MODELLING AND PREDICTION OF LOAD IN DISTRICT HEATING SYSTEMS

Ken Sejling

LYNGBY 1993
Ph.D. THESIS
No. 65

ISSN 0907-1083

Typeset in LaTeX.

Trykt af IMSOF - DTH
Bogbinder Hans Meyer

PREFACE

The present thesis has been prepared at the Institute of Mathematical Statistics and Operations Research (IMSOR), Technical University of Denmark, in partial fulfillment of the requirements for earning the degree of Ph.D. in engineering.

The thesis discusses issues in modelling and prediction of heat load in district heating systems. It is implied that the reader has a basic knowledge of statistical dynamic modelling theory. Furthermore, knowledge about the physics of district heating systems will be beneficial to the understanding of the modelling issue.

The intention of the work has been to investigate and propose techniques for modelling and prediction of heat load. A number of different approaches are discussed, each one having its justification in being suitable for certain aspects of the modelling and prediction issue. However, the thesis does not pretend to give a complete survey of applicable techniques.

Lyngby, May 1993

Ken Sejling

Some of the work in this thesis has previously been published in:

Jonsson, G., Palsson, O.P. & Sejling K. (1992). Modeling and Parameter Estimation of Heat Exchangers - A Statistical Approach. *Journal of Dynamic Systems, Measurement, and Control,* **114**, 673-679.

Sejling, K. (1991). *Application of Nonparametric Regression Methods on Load Data in Energy Systems.* Technical Report 19/1991, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark.

Madsen, H., Palsson, O.P., Sejling, K. & Søgaard, H.T. (1990). *Models and Methods for Optimization of District Heating Systems, Part I: Models and Identification Methods.* Institute of Mathematical Statistics and Operations Research, Technical University of Denmark.

Madsen, H., Palsson, O.P., Sejling, K. & Søgaard, H.T. (1992). *Models and Methods for Optimization of District Heating Systems, Part II: Models and Control Methods.* Institute of Mathematical Statistics and Operations Research, Technical University of Denmark.

Sejling, K., Madsen, H., Holst, J., Holst, U. & Englund, J-E. (1994). Methods for Recursive Robust Estimation of AR Parameters. To be published in *Computational Statistics & Data Analysis*, primo 1994.

ACKNOWLEDGEMENTS

This thesis had not come into existence had it not been for the support and contributions of several people.

First of all I would like to thank my supervisors, Assoc. Prof., Ph.D. Henrik Madsen (IMSOR) and Assoc. Prof., Techn. Dr. Jan Holst (Department of Mathematical Statistics, Lund Institute of Technology, Lund, Sweden) for constructive criticism and lasting enthusiasm in search for improvements.

Furthermore, I wish to thank Henning T. Søgaard, Arne Henningsen, Ólafur P. Pálsson, Gudmundur R. Jónsson, Henrik Melgaard, Atli Benonysson, Ulla Holst and Jan-Eric Englund for pleasant and beneficial cooperation. In particular I am indebted to Gudmundur R. Jónsson, Arne Henningsen and Ólafur P. Pálsson for reading proofs of the manuscript.

During the preparations I have repeatedly appreciated the excellent facilities at IMSOR, and especially the group of systems programmers (super users) deserve thanks for providing and maintaining a superior computer system.

I wish to express my gratitude to those who supported the research of this thesis financially:

- The District Heating Programme under the Energy Research Cooperation of the Nordic Council of Ministers for financial support and incitement to cooperation among researchers in the Nordic countries.
- The Research Programme of the Danish Ministry of Energy for financial support under contracts EFP 1323/89-14 and EFP 1323/91-10.
- The power plant company I/S Vestkraft and the distribution companies in Esbjerg, Forsyningsvirksomhederne i Esbjerg and Varde Kommunale Værker, for letting their system be our experimental setup,

and especially O. Majland and M. Nohr for planting the seed which started the research in modelling and optimization of district heating systems at IMSOR, and lead to our mutual cooperation.

- The Department of Heat and Power Engineering at Lund Institute of Technology for purchasing measuring equipment and providing data from their heat exchanger test rig as well as from Kulladal in the Malmö district heating net.

During my stay in 1987-1988 at the Department of Mathematical Statistics, Lund Institute of Technology, I benefited from a number of Ph.D. courses as well as in matters on probability and statistics in general. I want to show my gratitude for that.

Last, but not least I am grateful to my fiancée, my family and my friends for supporting me all the way in my preparations of the present thesis.

SUMMARY

The subject of the present thesis is the construction of a dynamic model of the heat supply to district heating systems, and further, the investigation and development of methods for making on-line predictions of the heat demand. A particular problem, which is of importance for control of plant supply temperature, is the model description of how the variation of plant supply temperature is determining for the variation of the heat supply.

A district heating system is a distributed dynamic system with considerable transport delays in the distribution of heat. In Chapter 2 an introductory account of the configuration of a typical district heating system is given. The delimitation of the system to the surroundings is specified, and the external variates influencing the system are discussed.

Heat exchangers are used in district heating systems for heat transfer between primary transmission net and attached secondary distribution systems, and between distribution systems and subscriber installations. Chapter 3 describes the modelling of a heat exchanger carried out by expressing conservation of energy and mass in each compartment of a lumped system description. The compartmental model structure is used with the purpose of obtaining an approximation to the distributed heat transfer in the heat exchanger. Expressing conservation of energy and mass for each compartment leads to a set of differential equations which, by introduction of process and measurement noise, constitutes a stochastic time-continuous state space model for the outlet temperatures of the heat exchanger. The fact that the heat transfer depends on temperature and flow rate of the water implies nonlinear state dependence on these variates. In the chapter approximations within the class of linear state space models are investigated. The estimation is carried out in the time-continuous model formulation meaning that the parameters are physically interpretable. An advantageous implication of this is that it is possible to assess and compare the

parameter estimates with values obtained from physical constants, system dimensions etc.

If a model describing the supply of heat and its dependence on external effects, e.g. ambient air temperature, should be based on a combination of physically based component models, the complete model would end up being exceedingly complex. In the present investigation this is not considered to be a viable way of obtaining a model for the supply of heat. Instead different classes of models are considered, and within each class of model the appropriate structure and parameterization are determined using statistical methods. Chapter 4 describes the application of nonparametric regression to the disclosure of the heat supply dependence on time of day, ambient air temperature and supply temperature. The insufficiency of the linear model class is made clear, and the character of required model extensions is indicated.

Chapter 5 contains an investigation of the capability of various model classes to describe the dynamic relations between the required heat supply and explanatory variates. A model within the class of linear transfer function models is estimated. Extensions to the linear transfer function models are estimated consisting of the addition of (1) transfer functions from squared ambient air temperature and mass flow scaled supply temperature and (2) transfer functions from ambient air temperature and supply temperature, where the gains have a smooth threshold dependence on ambient air temperature and inverse mass flow, respectively. In addition neural network models of feed forward type with two layers of neurons are estimated. The result is that all of the nonlinear models are capable of giving an improved description compared to the linear model. When it comes to validation of the estimated models on a different set of data the nonlinear transfer function models show to give the best performance. The modelling results are supplemented by a description of how they are used for prediction, and the use of j-step predictors and multi-step predictors are motivated.

Chapter 6 describes the results of RLS estimation with exponential forgetting of linear models for j-step prediction of heat demand. Exponential

forgetting is applied to allow the linear model approximate the nonlinear relations at the operating point. Likewise, exponential smoothing of the heat production of each hour of the week superposed a value for each hour of the day (for faster adaption to level changes) is investigated. This is introduced into a state space model with 192 state variables, for which the Kalman filter is applied for update of the state vector. The conclusion from the investigation is that for prediction on shorter term than 24 hours the best prediction ability is obtained using RLS estimation with exponential forgetting of a linear model with trigonometric diurnal profile. For increasing prediction horizon, j, the performance of the two methods approaches each other. This is due to the observed result that the weight of the predictions on a diurnal profile increases with j, and consequently, as both contain a diurnal profile, their prediction abilities turn out to approach each other.

In an on-line implementation of a model it is evident that collected data may be contaminated with errors occurring either in the recording or transmission of data or, for instance, in the case of a heat supplying system as a consequence of interruption of operation. Robust estimation methods may be applied to minimize the influence of erroneous data (outliers) on the parameter estimates. Chapter 7 discusses the issue of robust estimation for outliers of innovation type (deviations from assumed innovation density) and additive type (errors in observations). An approach for the derivation of recursive robust estimation algorithms, having the estimator given by a criterion to be minimized, is proposed and used to obtain two algorithms for recursive robust estimation with exponential forgetting of AR parameters. In a simulation study they are compared to RLS and a modification of RLS, where observations are treated as missing if the prediction error, compared to an estimated scale parameter, exceeds a specified bound. The simulations show that the derived algorithms give estimation results, which are similar to their off-line counterparts. However, it is also demonstrated that the crucial effect of additive outliers, resulting in severe bias on parameter estimates, is only down-weighted and not removed by any of the proposed estimation methods.

SAMMENFATNING (SUMMARY IN DANISH)

Emnet for nærværende afhandling er opbygningen af en dynamisk model for varmetilførslen til fjernvarmesystemer, og derudover undersøgelsen og udviklingen af metoder til on-line beregning af prædiktioner af varmebehovet. Et specielt problem, som er vigtigt i forbindelse med styringen af fremløbstemperaturen på det varmeproducerende værk, er modelbeskrivelsen af, hvordan fremløbstemperaturen fra værket er bestemmende for varmetilførslen.

Et fjernvarmesystem er et distribueret dynamisk system med betragtelige transporttider i fordelingen af varme. Indledningsvis redegøres i Kapitel 2 for opbygningen af et typisk fjernvarmesystem. Systemets afgrænsning til omgivelserne specificeres, og de herfra kommende eksterne påvirkninger diskuteres.

Varmeveklsere anvendes i fjernvarmesystemer til varmeoverførsle mellem primære transmissionsnet og tilknyttede sekundære distributionssystemer, og mellem distributionssystemer og forbrugerinstallationer. Kapitel 3 beskriver modelleringen af en varmeveksler gennemført ved at udtrykke energibevarelse for hvert afsnit af en sektioneret systembeskrivelse. Sektioneringen i systembeskrivelsen anvendes for at opnå en approksimation til den distribuerede varmeoverførsel i varmeveksleren. Udtrykkene for energibevarelse fører til et sæt af differentialligninger, som, ved at introducere støj på proces og målinger, udgør en stokastisk tidskontinuert tilstandsbeskrivelse for udløbstemperaturerne i varmeveksleren. Den kendsgerning, at varmeoverføringskoefficienten afhænger af temperatur og gennemstrømningshastighed, medfører en ikke-lineær tilstandsafhængighed af disse variable. I kapitlet undersøges approksimationer indenfor klassen af lineære tilstandsmodeller. Estimationen gennemføres i den tidskontinuerte modelformulering, hvilket betyder, at parametrene er fysisk fortolkbare. En fordelagtig følge af dette er, at det er muligt at vurdere og sammen-

ligne parameterestimaterne med værdier opnået v.h.a. fysiske konstanter, systemdimensioner etc.

Hvis en samlet model for varmetilførslen og dens afhængighed af eksterne påvirkninger, som f.eks. udetemperatur, skulle baseres på en kombination af fysisk baserede komponentmodeller ville modellen blive særdeles kompliceret. I denne undersøgelse skønnes dette at være en ufarbar vej. I stedet betragtes kendte afgrænsede modelklasser, og indenfor hver klasse søges passende modelordener. Kapitel 4 beskriver anvendelsen af ikke-parametriske regression til afdækning af varmetilførslens afhængighed af tidspunkt på døgnet, udetemperatur og fremløbstemperatur. Utilstrækkeligheden af den lineære modelklasse gøres åbenbar, og karakteren af modeludvidelse indikeres.

Kapitel 5 indeholder en undersøgelse af forskellige modelklassers evne til at beskrive de dynamiske sammenhænge mellem varmetilførsel og forklarende variable. En model i klassen af lineære overføringsfunktionsmodeller estimeres. Desuden undersøges udvidelser til den lineære modelklasse bestående i indførelsen af (1) overføringsfunktioner fra kvadratled af udetemperatur og gennemstrømningsskaleret fremløbstemperatur og (2) udglattede tærskelfunktioner, hvor disse afhænger af hhv. vandgennemstrømning og udetemperatur, i overføringen fra fremløbstemperatur og udetemperatur. Ligeledes estimeres parameteren i en feed forward neural netværksmodel med to lag af neuroner. Disse modelklasser viser sig alle at give en forbedret tilpasning til data i forhold til den lineære model. Når de estimerede modeller afprøves på et andet datasæt viser det sig, at de to ikke-lineære udvidelser til overføringsfunktionsmodellen giver den bedste beskrivelse. Endelig motiveres anvendelsen af j-trins- og multitrinsprædiktorer.

Kapitel 6 beskriver resultateteterne af rekursiv mindste kvadraters estimation med eksponentiel glemsel af lineære modeller til j-trins prædiktion af varmebehov. Den eksponentielle udjævning anvendes for at tillade den lineære model approksimere de ikke-lineære relationer i arbejdspunktet. Ligeledes undersøges eksponentiel udjævning af varmeproduktionen for hver time i ugen overlejret en værdi for hver time i døgnet

(for hurtigere tilpasning til niveauændringer). Dette formuleres i en tilstandsmodel med 192 tilstandsvariable, i hvilken Kalman-filteret bruges til opdatering af tilstandsvektoren. Konklusionen på undersøgelsen er, at for prædiktion på kortere sigt end 24 timer opnås den bedste prædiktionsevne med rekursiv mindste kvadraters estimation og eksponentiel glemsel af en lineær model med trigonometrisk døgnprofil. For voksende horisont j nærmer prædiktionsevnen af de to metoder sig til hinanden. Dette skyldes, prædiktionernes vægtning af en døgnprofil vokser med j, og derfor viser det sig, eftersom de to metoder begge indeholder en døgnprofil, at prædiktionsevnerne nærmer sig hinanden.

I en on-line implementering af en model er det givet, at opsamlede målinger kan indeholde fejl. Disse kan opstå i opsamling eller transmission, eller, f.eks., som følge af driftforstyrrelser i et varmeforsynende system. Robuste estimationsmetoder kan anvendes til at minimere indflydelsen på en models parameterestimater af fejl i data (outliers). Kapitel 7 diskuterer robust estimation for datafejl af innovation type (afvigelser i tæthedsfunktionen for innovationer) og additiv type (fejl i observationer). En metode til udledningen af rekursive robuste estimationsmetoder udfra et robust kriterium anvendes til at udelede to algoritmer til robust estimation med eksponentiel glemsel af AR processer. I et simulationsundersøgelse sammenlignes disse med RLS og en udvidet version af RLS, hvor målinger behandles som manglende observationer, hvis den tilhørende prædiktionsfejl, i forhold til en estimeret skala parameter, overstiger en given grænse. Simulationerne viser, at de udledte algoritmer giver estimationsresultater, som er tæt på dem, som fås med oprindelige off-line metoder. approksimationer i udledningen forsvinder med antallet af observationer. Dog gøres det også klart, at den uheldige effekt af additive outliers (bias) kun er nedskaleret og ikke fjernet af nogle af de foreslåde metoder.

