noise process, e.g., an ARMA(p, q) process. This gives rise to the so-called *Box-Jenkins transfer function model*, which can be seen as a combination of (1.12) and (1.14) on the previous page. It is relatively straightforward to include a number of input processes by adding the corresponding number of extra convolutions on the right hand side of (1.16).

An important assumption related to the Box-Jenkins transfer function models is that the output process does not influence the input process. Hence for the heat dynamics of a building example in Section 1.1.3, a transfer function model for the relation between the outdoor air temperature and the indoor air temperature can be formulated. This model can be extended to also include the solar radiation and the heat supply (provided that no feedback exists from the indoor air temperature to the heat supply).

In the case of multiple processes with no obvious split in input and output processes, the multivariate approach must be considered. In Chapter 9 the *multivariate linear process* is introduced as an m-dimensional stochastic process $\{\boldsymbol{Y}_t\}$ defined by the multivariate convolution

$$\boldsymbol{Y}_t = \sum_{k=0}^{\infty} \boldsymbol{\psi}_k \boldsymbol{\varepsilon}_{t-k} \tag{1.17}$$

where $\boldsymbol{\psi}$ is a coefficient matrix, and $\{\boldsymbol{\varepsilon}_t\}$ the multivariate white noise process. This formulation is used in Chapter 9 as the background for formulating the *multivariate ARMA(p, q) process* (also called the Vector-ARMA(p, q) process) and other related models.

As mentioned previously, the muskrat-mink case represents a problem which must be formulated as a multivariate process, simply because the population of minks influences the population of muskrats and vice versa.

Until now all the models can be considered as input-output models. The purpose of the modeling procedure is simply to find an appropriate model which relates the output to the input process, which in many cases is simply the white noise process. An important class of models which not only focuses on the input-output relations, but also on the internal state of the system, is the class of *state space models* introduced in Chapter 10.

A state space model in discrete time is formulated using a first order (multivariate) difference equation describing the dynamics of the *state vector*, which we shall denote $\boldsymbol{X}_t$, and a static relation between the state vector and the (multivariate) observation $\boldsymbol{Y}_t$. More specifically the *linear state space model* consists of the *system equation*

$$\boldsymbol{X}_t = \boldsymbol{A}\boldsymbol{X}_{t-1} + \boldsymbol{B}\boldsymbol{u}_{t-1} + \boldsymbol{e}_{1,t}, \tag{1.18}$$

and the *observation equation*

$$\boldsymbol{Y}_t = \boldsymbol{C}\boldsymbol{X}_t + \boldsymbol{e}_{2,t}, \tag{1.19}$$

where $\boldsymbol{X}_t$ is the m-dimensional, latent (not directly observable), random *state vector*. Furthermore $\boldsymbol{u}_t$ is a deterministic *input vector*, $\boldsymbol{Y}_t$ is a vector of observable (measurable) stochastic output, and $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ are known matrices of suitable dimensions. Finally, $\{\boldsymbol{e}_{1,t}\}$ and $\{\boldsymbol{e}_{2,t}\}$ are vector white noise processes.

For linear state space models the *Kalman filter* is used to estimate the latent state vector and for providing predictions. The *Kalman smoother* can be used to estimate the values of the latent state vector, given all N values of the time series, for $\boldsymbol{Y}_t$.

To illustrate an example of application of the state space model, consider again the heat dynamics of the test building in Section 1.1.3. **HMnJHt95** shows that a second order system is needed to describe the dynamics. Furthermore it is suggested to define the two elements of the state vector as the indoor air temperature and the temperature of the heat accumulating concrete floor. The input vector $\boldsymbol{u}_t$ consists of the ambient air temperature, the solar radiation, and the heat input. Only the indoor air temperature is observed, and hence, $\boldsymbol{Y}_t$ is the measured indoor air temperature. Using the state space approach gives us a possibility of estimating the temperature of heat accumulating in the concrete floor using the so-called Kalman filter technique.

In general, the parameters of the models are assumed constant in time. However, in practice, it is often observed that the dynamical characteristics change with time. In Chapter 11 *recursive and adaptive* methods are introduced. Basically the adaptive schemes introduce a time window related to the data, such that newer data obtains more influence than older data. This leads to methods for adaptive forecasting and control.