Symbols, Abbreviations and Notational Conventions

Symbols

N	number of observations
n	order of model
p	order of AR polynomium
q	order of MA polynomium
$x(t)/x_i$	external/explanatory variable
$\boldsymbol{x}(t)$	regressor vector
$u(t)$	control variable
$y(t)/y_i$	dependent variable
$o(t)$	output of neuron
$z(t)$	observation of dependent variable
$e(t)/e_i$	independent/uncorrelated and identically distributed random variable
k	time delay
$\varepsilon(t)$	modelling error
$\varepsilon(t+j \mid t)$	j-step prediction error
$\boldsymbol{\theta}$	parameter vector
$w_{i,j}, b_i$	parameters in neural net
$V(\boldsymbol{\theta})$	criterion dependent on θ
$\boldsymbol{M}$	Fisher information matrix
$\boldsymbol{H}$	Hessian
$\boldsymbol{P}$	Covariance matrix
λ	forgetting factor
σ^2	variance
σ_ε^2	prediction error variance

s_ε	empirical standard deviation
$T_h(t)$	temperature of hot flow in heat exchanger
$T_c(t)$	temperature of cold flow in heat exchanger
$T_m(t)$	temperature of metal separation in heat exchanger
$\Delta T(t)$	heat transfer driving force in heat exchanger
$\dot{m}(t)$	mass flow
$p(t)$	heat load/supply
$a(t)$	ambient air temperature
$s(t)$	supply temperature
$r(t)$	return temperature
$w(t)$	wind velocity
M	mass
c	specific heat
U	overall heat transfer coefficient
h	heat transfer coefficient
A	heat transfer area
n_s	number of compartments
α	heat flow parameter
γ	scaled mass parameter
y	parameter for mass flow dependence
$m(u)$	regression function
$K(u)$	kernel
$W(u)$	weight function for nonparametric estimate
h	bandwidth
d_A	average squared error
$\phi(u)$	orthonormal functions
$P(u)$	Legendre polynomials
$\rho(u)$	weight function for robust M-estimation
$\psi(u)$	$= \frac{d}{du}\rho(u)$
$\chi(u)$	weight function for robust M-estimation of scale
$\boldsymbol{A}(t)/\boldsymbol{A}_N$	Robust estimat of regressor covariance

$\boldsymbol{A}$	system matrix of state space model in continuous time
$\boldsymbol{\Phi}$	system matrix of state space model in discrete time
$\boldsymbol{B}$	input matrix of state space model in continuous time
$\boldsymbol{\Gamma}$	system matrix of state space model in discrete time
$\boldsymbol{C}$	measurement matrix of state space model
$\boldsymbol{R}$	covariance matrix for noise in state space model
$\boldsymbol{K}$	Kalman gain
$\boldsymbol{w}(t)$	Wiener process

ABBREVIATIONS

AO	additive outliers
AR	autoregressive
ARMAX	autoregressive moving average with external input
AIC	Akaike information criterion
BIC	Bayes information criterion
CHP	combined heat and power plant
CV	cross validation
DH	district heating
EFF	asymptotic efficiency
FPE	final prediction error (criterion)
IO	innovation outliers
LMTD	log mean temperature difference
L	likelihood
LS	least squares
LSCV	least squares cross validation
MA	moving average
ML	maximum likelihood
NO	no outliers
Nu	Nusselt number
PR	Prandtl number
Re	Reynolds number

RH1	recursive Huber algorithm (version 1)
RH2	recursive Huber algorithm (version 2)
RKW	recursive Krasker & Welsch algorithm
RLS	recursive least squares
RMO	recursive missing observation algorithm
RLS	recursive least squares
SS	sum of squares (measure)
SS%	relative sum of squares (measure)

NOTATIONAL CONVENTIONS

The following notation is adopted in the thesis

F	distribution function
f	density function
$N(\mu;\sigma^2)$	normal distribution with mean μ and variance σ^2
$AsN(\mu;\sigma^2)$	asymptotic normal distribution
$NIID(\mu;\sigma^2)$	independent and identically dist. normal variables
q^{-k}	backshift operator defined by $q^k x(t) = x(t-k)$
∇_k	difference operator defined by $\nabla_k x(t) = x(t) - x(t-k)$
∇	short for ∇_1
∇_θ	the derivative w.r.t. θ
$I_{\{\cdot\}}$	Indicator function
$\boldsymbol{X}$	boldface capital letters are matrices
$\boldsymbol{x}$	boldface non-capital letters are vectors
x	non-boldface non-capital letters are univariats
log	natural logarithm
arg min	the minimizing argument

Contents

1

INTRODUCTION

The investigation in the present thesis deals with the problem of obtaining a dynamic model describing the variations of the heat supply in a district heating (DH) system. Especially, the description of how the heat supply depends on the variations of plant supply temperature is a quality of the model, which is of relevance for control of supply temperature. The thesis further deals with the investigation and development of methods for prediction of system heat demand up to 24 hours ahead. This covers exponential weighting methods and recursive robust algorithms for estimation of AR parameters.

A first important issue, relating to the modelling and prediction problem, is that of defining the terms heat demand, heat consumption, heat supply and heat load. This is done in Chapter 2 in connection with an outline of a typical DH system and an account of the external processes influencing the system.

The intentions of the investigations in the thesis is to provide means for saving energy in the production and distribution of heat in DH systems. Basically to run the heat production and distribution in accordance with a given optimal strategy, it is necessary to have a complete and exact description of the DH system. In addition, since the system is dynamic with built-in transport delays, meaning that momentary control decisions influence future states, it is required that the system description is dynamic. Moreover, it is also necessary to know in advance, what the heat demand

of the consumers is going to be, or alternatively, how the external variates, having influence on the heat demand of the consumers, develop in the future. The purpose of the thesis is to investigate and develop methods for the construction of approximations to the exact modelling of future heat supply variations.

1.1 OUTLINE OF THE THESIS

The issues in modelling of heat supply and prediction of heat demand reaches from the initial phase of determining an appropriate model type and order to the application of the model for on-line prediction. Traditionally the process of obtaining a sufficient dynamic model, using statistical methods, is separated into three phases, see e.g. Box & Jenkins (1976), viz.

1. Identification of type, structure and order of model.
2. Estimation of model parameters.
3. Validation of the estimated model, i.e., test of whether the model gives a sufficient description of data.

This separation of the model construction process is also adopted in this thesis.

An important initial step in developing a dynamic model is to use physical insight and knowledge about the system to get hold of the dynamic relations in the system. Chapter 2 serves this purpose in the current investigation.

In Chapter 3 a model of a heat exchanger in continuous time is developed. This is done based on physical considerations concerning conservation of energy and mass in a lumped parameter model. This kind of modelling is an example of a modelling process, where the use of physical insight goes far in the identification of model structure and order, and where it, indeed, pays off to use the thermodynamic theory of heat transfer in heat exchangers. The approach of carrying out the modelling and parameter

estimation in continuous time implies that the parameter estimates can be given a physical interpretation. This is a nice quality of the time-continuous modelling approach since it allows for a judgement of the reliability of the model, via a comparison of the estimated parameters to parameter values, which are to be expected from physical considerations and/or experimental results.

When system knowledge is limited, statistical methods can be used on system data to disclose the structure or even character of relations in the system. To determine the dynamic correlation structure, autocorrelation functions and cross correlation functions can be estimated (Box & Jenkins, 1976). The limitation of these methods is, however, that they only apply on linear relations. Chapter 4 describes how nonparametric regression methods can be used to estimate conditional expectations, which not necessarily are linear. This is done for the dependence of DH supply on time of day as well as supply temperature and ambient air temperature.

Chapter 5 describes the result of an investigation with the purpose of building a dynamic model which gives a sufficient description of the DH supply variations including its dependence on time of day and week, supply temperature and ambient air temperature. It is clear that there are other external processes influencing the heat demand in DH systems. These are, however, not considered here as the problem of modelling the nonlinear dependence is been given a higher priority. In the first place, a model within the class of linear transfer function models is determined. This model class can reasonably well be understood to be insufficient since nonlinear relations between the external variables and heat load do exist. This fact is also made clear from the nonparametric regression results in Chapter 4. The linear model class is extended to include specific nonlinear model structures, viz. models with transfer functions for second order explanatory variables and models with a smooth threshold in the gain of the transfer functions. Furthermore, models in the class of layered, feed-forward neural networks are estimated. The modelling approach of using linear as well as nonlinear transfer function models is semi-physical in nature. Indeed, physical considerations are used for determining explanatory variables and also for

assessment of the characteristics of the model estimate. However, the order of transfer functions and the parameters of the model are determined by statistical methods. In the neural network model building, physical considerations are only used when choosing input variables, whereas it is difficult to assess the validity of the estimated model.

Recursive methods for adaptive estimation of models and predictors are discussed in Chapter 6. The main reasons for bringing recursive, adaptive estimation into use and leaving the off-line models are twofold. First of all there is always a possibility that the system is subject to changes. For instance, networks are restructured and new subscribers are attached or detached DH systems continuously. This requires that a model is capable of adapting to such system changes. The use of recursive algorithms is a much more computationally efficient way of obtaining adaptivity in the estimation than repeated estimation of models using off-line methods. The second reason for using adaptive methods is the insufficiency of, at least linear, models in giving an overall description of the heat supply variations. The insufficiency of a model implies that, when it is estimated on a set of data, it will be an approximation to these data. In this case adaptive methods can be used to make the approximation of the insufficient model follow the operation point of the system.

The possibility of the occurrence of errors in recorded data motivates the use of methods for detection and estimation in the presence of such errors (outliers). Especially, when data are collected on-line for use in estimation, there are no other way of treating such outliers than in the algorithm taking care of the adaptive estimation. Chapter 7 develops and investigates algorithms for recursive robust estimation with exponential forgetting.

2

Description of a DH System

The purpose of the present chapter is to give an introductory description of the dynamics and relations in a DH system. In the first place, it is defined how the terms heat demand, heat consumption, heat supply and heat load should be interpreted in the present thesis. Furthermore, the delimitation of the system is defined, and it is discussed which external processes influence the system. In connection with the system outline the possibilities for optimizing by control and planning are discussed.

2.1 Dynamics and Delimitation of the DH system

The term heat demand is used for the amount of heating, which the consumer requires at the entrance of his installations subject to the prevailing conditions. The overall heat demand in a DH system is given as the sum of the heat demand of the individual consumers. The heat consumption is, similarly, the amount of heating that actually has been supplied to a consumer, and the overall system heat consumption is the sum of the heat consumption for all consumers. The heat supply to a DH system is the amount heat delivered to the transmission and distribution system from heat supplying units. Finally, the term heat load denotes the supply of heat for a given control of the system and subject to the condition that all consumers have their heat demand fulfilled. That is, a given heat supply may be insufficient, whereas the heat load is the same as the heat supply,

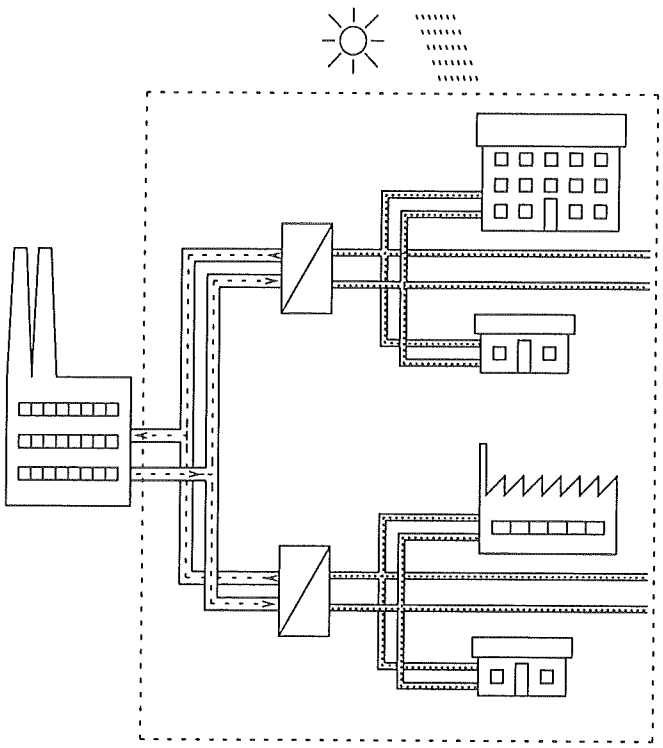


Figure 2.1: Sketch of a typical district heating system.

when there is no shortage of heating for any of the consumers.

The object for the modelling and forecasting study in this thesis is the heat supply to a given DH system conditioned on all consumers having their heat demand fulfilled, viz. the heat load. Figure 2.1 depicts a typical DH system. This system is delimited on the supply side at the point where the hot DH water leaves and the cold water enters a heat supplying plant.

It is clear that the necessary heat supply to a building, to keep up a stable indoor climate, depends on the ambient weather condition. This means that

explicit modelling of the influence of weather on the heat supply can most likely improve the performance of the model. This goes for a model of a single building and, consequently, also for a complete DH system. Therefore, whenever measurements of weather variables are at hand, e.g. ambient air temperature or solar radiation, these may be used as explanatory variables for describing the variations of heat load.

As a consequence of the use of weather measurements in modelling the heat load, it is natural to place the other limit of the system at the boundary surface, where the ambient weather condition influences the indoor climate of the buildings in the DH system. That is, this delimitation is placed in the outer layer of the external walls of the consumers' buildings. The overall delimitation of the typical DH system is illustrated in Figure 2.1 by the broken-line box.

Based on the chosen system delimitation the typical DH system consists of elements of the following type

- Installations and buildings of the consumers.
- Transmission and distribution lines.
- Possible heat exchangers between transmission lines and distribution system or between distribution network and consumers' installations.

The pumps of the system are intentionally omitted in this list. Actually is a great deal of the power, which is supplied to run the pumps, turned into frictional heat in the DH water. Especially in transmission lines is the power supply to the pumps of such an amount that it is a considerable contribution to heat supply (Benonysson, 1991). This matter is, however, not taken into account in the modelling studies in this thesis, at least not explicitly, first of all because the pumping power measurements were not available, but it is also considered to be of inferior influence.

Some systems may have an accumulator tank attached for service security reasons. This allows for optimization by loading the tank when the price of heat production is low and unloading when the price is high. The existence

of a tank is not relevant for the modelling issue as the tank usually is located at the production plant, and, as such, it can be considered to be exterior to the delimitation of the system.

2.1.1 BUILDINGS AND INSTALLATIONS

There are two conceptually different types of heat demand:

1. Heating.
2. Hot tap-water.

HEATING

The consumers require supply of heat for heating of the indoor air. The amount of heating depends on the desired degree of comfort with respect to indoor temperature. That is, the consumers' choice of degree of comfort is determining for the relation between heating and indoor temperature. The ambient weather condition in turn affects the indoor temperature through the building dynamics. Madsen (1985) discusses the variates influencing the indoor temperature through the dynamics of buildings. The list of influential variates of the ambient weather is

- ▷ Ambient air temperature.
- ▷ Short-wave radiation (Solar radiation).
- ▷ Wind speed and direction.
- ▷ Long-wave radiation.
- ▷ Precipitation.

The variates with the clearest influence are the ambient air temperature and the short-wave radiation. The temperature affects the indoor climate via heat conduction in the outer walls, roofs and windows, and by free

and forced ventilation. The heat conduction is dependent on the degree of insulation, and the ventilation depends on the leakages of the building. The short-wave radiation is responsible for the heating of indoor air and material as a consequence of the incident radiation through the windows of the building. Furthermore, the short-wave radiation heats the outer surfaces implying that the heat conduction through the walls, roofs and windows is lowered.

The wind affects the indoor temperature of the building through the ventilation. That is, dependent on speed and direction the wind is responsible for pressure differences around the building, which in turn implies an air flow through leakages of the building. The heat conduction through the walls, roofs and windows is also affected by the wind speed as the degree of cooling of the outer surfaces, via convection, increases with the wind speed. The influence of wind speed is generally of minor importance compared to that of ambient air temperature and solar radiation as buildings in Denmark normally are well insulated with a minimum of leakages.

The long-wave radiation emitted from outer surfaces of the buildings may, especially at clear nights and at low wind speeds, give rise to lowering of the surface temperatures, which consequently leads to an increase in the heat conduction through the walls and roofs. Except for buildings with nearly flat or poorly insulated roofs, the effect is, however, of inferior importance.

The influence of precipitation is due to fact that the heat conductivity of the walls and roofs depends on their degree of humidity. The heat resistance is typically lowered if a wall is wet.

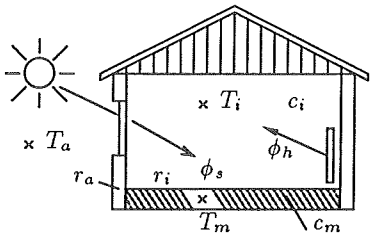
The list of influences originating from sources inside the buildings is

- Light.
- Machines.
- Persons.
- Radiators.

The contribution to the heat supply of both light, machines and persons may be of significant influence in buildings, where there is a large degree of lighting, many machines or many persons. The significance of the heat supply from these sources compared to the heat supply from radiators increases with the degree of insulation of the building since the total heat demand decreases with the insulation degree. For poorly insulated buildings the heat supply from the radiators will be dominating, but for well insulated buildings, which is the normal case in Denmark, the contributions from the other sources may be comparable to the heat supply from radiators.

It is clear from the discussion above that the ambient weather condition has a significant influence on the required heating in buildings. However, it can also be concluded that the part of the required heating, which is contributed from radiators, is significantly influenced by the heat contributions from light, machines and persons. This actually means that the dynamic relation between the ambient weather and the required heating from radiators in a building is varying concurrently with the changes of the heat contributions from other sources. However, for lighting of buildings, for the operation of machines and for the presence of persons in buildings, it is the case that it is more or less likely for a certain time of day, week and year. This means that it can be assumed to follow a probabilistic distribution implying that for an increasing number of buildings/consumers the total heat supply from these sources will tend to follow a systematic variation, i.e. it will tend to have a mean value for each time of day, week and year around which the variation decreases with the number of buildings. The mean value will not vary in the same way as the ambient weather meaning that in a modelling situation it is likely that the two contributions can be separated.

Based on this discussion, it can be concluded that the heat consumption shows a significant dependence on ambient weather condition. This dependence is, for Danish conditions, expected to be nonlinear since for ambient air temperatures above a level around 17 °C, heating is generally not necessary, but for decreasing ambient air temperatures the required indoor heating will increase. For such levels of ambient air temperatures, at which heating is necessary, the heat demand depends on the ambient weather as



T_m Temperature of large heat accum. medium.
T_i Temperature of indoor air.
T_a Temperature of amb. air.
c_m Heat capacity of large heat accum. medium.
c_i Heat capacity of indoor air.
ϕ_h Heat input from radiator.
ϕ_s Heat input from solar radiation.
r_a Resistance against heat transfer to amb. air.
r_i Resistance against heat transfer from indoor air to large heat accum. medium.

Figure 2.2: Sketch of building with interior dominating heat capacity and indication of heat fluxes.

discussed above.

Modelling results for the heat dynamics of buildings have shown that it is possible to describe the indoor temperature satisfactorily by a model with two time constants based on differential equations for a lumped parameter description, see Madsen, Nielsen, & Saxhof (1992a). The actual structure of such a model is specific for a given type of building. The important feature is, however, that typically there is one time constant for heavy-mass material in the building, e.g. walls and furniture, and another time constant for indoor air and light-mass material.

Figure 2.2 depicts an example of a building, where the dominating heat capacity is placed in the interior of the building. The pair of differential equations for the heat fluxes in the model is

$$c_m \frac{dT_m(t)}{dt} = \frac{1}{r_i}(T_i(t) - T_m(t)) \tag{2.1}$$

$$c_i \frac{dT_i(t)}{dt} = \frac{1}{r_a}(T_a(t) - T_i(t)) + \frac{1}{r_i}(T_i(t) - T_m(t)) + \phi_s(t) + \phi_h(t). \tag{2.2}$$

It is clear that this model does not describe the influence of all the weather variables listed above.

Based on data consisting of measurements of indoor temperature, ambient air temperature and heat fluxes from radiator and the sun, the parameters of this model can be estimated by the Maximum Likelihood method outlined in Appendix A. Madsen et al. (1992a) report a number of estimation results for different types of buildings. These show that the long time constant in the transfer function from ambient air temperature to indoor air temperature for typical one family houses lies in the range from 24 hours to one week. This means that it is to be expected that the transfer function from ambient air temperature to overall heat consumption in a DH system also will be dominated by a time constant of the same order of magnitude. The relation between ambient air temperature and heat supply is actually also dependent on how the heat supply is controlled by the consumers. However, it is likely that the consumers control the heat supply to keep a stable indoor temperature - apart from night set-back etc. - meaning that the long-term dependence on the ambient air temperature is not influenced by the control.

The other time constant of the estimated models is typically below one hour. This means that there is also a fast reaction on variations in heat input from radiator and sun, and from changes in ambient air temperature as well. For the latter this fast influence is mostly due to the air leakages in the buildings. The fact that weather has both a diurnal and an annual cycle implies periodicity in the heat consumption with the very same periods.

HOT TAP-WATER

Another source of periodicity in the heat consumption is the behavior of the consumers themselves, when they are using hot tap-water. The use of hot tap-water at each consumer is by nature stochastic. However, at certain hours of the day the use of hot tap-water is more or less likely meaning that it can be assumed to follow a probabilistic distribution for each time of day, week and year. Moreover, it is also likely that the distributions for different consumers are similar implying that as the number of consumers is increasing, the total heat consumption due to hot tap-water consumption

will contain less relative variation around a mean value for each type of day, week and year. Consequently, the part of the consumption reflecting the use of hot tap-water, can be expected to have diurnal, weekly and annual cycles, and thus the same will apply to the total heat consumption.

2.1.2 TRANSMISSION AND DISTRIBUTION LINES

The heat is supplied to the subscribers of the system through a network of transmission and distribution lines. The DH water possesses the capacity of accumulating considerable amounts of energy, For instance, the distribution system in Esbjerg contains 31.500 m^3. An increase of, for instance, five degree centigrade of the water in the supply line (half of the water) corresponds to an increase in energy accumulation of 321 GJ. Comparing this to the usual heat consumption in the system, being 50-300 MJ/s, it is seen that this amount of energy corresponds to the consumption in 20-100 minutes.

Recognizing that the distribution system is a heat buffer and that varying supply temperature implies changes in the accumulated heat in the DH water gives that the heat supply not necessarily has to be controlled to match the heat demand in the system. That is, the fact that the distribution system is a heat buffer allows for differences - apart from heat losses - between heat supply and heat consumption. As such there is a possibility of using the distribution system in a control of the heat supply and the supply temperature in accordance with an optimal strategy.

The power supply to the pumps, and its distribution among pumps, is determining for the mass flows in the system. Each pump has a characteristic, which gives the flow rate and pressure increase dependent on the resistance in the system. This means that the flow rates in the system depend on the power supply to the pumps as well as on the heat consumption in a complicated way.

SYSTEM OPERATION IN FEEDBACK

It is clear from the previous discussion that a DH system is operated subject to different feedback mechanisms. First of all, the consumers operate their radiators to keep a stable indoor temperature. Secondly, the DH system is operated in such a way that all consumers have their heat demand fulfilled. This is obtained by control of the pumps in the system and by control of the supply temperature.

The pumps are controlled to keep the pressure difference between supply pipe and return pipe above a given level all over system. The pressure difference is necessary to make the water pass through consumer installations. The control of the pumps is typically implemented as an automatic feedback from on-line measurements of pressure differences at specific locations in the DH net. Certainly does the pressure drop along a distribution line depend on the heat consumption, for instance, is there a pressure drop over radiators. This means that there is feedback from heat consumption to power supply to the pumps, which, however, is complex and presumably nonlinear.

The supply temperature is normally controlled to secure a supply temperature at the consumers above a certain temperature, dependent on the ambient air temperature. The relation between minimum supply temperature at the consumers and ambient air temperature is typically as sketched in Figure 2.3, viz. a constant level for ambient air temperatures above a certain point and an increasing supply temperature for decreasing ambient air temperatures below this point. This curve form has to do with design of the heating systems in the buildings. The supply temperature may, however, also be controlled to accumulate additional energy in the distribution system either to meet expected near-future peaks in the heat consumption or to make use of varying production cost (optimization by redistribution in time of heat supply).

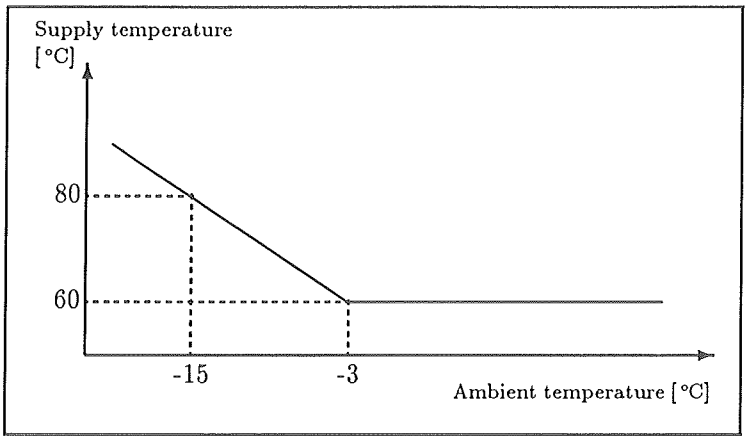


Figure 2.3: Typical curve for temperature demand at the consumer.

MODELLING OF THE DH NET

Benonysson (1991) investigates in detail physically based models for the propagation of temperature variations in a single pipe line. The result is that to describe the outlet temperature variation as a function of inlet temperature the following are of importance

1. The time of water propagation from inlet to outlet depends on the course of flow variations during the propagation time. As a first approximation the propagation time is assumed proportional to the inverse average flow during the propagation time.
2. The temperature drop is described sufficiently as decaying exponentially with the time of propagation. The decay is towards the temperature of the ambient medium.
3. The heat capacity of the pipe and the surroundings implies a low-pass filtering of the temperature variations which is sufficiently described by a first order model.

Benonysson (1991) forms a complete model for a transmission network with a number of attached heat exchanger stations by connecting single pipe models into a node model. In this approach it is required to know the complete flow distribution in the DH net. This in turn requires that the heat consumption at all heat exchanger stations is known.

Søgaard (1993) applies another approach for description of the relation between supply temperature at a supply point and the supply temperature at specific locations in a distribution net. In this approach the model is established by using statistical methods for identification and estimation of model parameters on data from the DH system in Esbjerg, Denmark. It turns out that the model describes the same dependencies as the physically based models. The statistical model has the structure

$$A_t(q^{-1})s_o(t) = q^{-k}B_t(q^{-1})s_i(t) + e(t), \tag{2.3}$$

$s_i(t)$ being the temperature at the supply point, and $s_o(t)$ the supply temperature at a specific location in the system. The parameters of A and B have to be time-varying primarily due to the varying flow in the system. q^{-1} is the backshift operator defined by $q^{-k}x(t) \equiv x(t-k)$.

Benonysson (1991) reaches to a model of the same structure for each pipe line, in which the parameters are determined from the distribution of consumption and flow in the overall transmission system. Søgaard (1993), on the other hand, presents statistical methods for estimation of model parameters and tracking of the time delay, k. This requires use of adaptive methods for simultaneous estimation of k and model parameters. The heat capacity of the pipe lines and the surroundings is described by the time constants of the autoregressive polynomium, and the temperature drop due to heat losses is described by the stationary gain being a little below one. For instance, a stationary gain of 0.98 in a model for absolute degrees centigrade measurements corresponds to a temperature drop of 2 °C if the supply temperature is 100 °C.

CONTROL OF SUPPLY TEMPERATURE

The delivery of heating through the transmission and distribution net takes place with the velocity of the water. This means that the water is delivered with considerable transport delays. Varying heat consumption can be fulfilled either by varying mass flow or by varying supply temperature. Changes in mass flow propagate with a velocity close to that of sound, whereas changes in supply temperature propagate with the velocity of the water. Consequently, when the variations of the heat consumption are met by variations in mass flow, the heat production does in practice follow the variations of the overall heat consumption of the system. There is, of course, a discrepancy due to the heat losses from the water on its way to the consumers.

For optimization purposes, however, it may be profitable to run the system in such a way that the supply temperature is at a minimum and/or the mass flow is at its maximum (determined by physical limitations of distribution system and pumps). This requires that the geographically distributed character of the consumption is taken into account, since temperature changes, appearing in concordance with the control strategy, reaches the consumers with differing transport delays. To implement the control strategy it is required to have a model, which describes how the heat production depends on a given course of variation of the supply temperature. Such a model must necessarily contain an, at least implicit, description of the geographically distributed consumption.

In Søgaard (1993) the *generalized predictive control* (GPC) strategy is discussed in detail with respect to control of supply temperature subject to a restriction of a maximum water flow in the transmission line leaving the power plant. Essentially the idea of this control strategy is to take into account not only the one-step ahead prediction, but instead prediction of a sequence of the process to be controlled.

In the present case the control variable is the supply temperature, and the process to be controlled is the water flow. Actually the system will not allow

the water flow to exceed the maximum due to the physical limitations, but if it happens that the consumers' demand of water flow add up to exceed the maximum, the result will be that the most distant consumers will not have their heat demand fulfilled. To prevent this situation to occur it will be necessary to start heat production at one or more peak load stations, if such back-up units exist. Accordingly, the purpose of the GPC controller is to avoid or minimize production at peak load stations, while at the same fulfilling the demands of the consumers.

Since the process to be controlled is the water flow a conversion from predicted heat production to predicted water flow has to be carried out, the conversion using future supply and return temperatures.

For the successful implementation of the GPC strategy for control of supply temperature it is of crucial importance to have a model giving a physically true description of the dependence of the required heat supply on supply temperature. When a model of this quality is at hand it can be used to give the prediction of the heat load as a function of the particular sequence of supply temperature, which are to be determined by the controller.

2.1.3 HEAT EXCHANGERS

Heat exchangers are used in DH systems for transfer of heat, typically from primary transmission net to secondary distribution net. Some systems, at least in Denmark, do not have heat exchangers. This is, for instance, the case for the DH system in Esbjerg. Heat exchangers are also used in domestic installations for heating of tap-water.

Chapter 3 contains a modelling study of a small-scale heat exchanger, which is normal in Swedish domestic installations. It turns out that the dominating time constant of this heat exchanger does not exceed 10 seconds – it depends on the flow through the heat exchanger. This means that the dynamics of such heat exchangers is negligible compared to the dynamics of buildings and transport delays of distribution systems.

Benonysson (1991) considers the dynamics of heat exchangers placed in the separation between transmission lines and secondary distribution nets in the DH system in Ishøj, Denmark. He concluded that it is sufficient to use the steady state relations of the heat exchangers in a modelling study based on 15 minutes sampling interval.

The estimation approach in Chapter 3 provides a method for determining the dynamics of a given heat exchanger. As such it can be used for assessing the dynamics of given heat exchangers in DH systems. This is, for instance, important when deciding whether it is required to model the dynamics of the heat exchangers or the steady state relations are sufficient.

3

Heat Exchanger Modelling in Continuous Time

A heat exchanger is a distributed system for which the dynamics may be described by physical conservation laws for mass, energy and momentum. One important property of a heat exchanger is that the dynamic response depends upon the operating points of the mass flows and temperatures. For instance, transport lags depend on mass flow, and the heat transfer coefficient is a function of both temperature and mass flow. If the purpose of modelling the heat exchanger is to obtain a description of the dynamic relations, being valid over a wide range of operating conditions, it is of crucial importance to take into account these, generally, nonlinear relations.

As the literature clearly demonstrates heat exchangers have been modelled in many ways, see e.g. Bittanti, Cividini, & Scattolini (1982), Humo & Popovic (1982) and Masada & Wormley (1982) for three basically different modelling approaches. Bittanti et al. (1982) use ARMAX modelling for the dependence of outlet temperatures on mass flows. Humo & Popovic (1982) apply the Laplace transformation with respect to time on partial differential equations for temperature, and obtain a solution using finite difference approximation.

The approach for deriving the model formulation in Masada & Wormley (1982) is the same as applied in this chapter. That is, physical considerations, viz. energy and mass conservation, along with empirical relations for

the heat transfer coefficient are applied in a compartmental model structure to give a system of ordinary differential equations. This approach has been used by many authors, e.g. Shoureshi & Paynter (1983), Gummérus (1988), Jonsson & Holst (1989), Palsson (1989) and Steiner (1989). In series of works (Jonsson, 1990), (Jonsson & Palsson, 1991),(Jonsson, Palsson, & Sejling, 1992) and (Jonsson & Palsson, 1993) different parameterizations of the compartmental model have been investigated.

Introducing additive random variables with distributional assumptions in these equations (process noise) and in measurement equations (measurement noise) results in a stochastic model formulation. In this chapter the parameters of the model are estimated by use of the maximum likelihood (ML) method.

3.1 THE BASIC HEAT EXCHANGER MODEL

Basically the model describes the heat exchanger as two flows of liquid or gas separated by a layer of metal. The fact that the heat exchanger is a distributed system prescribes that the heat transfer process is appropriately described by partial differential equations. However, the distributed character of the heat transfer can be approximated by introducing compartments in the system (lumping), i.e. the heat exchanger is divided into compartments of equal size, with the temperatures at the flow exit of the compartments being the model states.

This modelling approach is illustrated in Figure 3.1 for a counter flow heat exchanger, with n_s being the number of compartments on each side. $\dot{m}_h(t)$ and $\dot{m}_c(t)$ are the mass flows on the hot and the cold side, respectively. Similarly, $T_{h,in}(t)$, $T_{h,out}(t)$ and $T_{c,in}(t)$, $T_{c,out}(t)$ are the inlet and outlet temperatures on each side. $(T_{h,i}(t), T_{c,i}(t)), i \in [1; n_s]$ are the temperatures at the points in each compartment, where the flow leaves the respective hot and cold compartments, and $T_{m,i}(t), i \in [1; n_s]$ are the temperatures of the intermediate metal layer compartments. The compartments are all numbered in the direction of the hot water flow.

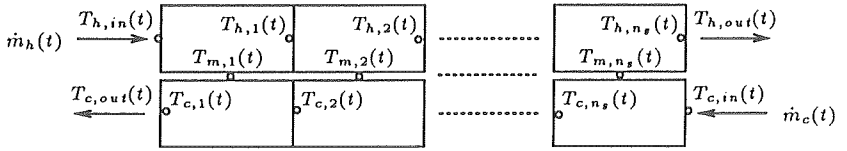


Figure 3.1: The compartmental counter flow heat exchanger model.

The main assumptions in forming the model, used and justified by many authors, (Ito & Masubuchi, 1978) and (Masada & Wormley, 1982), are

1. The heat transfer to the surroundings is negligible.
2. In the metal there is no heat conduction in the direction of the flow.
3. In the fluids there is no heat conduction at all.

In addition, in the present investigation it is assumed that there are no phase shifts in the fluids. Subject to these assumptions and by use of energy and mass balance equations for each compartment of the model, a system of differential equations can be formulated constituting the deterministic part of a dynamic model.

The first assumption for the model formulation is reasonable as heat exchangers usually are well isolated, meaning that heat losses are of minor importance compared to the heat transferred from the hot flow to the cold flow. The justification for the second and the third assumption is that in the directions of the flow, the heat transport of the water is of much greater importance than the heat conduction in the flow and in the metal.

The following paragraphs contain an account of the resulting differential equations for the two cases, where either the heat capacity of the intermediate metal layer is neglected or taken into account by the model.

The heat transfer between the hot and cold water is driven by the temperature difference. The temperature is, however, not the same in the length of each compartment, and accordingly, the issue is how the actual spatially distributed temperature difference for each pair of opposite compartments should be expressed by use of model states and inlet temperatures. Different possibilities for modelling the driving force, accounting for the heat transfer between opposite compartments, are discussed in a subsequent paragraph.

From the theory of heat transfer (Holman, 1981) it is well-known that the heat transfer coefficient depends on temperature and mass flow. In the final part of this section empirical relations for this dependence are discussed.

Section 3.2 describes how the differential equations with the physical relations of this section can be turned into a state space formulation, and how the parameters of this model can be estimated.

NEGLECTING THE HEAT CAPACITY OF THE METAL

Neglecting the effect of the metal heat capacity corresponds to a model in which the separation between the two flows has no heat resistance and no heat capacity. This gives for the compartments of number i the following differential equations

$$\frac{M_h}{n_s} c_h \frac{dT_{h,i}(t)}{dt} = \dot{m}_h(t) c_h [T_{h,i-1}(t) - T_{h,i}(t)] - \frac{A_h}{n_s} U \Delta T_i(t) \quad (3.1)$$

$$\frac{M_c}{n_s} c_c \frac{dT_{c,i}(t)}{dt} = \dot{m}_c(t) c_c [T_{c,i+1}(t) - T_{c,i}(t)] + \frac{A_c}{n_s} U \Delta T_i(t), \quad (3.2)$$

for i ranging from 1 to n_s. The inlet temperatures correspond to compartments 0 and $n_s + 1$, respectively, i.e. $T_{h,0}(t) = T_{h,in}(t)$ and $T_{c,n_s+1}(t) = T_{c,in}(t)$, and the outlet temperatures equal the temperatures of the last compartments, i.e. $T_{h,n_s}(t) = T_{h,out}(t)$ and $T_{c,1}(t) = T_{c,out}(t)$.

M_h and M_c are the masses of the flow media in the heat exchanger, A_h and A_c are the heat transfer areas, and c_h and c_c are the respective specific

heats for the flow media on the hot and cold side. In the case, where the flow is in liquid phase, the specific heat can reasonably be assumed to be constant. The discrimination between the specific heat for the two sides is motivated by the possibility of having different liquids on the two sides or even different phases (liquid/gas).

U is the overall coefficient for the heat transfer from one flow to the other, written, for simplicity, as an independent parameter though it actually is a function of mass flows and temperatures. $\Delta T_i(t)$ is the driving force for the heat transfer representing the temperature difference between each pair of hot and cold compartments. In the current case with no modelling of the heat capacity of the intermediate metal the energy transfer between two pairwise compartments leads to the equality $UA_h/n_s = UA_c/n_s$.

MODELLING THE HEAT CAPACITY OF THE METAL

The metal layer in between the two mass flows is taken into account by recognizing the heat capacity of the bulk of metal and introducing a differential equation for the temperature of each metal compartment. The heat resistance of the metal, however, is still assumed to be negligible. In this case the system of equations takes the form

$$\frac{M_h}{n_s} c_h \frac{dT_{h,i}(t)}{dt} = \dot{m}_h(t) c_h [T_{h,i-1}(t) - T_{h,i}(t)] - \frac{A_h}{n_s} h_h \Delta T_i^h(t) \quad (3.3)$$

$$\frac{M_m}{n_s} c_m \frac{dT_{m,i}(t)}{dt} = \frac{A_h}{n_s} h_h \Delta T_i^h(t) - \frac{A_c}{n_s} h_c \Delta T_i^c(t) \quad (3.4)$$

$$\frac{M_c}{n_s} c_c \frac{dT_{c,i}(t)}{dt} = \dot{m}_c(t) c_c [T_{c,i+1}(t) - T_{c,i}(t)] + \frac{A_c}{n_s} h_c \Delta T_i^c(t). \quad (3.5)$$

M_m is the mass of the metal between the flows, and c_m is the specific heat of the metal. Furthermore, h_h and h_c are the heat transfer coefficients between the hot flow and the cold flow, respectively, and the metal layer. Finally $\Delta T_i^h(t)$ and $\Delta T_j^c(t)$ represent the temperature differences, which constitute the driving forces for the heat transfer from the hot water to the metal and from the metal to the cold water, respectively.

It should be noted that Equations (3.1)-(3.2) as well as Equations (3.3)-(3.5) also apply to parallel flow heat exchangers by suitable reversals of compartment indices.

EXPRESSIONS FOR THE HEAT TRANSFER DRIVING FORCE

For the model, where the effect of the heat capacity of the metal is neglected Steiner (1989) and Palsson (1989) compared the following three possibilities for expressing the effective temperature difference that drives the transfer of heat from hot water compartments to cold water compartments

1. $\Delta T_i(t) = T_{h,i}(t) - T_{c,i}(t)$
2. $\Delta T_i(t) = \frac{1}{2}[T_{h,i-1}(t) + T_{h,i}(t)] - \frac{1}{2}[T_{c,i+1}(t) + T_{c,i}(t)]$
3. $\Delta T_i(t) = \dfrac{[T_{h,i-1}(t) - T_{c,i}(t)] - [T_{h,i}(t) - T_{c,i+1}(t)]}{\log([T_{h,i-1}(t) - T_{c,i}(t)]/[T_{h,i}(t) - T_{c,i+1}(t)])}.$

The first possibility corresponds to the temperature being constant in the full length of each compartment. As the temperature evidently varies in the direction of the flow - decreases on the hot side and increases on the cold side, at least for constant inlet temperatures and mass flows - it is obvious that applying the first possibility gives rise to an erroneous expression for the heat transfer. In steady state the expressed temperature difference, and hereby the heat transfer, will be too small, and for this reason it is to be expected that when estimating model parameters from data, the estimates will be biased towards values that will tend to neutralize the erroneous expression for the temperature difference. This is demonstrated in Jonsson & Holst (1989) and Palsson (1989). In the second possibility the temperature variation in the flow direction of each compartment is assumed to be linear, and the third possibility corresponds to an exponential temperature distribution in each compartment.

It is desirable that the driving force description of the model has a structure that allows the parameters to be estimated to coincide with what can be determined from steady state relations. If it is assumed that the tem-

peratures are exponentially distributed in the directions of the flow, which indeed is reasonable, the following steady state relations between inlet and outlet temperatures applies

$$\dot{m}_h c_h [T_{h,in} - T_{h,out}] = AU\Delta T_{LMTD} = \dot{m}_c c_c [T_{c,out} - T_{c,in}]. \qquad (3.6)$$

T_{LMTD} is the *log mean temperature difference* (Holman, 1981) given by

$$T_{LMTD} = \frac{(T_{h,in} - T_{c,out}) - (T_{h,out} - T_{c,in})}{\log\left((T_{h,in} - T_{c,out})/(T_{h,out} - T_{c,in})\right)}. \qquad (3.7)$$

Indeed, it can be shown that (3.6) can be satisfied for any number of compartments, when the third alternative is applied. Assuming linear temperature distribution in each compartment (the second alternative) gives a satisfying approximation with only a few compartments on each side, whereas for the assumption of constant temperature in each compartment (the first alternative) a greater number of compartments might be necessary to approximate the exponential distribution.

Using the third alternative implies that the driving force expressions are nonlinear in inlet temperatures and compartment temperatures. The consequence is that the differential equations for the compartment temperatures are nonlinear. Therefore the use of the standard Kalman filter for computation of state prediction and reconstruction is not appropriate as the recursions of the Kalman filter rely on the assumption of the differential equations being linear. The nonlinearity of the differential equations calls for the use of the techniques that take into account this nonlinearity, see Jonsson & Palsson (1993) for the application of the extended Kalman filter to this issue. To remain in the class of linear differential equations the second alternative is applied in the current investigation.

For the model with differential equations for the metal compartments the linear approximation to the temperature distributions gives the following driving force expressions

- $\Delta T_i^h(t) = \frac{1}{2}[T_{h,i-1}(t) + T_{h,i}(t)] - T_{m,i}(t)$
- $\Delta T_i^c(t) = T_{m,i}(t) - \frac{1}{2}[T_{c,i+1}(t) + T_{c,i}(t)]$.

EMPIRICAL RELATIONS FOR THE HEAT TRANSFER COEFFICIENT

The main reason for the nonlinearity in heat exchangers is the strong dependence of the heat transfer coefficients on mass flow and temperature. Holman (1981) gives a thorough treatment of this matter. In order to account for the nonlinearity, empirical relations for the heat transfer coefficients can be used and incorporated in the model parameters.

In the present case the empirical relation

$$Nu = CPr^x Re^y, \tag{3.8}$$

describing the dependence of the heat transfer coefficient on mass flow, temperature, dimensions and physical constants, is used. The values of C, x and y depend on the type of flow, e.g. turbulent or laminar flow. It is assumed that this relation applies to both of the mass flows of the heat exchanger for the same values of x and y.

Nu, Pr and Re are dimensionless expressions called Nusselt, Prandtl and Reynolds number, respectively. Inserting the expressions for Nu, Pr and Re gives

$$\frac{h(T,\dot{m})d}{k(T)} = C\left(\frac{c\mu(T)}{k(T)}\right)^x \left(\frac{u(\dot{m})d\rho}{\mu(T)}\right)^y. \tag{3.9}$$

$h(T,\dot{m})$ is the heat transfer coefficient, T is temperature and $\dot{m}$ is mass flow. $u(\dot{m})$ is the velocity given as the mass flow divided by the cross-sectional area and density, i.e. $u(\dot{m}) = \frac{\dot{m}}{\rho A_{cross}}$. $\mu(T)$ is the dynamic viscosity, and $k(T)$ is the thermal conductivity both with dependence on temperature. c is the specific heat (assumed to be independent of temperature and mass flow), d is an expression for the diameter of the cross section of the flow, and C is a constant.

By rearranging Equation (3.9) the expression for the heat transfer becomes

$$h(T,\dot{m}) = \underbrace{C\frac{c^x d^{(y-1)}}{A_{cross}^y}}_{i}\underbrace{\mu(T)^{(x-y)}k(T)^{(1-x)}}_{ii}\underbrace{\dot{m}^y}_{iii}. \tag{3.10}$$

The term (i) is constant, i.e. it is independent of the mass flows and temperatures. (ii) is temperature dependent, and (iii) is mass flow dependent.

Jonsson & Palsson (1993) proposes and reports results showing that the temperature dependence can be approximated reasonably well by a linear form giving the following expression for the heat transfer

$$h(T, \dot{m}) = K'(1 + bT)\dot{m}^y. \tag{3.11}$$

If the temperature dependence is neglected, this is simplified to

$$h(\dot{m}) = C'\dot{m}^y. \tag{3.12}$$

These expressions for the heat transfer coefficient apply to the heat transfer between one of the flows and the metal separation. This means that they apply to the models that take into account the heat capacity of the metal. In the models, where the heat capacity of the metal is neglected, however, it is necessary, when applying either of the heat transfer expressions, to recognize that there actually is one coefficient for the heat transfer between the hot water and the metal, and another coefficient for the heat transfer between the metal and the cold water. Furthermore, there is a thermal resistance in the metal between the two media, which, however, is considered to be negligible. This implies that an expression for the overall heat transfer coefficient U can be obtained as the serial relation for the resistance of the heat transfer from the hot water to the cold water given by

$$\frac{1}{U} = \frac{1}{h_h} + \frac{1}{h_c}. \tag{3.13}$$

3.2 STATE SPACE DESCRIPTION OF THE HEAT EXCHANGER

For both types of model structures (with and without modelling the effect of the metal), and for a given number of compartments the complete system

of differential equations can be written in state space form as

$$\frac{d}{dt}\boldsymbol{T}(t) = \boldsymbol{A}\boldsymbol{T}(t) + \boldsymbol{B}\boldsymbol{T}_{in}(t), \tag{3.14}$$

where $\boldsymbol{T}(t)$ is the state vector for the system of differential equations, and

$$\boldsymbol{T}_{in}(t) = \begin{bmatrix} T_{h,in}(t) \\ T_{c,in}(t) \end{bmatrix}. \tag{3.15}$$

$\boldsymbol{A}$ is the system matrix of dimensions $(2n_s, 2n_s)$ for the model without the effect of the metal and $(3n_s, 3n_s)$ for the model, where the heat capacity of the metal is taken into account. $\boldsymbol{B}$ is the input matrix of dimensions $(2n_s, 2)$ or $(3n_s, 2)$ depending on the option of modelling the metal. $\boldsymbol{A}$ and $\boldsymbol{B}$ are both functions of mass flows and compartment temperatures.

As an example, the state space formulation is drawn up for the differential equations corresponding to the model with three compartments on each side, for which the heat capacity is neglected, and where the linear approximation to the driving force is used. For the state vector

$$\boldsymbol{T}(t) = \begin{bmatrix} T_{h,3}(t) & T_{h,2}(t) & T_{h,1}(t) & T_{c,1}(t) & T_{c,2}(t) & T_{c,3}(t) \end{bmatrix},$$

and by introducing the parameters $\alpha_h = A_h U/3c_h$ and $\alpha_c = A_c U/3c_c$, the system matrix $\boldsymbol{A}$ is equal to

$$\begin{bmatrix}
-\frac{\dot{m}_h(t)}{M_h/3} - \frac{\alpha_h/2}{M_h} & \frac{\dot{m}_h(t)}{M_h/3} - \frac{\alpha_h/2}{M_h} & 0 & 0 & 0 & \frac{\alpha_h/2}{M_h} \\
0 & -\frac{\dot{m}_h(t)}{M_h/3} - \frac{\alpha_h/2}{M_h} & \frac{\dot{m}_h(t)}{M_h/3} - \frac{\alpha_h/2}{M_h} & 0 & 0 & \frac{\alpha_h/2}{M_h} \\
0 & 0 & -\frac{\dot{m}_h(t)}{M_h/3} - \frac{\alpha_h/2}{M_h} & \frac{\alpha_h/2}{M_h} & \frac{\alpha_h/2}{M_h} & 0 \\
0 & 0 & \frac{\alpha_c/2}{M_c} & -\frac{\dot{m}_c(t)}{M_c/3} - \frac{\alpha_c/2}{M_c} & \frac{\dot{m}_c(t)}{M_c/3} - \frac{\alpha_c/2}{M_c} & 0 \\
0 & \frac{\alpha_c/2}{M_c} & \frac{\alpha_c/2}{M_c} & 0 & -\frac{\dot{m}_c(t)}{M_c/3} - \frac{\alpha_c/2}{M_c} & \frac{\dot{m}_c(t)}{M_c/3} - \frac{\alpha_c/2}{M_c} \\
\frac{\alpha_c/2}{M_c} & \frac{\alpha_c/2}{M_c} & 0 & 0 & 0 & -\frac{\dot{m}_c(t)}{M_c/3} - \frac{\alpha_c/2}{M_c}
\end{bmatrix}$$

and the input matrix is

$$\boldsymbol{B} = \begin{bmatrix} 0 & \frac{\alpha_h/2}{M_h} \\ 0 & 0 \\ \frac{\dot{m}_h(t)}{M_h/3} - \frac{\alpha_h/2}{M_h} & 0 \\ \frac{\alpha_c/2}{M_c} & 0 \\ 0 & 0 \\ 0 & \frac{\dot{m}_c(t)}{M_c/3} - \frac{\alpha_c/2}{M_c} \end{bmatrix}.$$

The heat exchanger model as given by (3.14) is deterministic. To allow for variations between the model and the true temperatures, a noise term is added. The model is then described by a stochastic differential equation and written as

$$d\boldsymbol{T}(t) = \boldsymbol{A}\boldsymbol{T}(t)dt + \boldsymbol{B}\boldsymbol{T}_{in}(t)dt + \boldsymbol{w}dt, \tag{3.16}$$

where $\boldsymbol{w}(t)$ is assumed to be a Wiener process with the incremental covariance $\boldsymbol{R}_w dt$. Stochastic differential equation models are described in, e.g. Åström (1970) and Gard (1988).

The measurements at sample instants $t_i, i = 1, \ldots, n$, are given by

$$\boldsymbol{T}_r(t_i) = \begin{bmatrix} T_{h,n_s}(t_i) \\ T_{c,1}(t_i) \end{bmatrix} + e(t_i) = \boldsymbol{C}\boldsymbol{T}(t_i) + e(t_i), \tag{3.17}$$

where $\boldsymbol{C}$ is a matrix of dimensions $(2, 2n_s)$ or $(2, 3n_s)$. In the above mentioned model the observation matrix will be given by

$$\boldsymbol{C} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

as the outlet temperatures are equal to $T_{h,3}(t)$ and $T_{c,1}(t)$.

The measurement errors $\{e(t_i)\}$ are assumed to be zero mean, normal i.i.d. random variables with covariance $\boldsymbol{R}_e$. It is further assumed that $\boldsymbol{w}(t)$ and

$e(t_i)$ are mutually independent. Finally inlet temperatures and mass flows are assumed to be constant throughout each sampling interval.

Subject to the above given assumptions the parameters can be estimated using the ML method as outlined in Appendix A. To be able to maximize the likelihood function for a given set of data, it is necessary to the convert the model into a time-discrete formulation that complies with the sampling intervals in data. This is obtained by integrating the continuous time model through the sampling intervals - not necessarily of equal length - which relying on the assumptions of constant input and constant $\boldsymbol{A}$ and $\boldsymbol{B}$ gives

$$\boldsymbol{T}(t+\Delta t) = \Phi(\Delta t)\boldsymbol{T}(t) + \Gamma(\Delta t)\boldsymbol{U}(t) + \boldsymbol{v}(t;\Delta t). \tag{3.18}$$

Applying the Kalman filter on this model results in a set predictions and prediction errors, which then is used in the calculation of the (log)-likelihood function. That is, for a given set of parameters in $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{R}_w$ and $\boldsymbol{R}_e$ the likelihood function is calculated, and accordingly the ML-estimate of the parameters is found by numerical minimization of the negative log-likelihood function. There exists, as outlined in Appendix A, the possibility of estimating the entries of the stationary Kalman gain $\boldsymbol{K}$ instead of the covariance matrices $\boldsymbol{R}_w$ and $\boldsymbol{R}_e$.

To start off the calculations of the Kalman filter initial values of mean and covariance of the state vector, $\boldsymbol{\mu}_0$ and $\boldsymbol{V}_0$, are required. In the current implementation $\boldsymbol{\mu}_0$ is estimated in parallel with the other parameters of the model, and $\boldsymbol{V}_0$ can be supplied, for instance, as a diagonal matrix with elements equal to 10. When the entries of the stationary Kalman gain are estimated instead of $\boldsymbol{R}_w$ and $\boldsymbol{R}_e$, and the recursions from the first step are carried out using the stationary Kalman gain, the initial value of the state covariance $\boldsymbol{V}_0$ is not necessary.

3.3 DATA

The heat exchanger used in this study is a small water-to-water counter flow plate heat exchanger from Alfa-Laval, type CB25, with 14 plates. It is part

of a heat exchanger setup located at the Department of Heat and Power Engineering, Lund Institute of Technology, University of Lund, Sweden.

In the technical description of the heat exchanger (Alfa-Laval, 1985) is given that for 14 plates the dimensions are $304 \times 103 \times 56$ mm, and the total weight of the metal is 2.4 kg. The heat exchanger has six layers of hot water alternating between seven layers of cold water. Each layer contains 0.05 l of water meaning that the mass of the hot water is close to 0.3 kg, and the mass of the cold water is close to 0.35 kg.

Figure 3.2 and Figure 3.3 show the set of data used in this chapter. The heat exchanger was excited by changing both of the mass flows as well as the cold inlet temperature. The sampling interval was one second and constant throughout the experiment, and it was carried out using the highest possible sampling rate for the measuring equipment of the test setup. The temperature measurements are in degrees centigrade, and the accuracy is in the first decimal, and the flow measurements are in kg/s with the accuracy in the third decimal. For this set of data the transport delay through the heat exchanger varies between 1.5 and 3 seconds for the hot water and between 0.8 and 1.3 seconds for the cold water. In the estimation of models only measurements with excitation is used, i.e. measurements for $t_i \in [21; 510]$.

The dynamics of the heat exchanger is rather fast due to its small dimensions. For instance, a second order model is expected to have two time constants lying in the range of 1-10 seconds. As the sampling rate is close even to the dominating time constant of the heat exchanger, it is evident that some high frequency variation is removed.

The dynamics of the temperature sensors has been assessed in an experiment, where the sensors were dropped into a water bath. Assuming that the actual temperature hereby shows a perfect step, the step response of each of the sensors can be estimated. Approximating this step response by that of a first order system with unit gain implies that the dynamics of each sensor can be represented by one time constant, see Palsson (1990). Such

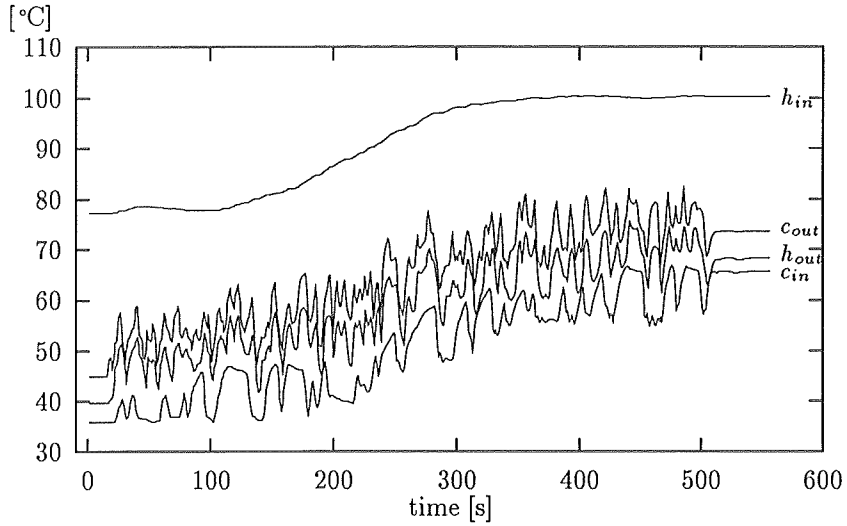


Figure 3.2: Temperature observations.

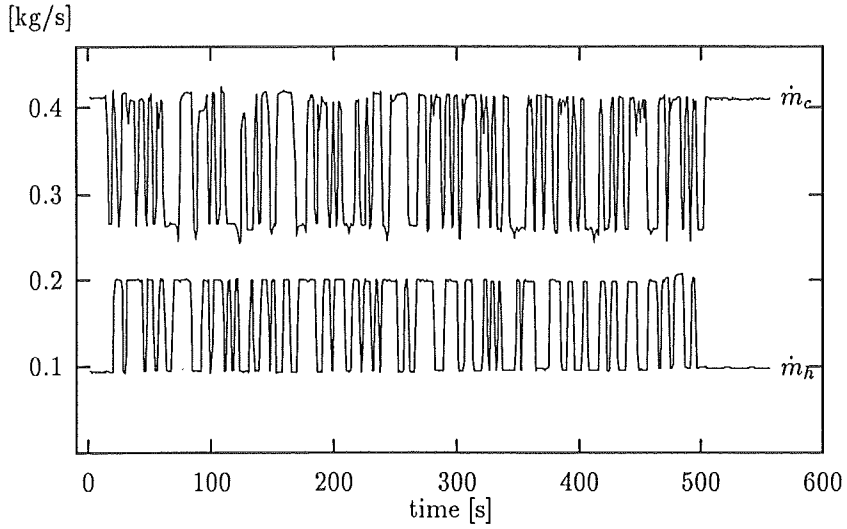


Figure 3.3: Mass flow observations.

experiments showed for each of the four sensors different time constants ranging from 0.73 to 1.5 seconds. This actually means that each of the four temperature sequences has passed through a filter with individual low-pass properties and phase shifts. Furthermore, the fact that the dominating time constant of the temperature sensors is approximately the same as the sampling interval implies that the resulting observations will be subject to a substantial low-pass filtering.

To obtain comparable measurements, it is assumed that the first order approximation of the sensors are applicable. Hereby it is given that the measurements has been passed through the first order model

$$T_f(t) = \frac{1-\varphi}{1-\varphi q^{-1}} T(t), \tag{3.19}$$

where the parameter is given by

$$\varphi = \exp\left(-\frac{\Delta t}{\tau}\right) \tag{3.20}$$

for sampling interval Δt and time constant τ. Hence, the back-filtered measurements are found as

$$T^*(t) = \frac{1-\varphi q^{-1}}{1-\varphi} T_f(t). \tag{3.21}$$

It is clear that the high-frequency information lost is not regained by doing this back-filtering. However, the differing phase shifts introduced by the sensors are compensated for. Before used for estimation, the temperature has been subject to this back-filtering.

3.4 ESTIMATION RESULTS

In the investigation of which model, within the proposals outlined above, appropriately describes the heat exchanger data, there are a number of alternatives with respect to model structure and parameterizations. First

of all, the alternative of modelling the metal separation or not exists. Secondly, there is the option of modelling the heat transfer coefficient as a constant or as being dependent on mass flow as given by (3.12). Combining these two options results in four model possibilities to be investigated

1. Neglecting the effect of the metal and neglecting the dependence of the heat transfer on mass flow.
2. Modelling the effect of the metal but neglecting the dependence of the heat transfer on mass flow.
3. Neglecting the effect of the metal but modelling the dependence of the heat transfer on mass flow.
4. Modelling the effect of the metal and modelling the dependence of the heat transfer on mass flow.

The estimation results for these four model approaches are reported in the following four sections. The dependence on temperature is investigated in Section 3.4.6 along with another proposal for model improvement consisting of the introduction of dependence on mass flow gradients.

During the investigation it turned out that estimation of a stationary Kalman gain matrix resulted in a higher maximum likelihood than estimation of $\boldsymbol{R}_w$ and $\boldsymbol{R}_e$ as diagonal matrices. Appendix A outlines how the parameters are estimated for both alternatives.

Actually the conditions for the convergence of Kalman gain to a stationary matrix are not fulfilled as the model is time-varying due to the dependence on mass flows and temperatures. The reason that it none the less payed off to estimate a stationary Kalman gain may be that the feedback from prediction errors to model states as a good approximation can be considered to be independent of the operating points.

The choice of estimating the stationary Kalman gain matrix rules out the possibility of assessing the elements of the covariance matrices of the model. However, generally the number of parameters of the covariance matrices exceeds the number of elements in the Kalman gain matrix. The implication

is that it is not possible to estimate $\boldsymbol{R}_w$ and $\boldsymbol{R}_e$ fully parameterized, and moreover, it is generally not possible to determine the restricted parameterization relating to the parameters of the Kalman gain matrix. When estimating $\boldsymbol{R}_w$ and $\boldsymbol{R}_e$ it is therefore necessary to impose restrictions on the matrix entries, e.g. by setting off-diagonal elements equal to zero.

It also turned out that it was profitable, with respect to likelihood maximization, to depart from the relation between input and output in the discretized model, prescribing that reactions to input variations are not seen in the output until the subsequent sample.

The model formulation is based on the assumption that the input variables remain constant during the sampling interval and equal to the sample values at the start of the interval. This means that the discretized model relates the output at one sample to the input at the previous sample. However, the assumption of constant input during the sampling interval is not fulfilled in the data used in the current investigation. Considerable changes in the input do occur between sample point, and moreover, the outlet temperatures can show almost immediate reactions to changes in inlet temperatures. As the model relates output to lagged input, it is the case that the model is not capable of describing immediate reactions to changes in input, at least when the changes occur between sampling points. Consequently, the predicted output reaction to changes in input may have a delay of a length up to one sample interval. The delay depends on where in between the sample points the change in input has occurred.

The improvement in likelihood result was obtained by changing the discretized model in such a way that the output is related to simultaneous input. This description corresponds to the assumption that all changes between sample points should have occurred immediately after sampling instead of immediately before. It is clear that neither of these assumptions are in accordance with reality as the inlet temperatures and mass flows do vary between sample points.

Further investigations showed that it was possible to obtain an even better likelihood result if the simultaneous input-output relation for the cold outlet and the original relation for the hot outlet were used. The explanation to this result is probably that the hot inlet shows a smooth variation, whereas the cold inlet is excited in a step-wise manner. The use of different input-output relations for the two flows implied, however, that the parameter estimates in some cases turned out to be unrealistic. Therefore, it is preferred to accept the (slightly) inferior likelihood results to obtain more credible parameter estimates. Indeed, it turned out that the parameter estimates depended on the applied input-output relation of the model.

In the case, where one or more estimated parameters turned out to be insignificant using a t-test on 5% significance level, the model was estimated over again with the parameter with the lowest t-ratio fixed at zero. This was continued until all estimates were significant on 5% level. The parameters fixed at zero are marked by (f) in the tables containing the estimation results. In the tables are also given Bayes Information Criterion (Schwarz, 1978)

$$BIC = -2 \log L + n_p \log N, \tag{3.22}$$

where $\log L$ is the log-likelihood, n_p is the number of estimated parameters, and N is the number of observations used in the estimation. BIC is applicable to the determination of the appropriate model order.

3.4.1 NO MODELLING OF METAL AND USING CONSTANT HEAT TRANSFER

For the first and most simple choice of modelling approach the differential equations in (3.1)-(3.2) are used with the heat transfer coefficient U being constant. For this model formulation it is recognized that four parameters in the differential equations can be estimated. In this investigation the four parameters are chosen as the masses of the amount of water in each

compartment, M_h and M_c, along with the two following parameters

$$\alpha_h = \frac{A_h U}{c_h} \quad \text{and} \quad \alpha_c = \frac{A_c U}{c_c}. \tag{3.23}$$

α_h and α_c express the time gradient for the temperature of the water mass for each side, respectively, of the heat exchanger per unit effective temperature difference (driving force). That is, these parameters are expressions for the cooling of the hot water mass and the heating of the cold water for an effective temperature difference of 1 °C.

Using the linear form of the effective temperature difference between opposite compartments gives for this parameterization the following set of differential equations

$$\begin{aligned} \frac{dT_{h,i}(t)}{dt} &= \frac{\dot{m}_h(t)}{M_h/n_s}[T_{h,i-1}(t) - T_{h,i}(t)] - \\ &\quad \frac{\alpha_h}{2M_h}\left([T_{h,i-1}(t) + T_{h,i}(t)] - [T_{c,i+1}(t) + T_{c,i}(t)]\right) \quad (3.24) \\ \frac{dT_{c,i}(t)}{dt} &= \frac{\dot{m}_c(t)}{M_c/n_s}[T_{c,i+1}(t) - T_{c,i}(t)] + \\ &\quad \frac{\alpha_c}{2M_c}\left([T_{h,i-1}(t) + T_{h,i}(t)] - [T_{c,i+1}(t) + T_{c,i}(t)]\right). \quad (3.25) \end{aligned}$$

For a given number of compartments these equations can be introduced into a state space model, as previously illustrated for $n_s = 3$.

The estimation results with one, two and three compartments are shown in Table 3.1. In the table is also listed the estimated entries of the stationary Kalman gain, where, e.g. $\boldsymbol{K}_{h_1,h_{out}}$ is the gain, which multiplied on the hot outlet prediction error, determines the update of the state temperature $T_{h,1}(t)$. The numbers in brackets are the estimated standard deviations of the parameter estimates, and an (f) marks that parameter has been fixed at the shown value.

For the model with only one compartment on each side the estimates of M_h and M_c are unrealistic. This is not surprising as a model with only one compartment on each side does not distinguish between a counter flow and

Statistic	Estimation Result		
	$n_s = 1$	$n_s = 2$	$n_s = 3$
α_h [kg/s]	0.29 (.02)	0.31 (.005)	0.32 (.006)
α_c [kg/s]	0.29 (.007)	0.34 (.007)	0.30 (.004)
M_h [kg]	4.1 (.9)	0.97 (.008)	0.97 (.02)
M_c [kg]	1.0 (.07)	0.95 (.01)	0.94 (.04)
$K_{h_3,h_{out}}$			2.4 (.2)
$K_{h_3,c_{out}}$			-0.6 (.1)
$K_{h_2,h_{out}}$		2.2 (.09)	0 (f)
$K_{h_2,c_{out}}$		-0.2 (.08)	3.5 (.3)
$K_{h_1,h_{out}}$	1.4 (.05)	-1.6 (.5)	3.5 (1.1)
$K_{h_1,c_{out}}$	0.4 (.05)	4.7 (.5)	-7.3 (1.7)
$K_{c_1,h_{out}}$	-0.5 (.09)	1.2 (.2)	-0.4 (.2)
$K_{c_1,c_{out}}$	2.1 (.08)	0.4 (.2)	3.1 (.2)
$K_{c_2,h_{out}}$		-5.0 (.5)	0 (f)
$K_{c_2,c_{out}}$		3.9 (.4)	0 (f)
$K_{c_3,h_{out}}$			-7.4 (.7)
$K_{c_3,c_{out}}$			7.8 (.6)
$-\log L$	1359.5	1205.1	1169.8
BIC	2781	2509	2457
$\hat{\sigma}^2_{\varepsilon_h}$ [°C^2]	1.00	0.56	0.48
$\hat{\sigma}^2_{\varepsilon_c}$ [°C^2]	1.65	0.95	0.96

Table 3.1: No modelling of metal and using constant heat transfer.

a parallel flow heat exchanger indicating that a one-compartment model is insufficient for a description of the dynamics and the distributed character of the system.

The ratio between M_h and M_c is from the physics expected to be 6/7. This is far from the case for $n_s = 1$. For the higher order models the mass estimates are close, but there is not seen any indication of the expected mass ratio. The estimates of M_h and M_c are considerably above the actual water masses, but this is not surprising as there is no explicit modelling of the 2.4 kg of metal, meaning that the mass of the metal is likely to contribute to the water mass estimates. For the α-parameters are seen the nice results that the estimates pairwise are close to each other and that they seem to be invariable with n_s.

The parameter estimates may also be compared to those found in Jonsson & Palsson (1993), where the same set of data was used for least squares estimation, with equally weighting of hot outlet and cold outlet prediction errors, of the same model. For $n_s = 2$ their estimates of M_h and M_c are 0.66(.03) kg and 0.48(.03) kg, and their estimates of α_h and α_c are 0.31(.003) kg/s and 0.32(.003) kg/s. Note that it has been necessary to rescale the estimates and their standard deviations given in Jonsson & Palsson (1993) to make them comparable to the present estimation results. It is seen that the estimates of the water masses are considerably different. The estimates in Jonsson & Palsson (1993) are closer to the masses given from the technical description of the heat exchanger, but on the other hand the estimates found here are pairwise almost the same. Actually it is difficult to reason from physical considerations, what is credible water mass estimates as the model does not take into account the mass of the metal. However, it is clear that each of the water mass estimates should reasonably be found in the interval between the given masses 0.3-0.35 kg and 1.5-1.55 kg, corresponding to the water masses plus half the metal mass. The estimates of the α-parameters are almost the same for the two estimation methods, which is nice as these are expressions for the overall heat transfer coefficient.

The estimated variances of the prediction errors show that the fit is better for the hot outlet than for the cold outlet. The possible explanation is the difference in excitation of inlet temperatures.

3.4.2 MODELLING THE METAL BUT USING CONSTANT HEAT TRANSFER

When the metal of the heat exchanger is taken into account, by introduction of differential equations for intermediate metal compartments, there are six parameters to estimate. Apart from the two mass parameters of the model without metal, viz. M_h and M_c, the following four parameters are used

$$\alpha'_h = \frac{A_h h_h}{c_h} \quad \text{and} \quad \alpha'_c = \frac{A_c h_c}{c_c}, \tag{3.26}$$

$$\gamma_h = \frac{M_m c_m}{c_h} \quad \text{and} \quad \gamma_c = \frac{M_m c_m}{c_c}. \tag{3.27}$$

M_m is the mass of the metal in the heat exchanger. The α'-parameters in (3.26) correspond to the α-parameters in (3.23), where the overall heat transfer coefficients have been substituted by the separate heat transfer coefficients applying to each flow in the heat exchanger. The α-parameters are, as previously, expressions for the heat flow, in this case, however, from the hot flow to the metal (α_h), and from the metal to the cold flow (α_c). The γ-parameters correspond to the mass of the metal scaled by the ratio between the specific heat of the metal and the specific heat of the two flows, respectively. This gives the following set of differential equations

$$\begin{aligned} \frac{dT_{h,i}(t)}{dt} &= \frac{\dot{m}_h(t)}{M_h/n_s}[T_{h,i-1}(t) - T_{h,i}(t)] - \\ &\quad \frac{\alpha'_h}{M_h}\left(\tfrac{1}{2}(T_{h,i-1}(t) + T_{h,i}(t)) - T_{m,i}(t)\right) \end{aligned} \tag{3.28}$$

$$\begin{aligned} \frac{dT_{m,i}(t)}{dt} &= \frac{\alpha'_h}{\gamma_h}\left(\tfrac{1}{2}(T_{h,i-1}(t) + T_{h,i}(t)) - T_{m,i}(t)\right) - \\ &\quad \frac{\alpha'_c}{\gamma_c}\left(T_{m,i}(t) - \tfrac{1}{2}(T_{c,i+1}(t) + T_{c,i}(t))\right) \end{aligned} \tag{3.29}$$

$$\frac{dT_{c,i}(t)}{dt} = \frac{\dot{m}_c(t)}{M_c/n_s}[T_{c,i+1}(t) - T_{c,i}(t)] - \frac{\alpha'_c}{M_c}\left(\tfrac{1}{2}(T_{c,i+1}(t) + T_{c,i}(t)) - T_{m,i}(t)\right). \quad (3.30)$$

For this model structure the estimation did not succeed. That is, both for $n_s = 1$ and $n_s = 2$ did some of the parameter estimates approach extremes either close to zero or unreasonably large values. This result is an indication that it is not possible estimate the parameters for the given structure of the model.

To analyze, whether the parameterization of this model structure is identifiable or not, the transfer functions from each input to each output are formulated using the six parameters of the differential equations (3.28)-(3.30). These transfer functions can be obtained from the state space model as

$$G(p) = C\,(pI - A)^{-1}\,B, \quad (3.31)$$

where p is the differentiation operator. These transfer functions are calculated for the model with one compartment for each flow ($n_s = 1$). By way of example, the transfer function from the hot inlet temperature to the hot outlet is given by the numerator polynomium

$$N(p) = \left(\left(p + \frac{\alpha'_h}{\gamma_h} + \frac{\alpha'_c}{\gamma_c}\right)\left(p + \frac{\dot{m}_c(t)}{M_c} + \frac{\alpha'_c}{2M_c}\right) - \frac{{\alpha'_c}^2}{2M_c\gamma_c}\right)\left(\frac{\dot{m}_h(t)}{M_h} - \frac{\alpha'_h}{2M_h}\right), \quad (3.32)$$

and the denominator

$$D(p) = \left(p + \frac{\dot{m}_h(t)}{M_h} + \frac{\alpha'_h}{2M_h}\right)\left(p + \frac{\dot{m}_c(t)}{M_c} + \frac{\alpha'_c}{2M_c}\right)\left(p + \frac{\alpha'_h}{\gamma_h} + \frac{\alpha'_c}{\gamma_c}\right) - \frac{{\alpha'_h}^2}{2M_h\gamma_h}\left(p + \frac{\dot{m}_c(t)}{M_c} + \frac{\alpha'_c}{2M_c}\right) - \frac{{\alpha'_c}^2}{2M_c\gamma_c}\left(p + \frac{\dot{m}_h(t)}{M_h} + \frac{\alpha'_h}{2M_h}\right). \quad (3.33)$$

By comparing the parameterization of all four transfer functions to the parameters, which actually can be observed, that is, the parameters in

$$G_{i,j}(p) = \frac{b_2^{(i,j)}p^2 + b_1^{(i,j)}p + b_0^{(i,j)}}{p^3 + a_2p^2 + a_1p + a_0}, \quad i \in (1,2), \quad j \in (1,2), \quad (3.34)$$

results in a set of nonlinear equations. This gives 15 equations expressing the relation between each of the parameters in (3.34) and the six parameters of the model parameterization. The crucial question is, whether the six parameters can be determined from the 15 parameters in (3.34).

Tedious manipulations give the result that there does not exist an infinite set of parameter values in the model parameterization for all of which the same input-output relations in the four transfer functions are obtained. This conclusion is based on a calculation of the rank of the matrix containing the derivatives of the relations between the parameters in (3.34) with respect to the six parameters in the model. This actually means that the dimension of the space spanned by the fifteen parameters is six, when it is projected down onto the space spanned by the six parameters.

The fact that the dimension of the model parameterization is six imposes severe restrictions on the 15 parameters of the four transfer functions of the model. This in fact may be the reason for the estimation problems meaning that the present model for the heat exchanger is not capable of giving a dynamic description, which matches the actual relations observed for the heat exchanger. That is to say, the restriction of the parameterization of the four transfer functions may have the implication that the actually observed input-output relations cannot be matched by the model. It may even be so that the parameterization only can provide a bad approximation to the actual relations, and as a consequence of that is it not possible to obtain reasonable estimates.

The conclusions, drawn from the analysis of the identifiability of the six model parameters, it is not disturbed by the fact that the varying mass flows enter the coefficients in the four transfer functions. The analysis is actually carried out as if the mass flows have constant values, and the parameters in (3.34) are constant. Time-variation of the mass flows does not imply that the calculated rank of the matrix described above is below six, at least not when the mass flows are different from zero. The time-variation does, however, require that the parameters of the transfer functions, which correspond to the actually observable dynamics, also are time-varying, if

the postulated model and the heat exchanger should match. Indeed, it is the case that the dynamics of the heat exchanger depend on the mass flows, but if the mass flow dependence actually is different from the parameterization of the model, it may an explanation to the estimation problems.

3.4.3 NO MODELLING OF METAL BUT USING MASS FLOW DEPENDENT HEAT TRANSFER

Modelling of the mass flow dependence of the heat transfer is obtained by assuming that $h = C'\dot{m}^y$ applies to both mass flows, in this investigation with the same value of y. Since the metal layer is neglected the serial representation for the overall heat transfer coefficient must be applied, resulting in the following expression for the overall heat transfer coefficient

$$U(\dot{m}_h(t), \dot{m}_c(t), y) = \left(\frac{1}{C'\dot{m}_h^y(t)} + \frac{1}{C'\dot{m}_c^y(t)}\right)^{-1} = C' \frac{\dot{m}_h^y(t)\dot{m}_c^y(t)}{\dot{m}_h^y(t) + \dot{m}_c^y(t)}. \quad (3.35)$$

Insertion of this in the parameters of the model, where the heat transfer is given as a constant parameter (3.23) results in five parameters to be estimated. M_h and M_c are the same mass parameters as used previously. The exponent y in the mass flow dependence is a new parameter, and yet another set of α-parameters, written with two primes, are

$$\alpha_h'' = \frac{A_h C'}{c_h} \quad \text{and} \quad \alpha_c'' = \frac{A_c C'}{c_c}. \quad (3.36)$$

These α-parameters can be interpreted as the time gradients for the temperature of the water mass on each side of the heat exchanger per unit effective temperature difference (driving force) and per unit of the flow expression $\dot{m}_h^y(t)\dot{m}_c^y(t)/(\dot{m}_h^y(t) + \dot{m}_c^y(t))$.

The system of differential equations for this model is obtained by interchanging α_h with $\alpha_h''\dot{m}_h^y(t)\dot{m}_c^y(t)/(\dot{m}_h^y(t) + \dot{m}_c^y(t))$ and α_c with $\alpha_c''\dot{m}_h^y(t)\dot{m}_c^y(t)/(\dot{m}_h^y(t) + \dot{m}_c^y(t))$ in the equations, where the heat transfer coefficients has been assumed to be independent of mass flows, (3.24)-(3.25).

Statistic	Estimation Result					
	$n_s = 1$		$n_s = 2$		$n_s = 3$	
α_h''	1.89	(.05)	1.55	(.03)	1.51	(.02)
α_c''	1.97	(.05)	1.60	(.03)	1.57	(.03)
M_h [kg]	0.71	(.03)	0.63	(.02)	0.61	(.02)
M_c [kg]	0.48	(.01)	0.49	(.01)	0.59	(.01)
y	0.78	(.02)	0.60	(.01)	0.57	(.01)
$K_{h_3,h_{out}}$					2.4	(.2)
$K_{h_3,c_{out}}$					-0.6	(.2)
$K_{h_2,h_{out}}$			2.0	(.1)	0	(f)
$K_{h_2,c_{out}}$			0.3	(.1)	4.1	(.6)
$K_{h_1,h_{out}}$	1.4	(.08)	-1.3	(.5)	0	(f)
$K_{h_1,c_{out}}$	0.3	(.07)	-2.5	(1.2)	-11.5	(1.1)
$K_{c_1,h_{out}}$	-1.4	(.1)	1.8	(.4)	-2.1	(1.0)
$K_{c_1,c_{out}}$	2.8	(.1)	0	(f)	5.2	(.5)
$K_{c_2,h_{out}}$			-3.1	(.7)	3.1	(1.2)
$K_{c_2,c_{out}}$			5.5	(.7)	0	(f)
$K_{c_3,h_{out}}$					-5.3	(.8)
$K_{c_3,c_{out}}$					6.8	(.7)
$-\log L$	1110.3		927.0		830.5	
BIC	2289		1953		1785	
$\hat{\sigma}^2_{\varepsilon_h}$ [°C^2]	0.79		0.68		0.63	
$\hat{\sigma}^2_{\varepsilon_c}$ [°C^2]	0.85		0.61		0.48	

Table 3.2: No metal modelling but modelling of mass flow dependent heat transfer.

Comparing Table 3.2 with Table 3.1 shows that modelling of the flow dependence of the heat transfer coefficient results in a clear improvement in likelihood and BIC. It is, however, remarkable that the hot outlet prediction error variance, contrary to the results here, is lower for the model with constant heat transfer coefficient (for $n_s = 2, 3$). A possible explanation is that y preferably could be modelled with different values for the two sides. This is not tried in the present investigation.

For the estimated models the lowest BIC is obtained for $n_s = 3$. All the parameters seem to vary with n_s, however, in such a way that they all tend to approach a limit for increasing number of compartments. For this parameterization it is also seen that the water mass estimates exceed the actual water masses in the heat exchanger; a result which most likely can be ascribed to neglecting the existence of the metal.

Jonsson & Palsson (1993) found for the same model with $n_s = 2$ the mass estimates $\hat{M}_h = 0.58(.02)$ and $\hat{M}_c = 0.52(.03)$, which are not that far from the estimates found here. Similarly, their estimate of y is 0.62(.03), which also compares nicely with the present y-estimate. Appropriate rescaling of the parameters for the heat transfer coefficient in Jonsson & Palsson (1993) gives that their estimates correspond to $\hat{\alpha}''_h = 1.58(.02)$ and $\hat{\alpha}''_c = 1.65(.02)$. This is also reasonably in accordance with the estimates found here.

3.4.4 MODELLING THE METAL AND USING MASS FLOW DEPENDENT HEAT TRANSFER

By introducing the mass flow dependence of the heat transfer, $h = C'\dot{m}^y(t)$, in the α'-parameters of the model (3.26), where the heat transfer coefficient is considered to be independent of mass flow, gives the requested model. In this model seven parameters in $\boldsymbol{A}$ and $\boldsymbol{B}$ are used. α''_h and α''_c are the same as in (3.36). Likewise M_h and M_c remain the same, and γ_h and γ_c are given in (3.27). Finally the exponent of the mass flow dependence, y, is estimated. Interchanging α'_h and α'_c in Equations (3.28)-(3.30) with $\alpha''_h \dot{m}_h^y(t)$

and $\alpha_c''\dot{m}_c^y(t)$, respectively, gives the system equations for the model with metal effect and mass flow dependent heat transfer coefficient. The results for this model structure with $n_s = 1, 2$ are shown in Table 3.3.

In contrast to the experience for the same model structure, however, without modelling of the mass flow dependent heat transfer, it is possible to estimate the parameters of this model. The probable explanation is that the mass flow dependence in the present model to a much higher degree is in accordance with the actual dependence in the heat exchanger. This proposes that it is an erroneous description of the mass flow dependence along with the modelling of the heat capacity in the metal, which is the explanation to estimation problems.

For the model with only one compartment, meaning a third order model, it is seen that the likelihood function and the prediction error variances are comparable to the results for the model with $n_s = 2$ and no metal modelling. The reason that this result is seen might be that the intermediate metal state takes over the role, from additional compartments in the water, of modelling the dynamics and transport lags.

For $n_s = 2$ is the *BIC* lower than for all previously estimated models. Furthermore, the estimates of the water masses are closer to the values, expected from the dimensions of the heat exchanger. The γ-estimates can be evaluated by converting them to estimates of the metal mass. This is done by multiplying with the ratio between the heat capacity of the water (~ 4.18 kJ/kg°C) and the heat capacity of the metal (steel) (~ 0.47 kJ/kg°C). This gives that the metal mass is estimated to 3.0-3.1 kg, which is rather close to the 2.4 kg given by the manufacturer. The α-estimates are, however, unrealistic as they are expected to be close to each other.

3.4.5 DISCUSSION

It is evident from the estimation results that the heat transfer coefficient has a clear dependence on the flow through the heat exchanger. For the models

Statistic	Estimation Result			
	$n_s = 1$		$n_s = 2$	
α''_h	1.09	(.05)	1.00	(.02)
α''_c	6.94	(1.2)	4.87	(.4)
M_h [kg]	0.57	(.02)	0.50	(.01)
M_c [kg]	0.17	(.03)	0.43	(.03)
γ_h [kg]	0.67	(.04)	0.34	(.04)
γ_c [kg]	0.70	(.04)	0.35	(.02)
y	0.67	(.01)	0.57	(.008)
$K_{h_1,h_{out}}$	1.7	(.05)	-4.1	(.4)
$K_{h_1,c_{out}}$	0	(f)	0	(f)
$K_{h_2,h_{out}}$			0	(f)
$K_{h_2,c_{out}}$			0	(f)
$K_{m_1,h_{out}}$	-0.4	(.1)	11.1	(1.1)
$K_{m_1,c_{out}}$	1.3	(.1)	5.3	(1.3)
$K_{m_2,h_{out}}$			50.5	(6.7)
$K_{m_2,c_{out}}$			11.5	(3.7)
$K_{c_1,h_{out}}$	0	(f)	0	(f)
$K_{c_1,c_{out}}$	0	(f)	0	(f)
$K_{c_2,h_{out}}$			-44.2	(5.7)
$K_{c_2,c_{out}}$			-11.7	(3.8)
$-\log L$	949.5		804.3	
BIC	1979		1726	
$\hat{\sigma}^2_{\varepsilon_h}$ [°C^2]	0.74		0.77	
$\hat{\sigma}^2_{\varepsilon_c}$ [°C^2]	0.68		0.40	

Table 3.3: Metal modelling and mass flow dependent heat transfer.

giving the best fit, i.e. no modelling of metal with $n_s = 3$ and modelling of metal with $n_s = 2$, the estimate of y turned out as 0.57 (.01) and 0.57 (.008), respectively. It is satisfying that the estimates are so close and with such small standard deviations. The estimates are, however, below the values recommended in the literature, see e.g. Kakac, Shah, & Bergles (1983), where y is given to lie in the interval $[0.65, 0.85]$ for a plate heat exchanger.

Figures 3.4-3.5 shows the prediction errors of the estimated model with mass flow dependent heat transfer but without metal effect for $n_s = 2$. It seems to be the case that prediction errors are of the same sign for sequences of substantial length. Furthermore, the autocorrelation of the prediction errors in Figures 3.6-3.7 show that much correlation is left in the prediction errors. For both sides it is also seen that for lag one the correlation is clearly significant and positive, whereas for lag three the correlation is even more significant and negative. This seems to indicate that the dynamic response to changes in inlet temperatures and/or mass flows is not described properly by the model. Moreover, estimated cross correlations between mass flows and prediction errors (not shown) disclose that the dynamic dependence on mass flows are not appropriately described.

One explanation to the insufficiency of the model structure could be that in the experiment the exchanger is subject to heavy excitation in the mass flows and in the cold inlet temperature as well. This might cause the appearance of transient flow variations inside the heat exchanger, e.g. propagation of pressure fronts, implying transient responses that is not accounted for by the model.

Another explanation to the shortcoming of the model is that it does not describe the transport delays, and the variation of the transport delays, in the length of the flow. This actually means that the dynamic description of the model serves the purpose of approximating two characteristics of the responses in the heat exchanger, viz. the heat capacities and the transport delays. The distributed heat transfer in the flow direction with distributed transport delay probably add up to give a response that deviates in form

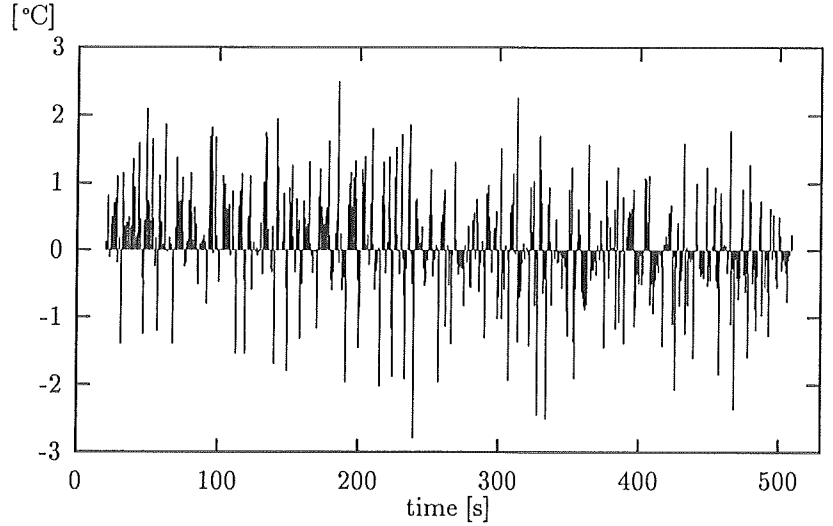


Figure 3.4: Hot outlet prediction errors for model with $n_s = 2$, heat transfer dependent mass flow and no metal modelling.

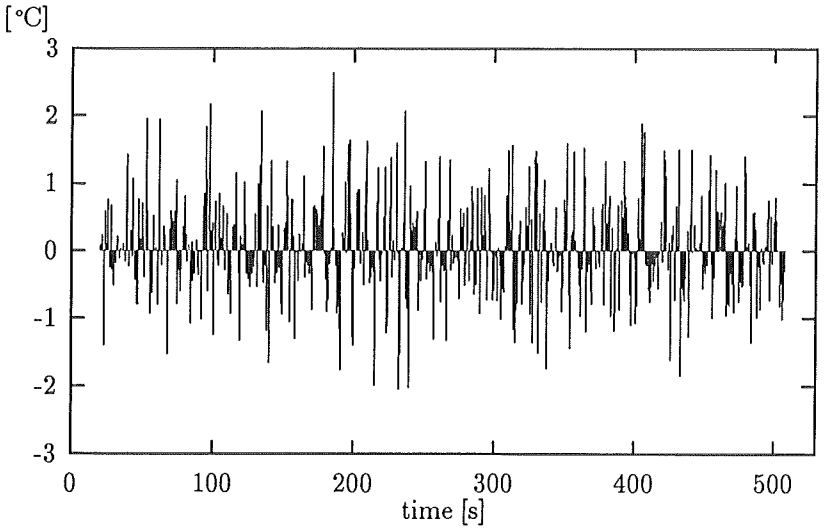


Figure 3.5: Cold outlet prediction errors for model with $n_s = 2$, heat transfer dependent mass flow and no metal modelling.

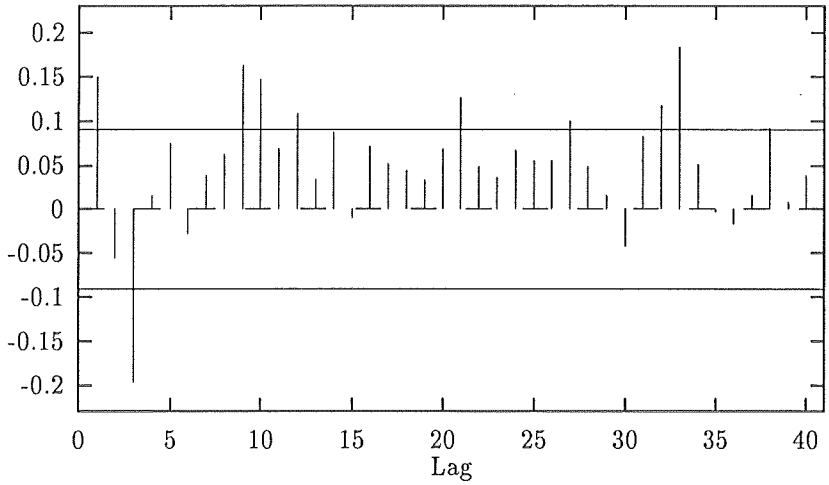


Figure 3.6: Autocorrelation of hot outlet prediction errors.

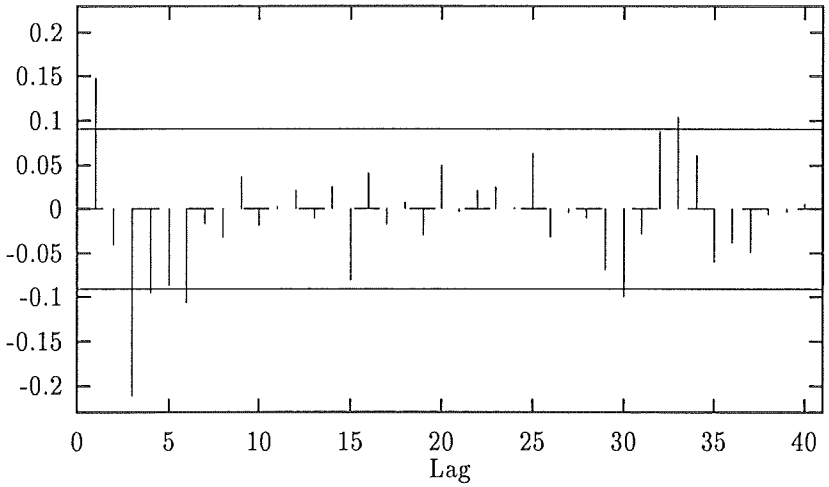


Figure 3.7: Autocorrelation of cold outlet prediction errors.

from that of a system with four time constants.

Yet another reason for the model insufficiency could be that the heat exchanger, from which data has been used in this chapter, is not a pure counter flow heat exchanger. In reality the flows propagate in directions differing from the counter flow, which is not in accordance with the model assumptions.

An additional explanation to the model insufficiency is that the inlet temperatures as well as the mass flows are not constant during sampling interval. Although the shifting of data, leading to the improved model fit, has been used, it does not exclude the erroneous influence of the deviation from the required constancy of input during sampling intervals.

For the cold outlet there seems to be a tendency that the prediction errors are more negative than positive in the first part of the data period and vice versa in the latest part of data. Similarly, for the hot outlet the prediction errors are mostly positive in the first part and tends to be more negative in the latest part of the data set. The existence of this pattern of the prediction errors leads along with the temperature variation in data to the conclusion that modelling of heat transfer dependence on temperature will be beneficial to the model fit.

3.4.6 MODEL IMPROVEMENTS

One obvious way of improving the model is to include the dependence of the heat transfer coefficient on temperature. Including this dependence of h in the parameters means that the model becomes nonlinear in the states, implying that the model assumptions to be fulfilled for using standard Kalman filtering are not met. Jonsson & Palsson (1993) investigate the application of extended Kalman filtering for the nonlinear filtering problem, and reports an improved fit. The present approach is to apply the simplifying approximation of modelling the heat transfer as depending on inlet temperatures instead of compartment temperatures to circumvene the

nonlinearity. Hereby the use of extended Kalman filtering is undue.

One idea for improving the modelling of the mass flow dependence is to extend the space of input variables by the differenced mass flows and apply a kind of black-box parameterization of the dependence on these variates. This idea comes from observing that changes in mass flow are followed by large prediction errors.

The state space formulation of the deterministic part of the model with two compartments, no modelling of metal, modelling of the heat transfer dependence on inlet temperatures and mass flows, and modelling of the dependence on differenced mass flows, takes the form

$$\frac{d}{dt}\begin{bmatrix} T_{h,2}(t) \\ T_{h,1}(t) \\ T_{c,1}(t) \\ T_{c,2}(t) \end{bmatrix} = \begin{bmatrix} -\frac{\dot{m}_h(t)}{M_h} - \frac{\alpha_h^* U}{2M_h} & \frac{\dot{m}_h(t)}{M_h} - \frac{\alpha_h^* U}{2M_h} & 0 & \frac{\alpha_h^* U}{2M_h} \\ 0 & -\frac{\dot{m}_h(t)}{M_h} - \frac{\alpha_h^* U}{2M_h} & \frac{\alpha_h^* U}{2M_h} & \frac{\alpha_h^* U}{2M_h} \\ 0 & \frac{\alpha_c^* U}{2M_c} & -\frac{\dot{m}_c(t)}{M_c} - \frac{\alpha_c^* U}{2M_c} & \frac{\dot{m}_c(t)}{M_c} - \frac{\alpha_c^* U}{2M_c} \\ \frac{\alpha_c^* U}{2M_c} & \frac{\alpha_c^* U}{2M_c} & 0 & -\frac{\dot{m}_c(t)}{M_c} - \frac{\alpha_c^* U}{2M_c} \end{bmatrix} \begin{bmatrix} T_{h,2}(t) \\ T_{h,1}(t) \\ T_{c,1}(t) \\ T_{c,2}(t) \end{bmatrix}$$

$$+ \begin{bmatrix} 0 & \frac{\alpha_h^* U}{2M_h} & B_{h_2,h} & B_{h_2,c} \\ \frac{\dot{m}_h(t)}{M_h} - \frac{\alpha_h^* U}{2M_h} & 0 & B_{h_1,h} & B_{h_1,c} \\ \frac{\alpha_c^* U}{2M_c} & 0 & B_{c_1,h} & B_{c_1,c} \\ 0 & \frac{\dot{m}_c(t)}{M_c} - \frac{\alpha_c^* U}{2M_c} & B_{c_2,h} & B_{c_2,c} \end{bmatrix} \begin{bmatrix} T_{h,in}(t) \\ T_{c,in}(t) \\ \nabla \dot{m}_h(t) \\ \nabla \dot{m}_c(t) \end{bmatrix}, \quad (3.37)$$

where U is given by

$$U = \frac{(1 + bT_{h,in}(t))\dot{m}_h^y(t)(1 + bT_{c,in}(t))\dot{m}_c^y(t)}{(1 + bT_{h,in}(t))\dot{m}_h^y(t) + (1 + bT_{c,in}(t))\dot{m}_c^y(t)}, \quad (3.38)$$

and the new α-parameters are

$$\alpha_h^\star = \frac{A_h K'}{c_h} \qquad \text{and} \qquad \alpha_c^\star = \frac{A_c K'}{c_c}. \tag{3.39}$$

Table 3.4 lists the results from estimation of the model with linear dependence of heat transfer on inlet temperatures (Model A), the model with additional modelling of the dependence on differenced mass flows (Model B) and the model with both extensions (Model AB).

It is obvious that both extensions imply a reduction in BIC as well as in the prediction error variances compared to the previous results.

The parameters describing the dependence on differenced mass flows can be interpreted as an additional change of the derivative of the states following a change in either of the mass flows. A likely explanation to the the improvement is that change of mass flows, resulting in large differenced mass flow, is an indicator that the heat exchanger is excited in way, which makes the actual dynamics of the heat exchanger depart from the compartmental model description. It may, for instance, be an indicator for transient flow variations, and allows for a better description of the actual dynamic responses. It is notable that seven out of eight of the B-estimates are clearly significant using a t-test.

Also the estimated linear heat transfer dependence on inlet temperatures is clearly significant. Furthermore, the b-estimate at 0.0042(.003) is close to estimate in Jonsson & Palsson (1993), being 0.006(.001), which was obtained by using the extended Kalman filter to describe the dependence on compartment temperatures. It is clear that the temperatures are subject to substantial changes in the length of the heat exchanger. This means that representing the temperatures of the compartments by the inlet temperatures is a source of error that is not negligible. This is probably the reason that the estimate in this investigation is closer to zero.

Figures 3.8-3.9 show the prediction errors from the estimation of Model AB. Both sequences of prediction errors seem to have been subject to a

Statistic	Estimation Result		
	Model A	Model B	Model AB
$\alpha_h^\star$ [kg/s]	1.14 (.02)	1.35 (.02)	1.11 (.02)
$\alpha_c^\star$ [kg/s]	1.19 (.02)	1.42 (.02)	1.16 (.03)
M_h [kg]	0.58 (.01)	0.80 (.03)	0.72 (.02)
M_c [kg]	0.47 (.02)	0.78 (.04)	0.70 (.02)
y	0.57 (.01)	0.51 (.01)	0.55 (.01)
b	0.0044 (.0003)		0.0042 (.0003)
$K_{h_2,h_{out}}$	2.3 (.08)	2.1 (.1)	1.7 (.1)
$K_{h_2,c_{out}}$	0 (f)	0 (f)	0.3 (.1)
$K_{h_1,h_{out}}$	-6.5 (0.6)	0 (f)	-2.1 (.7)
$K_{h_1,c_{out}}$	2.8 (0.7)	1.5 (.6)	2.3 (.7)
$K_{c_1,h_{out}}$	1.4 (.3)	1.4 (.3)	0.9 (.2)
$K_{c_1,c_{out}}$	0 (f)	0 (f)	0 (f)
$K_{c_2,h_{out}}$	0 (f)	-2.5 (.5)	0 (f)
$K_{c_2,c_{out}}$	2.7 (.4)	2.2 (.3)	1.2 (.3)
$B_{h_2,h}$ [°Cs/kg]		-29.0 (1.2)	-31.5 (1.2)
$B_{h_2,c}$ [°Cs/kg]		5.1 (.6)	5.9 (.6)
$B_{h_1,h}$ [°Cs/kg]		127.3 (10.0)	119.8 (9.7)
$B_{h_1,c}$ [°Cs/kg]		0 (f)	0 (f)
$B_{c_1,h}$ [°Cs/kg]		-25.7 (2.1)	-28.1 (1.9)
$B_{c_1,c}$ [°Cs/kg]		9.6 (1.6)	12.8 (1.4)
$B_{c_2,h}$ [°Cs/kg]		36.6 (6.9)	27.7 (7.5)
$B_{c_2,c}$ [°Cs/kg]		-51.2 (2.4)	-54.6 (2.3)
$-\log L$	871.6	500.7	455.8
BIC	1836	1132	1042
$\hat{\sigma}^2_{\varepsilon_h}$ [°C^2]	0.57	0.18	0.17
$\hat{\sigma}^2_{\varepsilon_c}$ [°C^2]	0.57	0.20	0.20

Table 3.4: Estimation results for models with heat transfer dependence on temperature and/or use of differenced mass flows as explanatory variables.

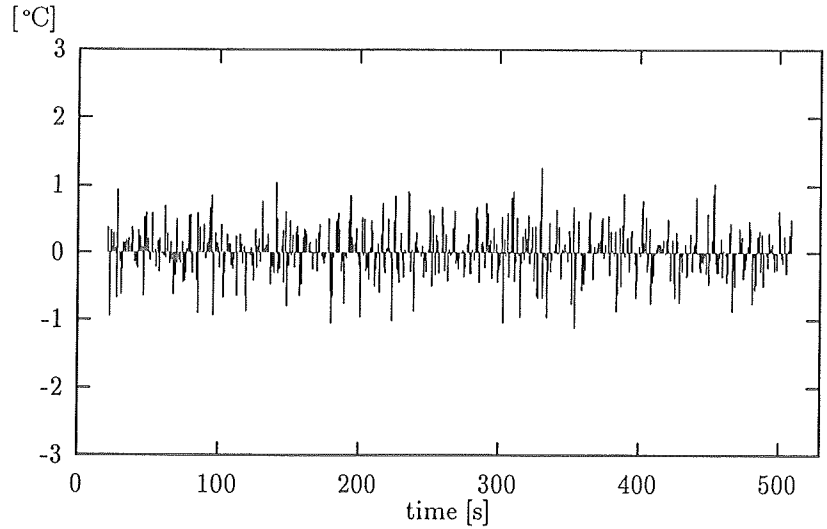


Figure 3.8: Hot outlet prediction errors for model with $n_s = 2$, heat transfer dependence on mass flow and temperature, and dependence on differenced mass flow.

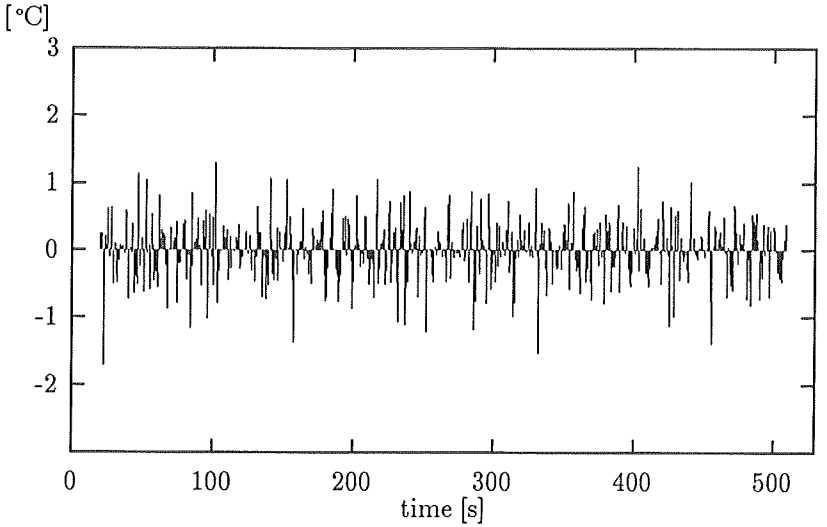


Figure 3.9: Cold outlet prediction errors for model with $n_s = 2$, heat transfer dependence on mass flow and temperature, and dependence on differenced mass flow.

general reduction in size. For both outlets the skewness along the time axis seems to have been removed by modelling the temperature dependence. It is also the picture that the prediction errors tend to be less correlated, but the autocorrelations in Figures 3.10-3.11 show that some correlation is left to be explained. The correlation at lag one is no longer significant, but the correlation at the following lags are still at the significance level and systematically negative.

A validation of the estimated model can be obtained by determining the overall heat transfer coefficient from the parameter estimates, and compare it with what can be obtained from steady state relations using measurements from a situation, where the inlet temperatures and mass flows have been constant for a long time (compared to the time constants). For Model AB the overall heat transfer coefficient can be obtained from the parameter estimates, using an observation vector from the initial phase of the experiment before the excitation was started. Using a heat transfer area $A = 12 \times 0.0255$ m^2, and a specific heat capacity $c_h = 4180$ J/kg°C gives

$$\begin{aligned} U &= \frac{\alpha_h^* c_h}{A} \frac{(1 + bT_{h,in})\dot{m}_h^y (1 + bT_{c,in})\dot{m}_c^y}{(1 + bT_{h,in}(t))\dot{m}_h^y + (1 + bT_{c,in})\dot{m}_c^y} \\ &= 3.62 \frac{\text{kW}}{\text{°Cm}^2}. \end{aligned}$$

The steady state relation (3.6) for the hot water gives, with the same values of A and c_h and the same observation vector,

$$U = 3.61 \frac{\text{kW}}{\text{°Cm}^2}, \tag{3.40}$$

showing excellent agreement between the two heat transfer values.

Seeing the agreement it may be asked whether there is anything gained by using the stochastic modelling approach along with statistical estimation of model parameters. First of all it is not possible to assess all the parameters of the differential equations in the same way as for the heat transfer coefficient. Secondly, the statistical approach offers parameter estimates as well as estimates of the variances of the determined parameters. This

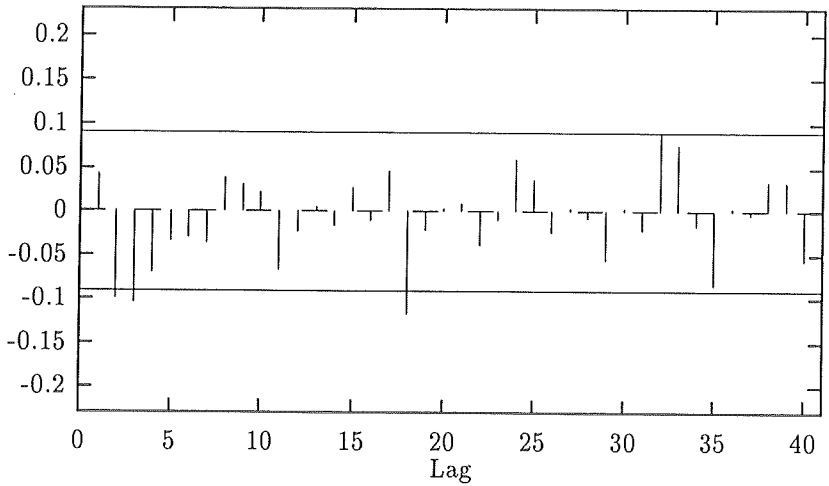


Figure 3.10: Autocorrelation of hot outlet prediction errors.

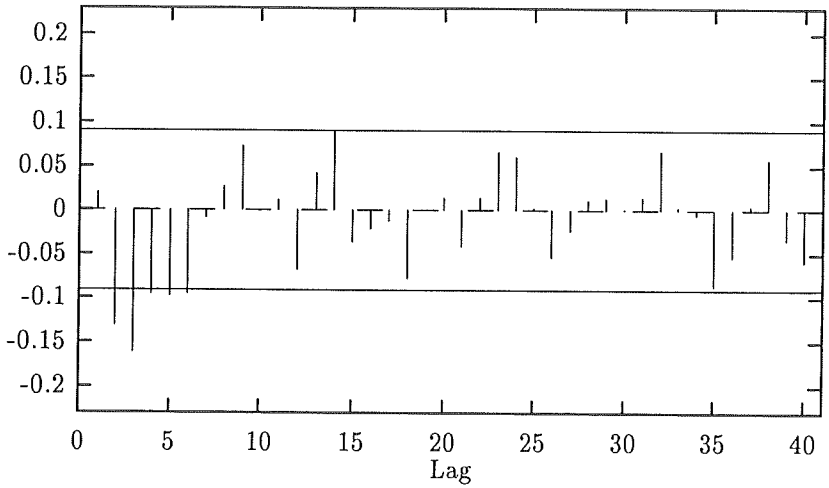


Figure 3.11: Autocorrelation of cold outlet prediction errors.

can, for instance, be used for supervision of the heat transfer capability of a heat exchanger. The statistical approach also allows for an efficient use of a set of data for the investigation of appropriate input-output relations, estimation of parameters, test of model assumptions and exploration of model insufficiencies. Finally, it is always possible using statistical methods to investigate whether it is possible to estimate a given parameter, and if so to estimate the parameters in a dynamic model that complies with the actual dynamic relations of the system.

3.5 CONCLUSION

The modelling and estimation results in this chapter is a demonstration of the application of continuous time linear stochastic modelling to measurements from a heat exchanger. Physical constants as well as parameters of empirical relations has been estimated using ML estimation for a lumped system description.

An important quality of the present approach is that given a set of data, which preferably scans the whole operating range of the heat exchanger, for both mass flows and temperatures, parameters of empirical formulas for the dependence on operating point can be estimated along with other parameters of the model.

Based on the data used in the present study it was found that it is profitable, in a dynamic model for a heat exchanger, to describe the dependence of the heat transfer coefficients on mass flows and temperatures. Furthermore, the distributed character of the heat transfer in the heat exchanger is approximated reasonably well by introducing a few compartments in the model formulation. It also turned out to be possible to estimate a model, which discriminated between the heat capacity of water and metal, but the estimates of the heat transfer coefficients were not realistic. Finally, the model was also improved by introducing a dependence on the differenced mass flows. This is probably due to the lacking capability of the model in describing transport lags and aberrant mass flows.

The results have indicated that the compartmental model of the heat exchanger, formed from energy and mass conservation, does not give a sufficient description of the dynamics, at least when the inlet temperatures and the mass flows are changed continuously. The insufficiency of the model has shown the unfavourable implication that the parameter estimates to a large extent depend on model structure and even model order.

The aim of modelling a heat exchanger in continuous time is mainly to be able to assess its physical characteristics. For instance, in DH systems is of great value to know the time constants of the heat exchanger to deduce the dynamic influence of the existence of between transmission and distribution net. Parameter estimates might also be used for assessment of heat transfer coefficients. Based on the results in this chapter it seems to be important, if the purpose is to compare different heat exchangers or to the determine changes of a heat exchanger over time, to use the same excitation in each experiments. This is in fact important for any system since the resulting estimates, or at least the uncertainty of the estimates, have significant dependence on the excitation of the system.

.....

All results in this chapter was obtained using CTLSM for estimation of continuous time linear stochastic models, see Melgaard & Madsen (1991). The calculations were carried out on a HP-9000/750 in double precision arithmetic.

.....

4

Nonparametric Regression for Identification Purposes

In some settings it is known that there exists a relationship between a response variable and one or more explanatory variables. As an example Chapter 3 has described the modelling of a heat exchanger where it is obvious that outlet temperatures depend on inlet temperatures as well as mass flows. This dependence has been modelled using physical considerations for assessment of model structure. However, it is not always possible to assess the character or even the existence of potential dependence relations. This situation demands that methods for disclosure and identification of model structure are applied.

In the present chapter nonparametric regression methods are applied with the purpose of identifying how the heat load in DH systems is related to the time of day, ambient air temperature and supply temperature. According to the discussion in Chapter 2 it is expected that the heat load do dependent on these variates.

The fact that there is an influence of time of day, and even of the type of day and the time of the year, has the implication that the system is not stationary. That is, the relations in the system are not constant in time. Using nonparametric regression for describing the relation between time of day and heat load, both when the time of day is used as the only explanatory variable and when the multivariate dependence on time of

day, supply temperature and ambient air temperature is estimated, can be regarded as a tool for determining the instationarity. Furthermore, nonparametric regression is also a tool for disclosing the nonlinearity of the relations between the explanatory variables and the heat load.

4.1 THE ISSUE OF IDENTIFICATION

For the class of linear model structures with normal distributed innovations there exists a number of procedures for identification of model orders. The most celebrated examples are the well-known methods for estimation of autocorrelation, partial and inverse autocorrelation applying to the determination of order of ARIMA models (Box & Jenkins, 1976). This is supplemented by procedures for identification of orders of transfer functions, see e.g. Edlund (1989) for different proposals for preliminary estimation of transfer function weights. As a solution to the general problem of choosing among different models with different numbers of parameters Akaike (1974) has applied an extension of the maximum likelihood principle leading to the well-known information criterion AIC, and a modification of this is the BIC (Schwarz, 1978), which can be shown to give consistent estimates of parameter linear models. Furthermore, Auestad & Tjøstheim (1990) propose the use of the final prediction error (FPE) criterion for determining the order of a general nonlinear model.

The advantage of the linear model structure with normal distributed noise is that a complete theory is available, and much experience has been gained in the application of ARIMA and transfer function models. On the other hand it is a fact that for most systems there are deviations from linearity and/or the assumption of normal distributed innovations. When the restriction of linear Gaussian models is relaxed there are unlimited modelling possibilities. Priestley (1988) and Tong (1990) list and investigate the properties of a number of nonlinear model structures. The issues of greatest importance are first of all the decision of whether extensions going beyond linear modelling is necessary and secondly, the determination of an

appropriate nonlinear model structure if linear modelling is considered to be insufficient.

In Robinson (1983) nonparametric regression methods are proposed for estimation of conditional expectations in non-Gaussian time series models, and Auestad & Tjøstheim (1990) study the identification of nonlinear time series using nonparametric estimates of the conditional mean and conditional variance. Tong (1990) outlines proposals for graphical analysis of data, the purpose of which is to be a guidance in disclosing possible nonlinear relations. For instance, reverse data plots can be used to detect time irreversibility of data. Mellentin (1992) demonstrates the use of neural network models to estimate the conditional expectation for a given set of data. This way of tackle the problem shows to be superior to nonparametric regression for threshold models. Rao & Gabr (1980) has a test in the bispectral density function for linearity of stationary time series, which is improved in Hinich (1982). Tong (1990) presents an overview of tests for deviations from the linear Gaussian model.

Nonparametric regression is a tool which is suitable for estimation of regression curves, where relationships either are completely unknown or where the substance of the dependence relations is indistinct. This means that nonparametric regression can be applied for identification of existing relationships leading to proposals for parametric model classes.

Following the discussion in Chapter 2, part of the load in a DH system can be assumed to have the characteristic that it consists of a mean value, which is a function of time and date, superposed zero mean random variables. In this chapter, using data from 16 terrace houses in Kulladal, a suburb of Malmö in Sweden, nonparametric regression methods are applied for estimation of heat load as a function of time of day. This investigation serves both as a comparison of a number of methods as well as a study of the applicability of these methods for modelling the diurnal variation of heat load in general.

One of the methods, kernel smoothing, is applied for estimation of the mul-

tivariate dependence on time of day, ambient air temperature and supply temperature in a large scale DH system with considerable transport delay. The aim of this study is to uncover the dependence on these explanatory variables, the character of which is known to go beyond linearity. Load data from the DH system in Esbjerg, Denmark are used in this part of the investigation.

4.2 OUTLINE OF NONPARAMETRIC REGRESSION

In the application of nonparametric regression it is assumed that a set of data $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^N$ can be described by the following regression relationship

$$y_i = m(\boldsymbol{x}_i) + e_i, \qquad i = 1, \ldots, N, \tag{4.1}$$

where $\boldsymbol{x}_i$ is a vector of explanatory variables, y_i is the response variable, and e_i is an additive stochastic variable with mean zero and in the first place, before considering the case of correlated observations, the innovations are are assumed to be independent. The goal is to determine an estimate of the regression curve or surface $m(\cdot)$, and the procedure of obtaining the estimate is referred to as smoothing.

Building a model of the regression function $m(\cdot)$ can be done both parametrically and nonparametrically. In the case where a parametric model is chosen, restrictions on the form of the regression curve are imposed. That is, the parametric model can only be estimated to fit the underlying regression function within a restricted class of functional relationships. On the other hand, when using nonparametric regression methods the degree of smoothness has to be chosen.

Some classes of nonparametric regression curves can also be formulated in parametric models, e.g. orthogonal series estimators. In this case the choice of degree of smoothing in the nonparametric method corresponds to the choice of number of terms in the orthogonal series.

The basic idea of smoothing is that the estimate for a specific value of the

explanatory variable, x, is a weighted average of response observations for which the explanatory variable is close to x. The reasoning behind this approach is that observations, y_i, for x_i close to x contain information about the value of the regression curve at x.

The local average forming the estimate at x is written as

$$\hat{m}(x) = \frac{1}{N} \sum_{i=1}^{N} W_{N,i}(x) y_i. \tag{4.2}$$

The weight $W_{N,i}(x)$ depends, of course, both on x and x_i. Furthermore, in choosing an optimal set of weights, the number of observations is of significance as is the smoothness of $m(x)$. This means that to choose the degree of smoothing in the estimation it is actually necessary to know the underlying regression curve.

The estimate in $\hat{m}(x)$ as given in (4.2) is actually a nonparametric least squares estimate at the point x. This is recognized from the fact that the solution to the least squares problem

$$\arg\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} W_{N,i}(x)(y_i - \theta)^2 \tag{4.3}$$

is given by

$$\hat{\theta}_{LS}(x) = \frac{\sum_{i=1}^{N} W_{N,i}(x) y_i}{\sum_{i=1}^{N} W_{N,i}(x)}. \tag{4.4}$$

This shows that at each x, $\hat{m}(x)$ is a scaled weighted LS location estimate, i.e. $\hat{m}(x) \equiv \hat{\theta}_{LS}(x) \times \frac{1}{N} \sum_{i=1}^{N} W_{N,i}(x)$.

4.3 KERNEL SMOOTHING

Kernel smoothing is probably the most frequently applied smoothing method. The reason is that representation of the weights in (4.2) is simple and easily comprehensible. The shape of the weight function is given by a density function with scale parameters that adjust the size of the weighting in the space of the explanatory variables. Usually this weight function, the kernel, is a continuous, bounded and symmetric function which integrates to one.

The weights in (4.2) are given by

$$W_{N,i}(x) = \frac{K_{h_N}(x - x_i)}{\hat{f}_{h_N}(x)}, \tag{4.5}$$

where the kernel estimator of the density of the explanatory variables is

$$\hat{f}_{h_N}(x) = \frac{1}{N} \sum_{i=1}^{N} K_{h_N}(x - x_i). \tag{4.6}$$

The kernel with scale parameter (bandwidth) h is given by

$$K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right). \tag{4.7}$$

The estimator given by (4.2) with the weights given by a kernel as in (4.5) is called the Nadaraya-Watson estimator, which by insertion is given as

$$\hat{m}_{h_N}(x) = \frac{\sum_{i=1}^{N} K_{h_N}(x - x_i) y_i}{\sum_{i=1}^{N} K_{h_N}(x - x_i)}. \tag{4.8}$$

From this formulation it is seen that the estimate at x of the regression function can be interpreted as a locally weighted average of the observed response variables scaled by the effective number of observations entering the nonparametric estimate.

In the case of more than one explanatory variable the nonparametric regression estimate will have a dimension corresponding to the number of explanatory variables, and the weighting of the observations will similarly be computed in a space with the same dimension. Hence, the applied kernel must be a function of the distance between the point on the curve estimate and the observation entering the estimate. The measure of the distance in the multi-dimensional space is determining for the weighting of each of the explanatory variables. Usually the applied kernel is given as a product of single-variate kernels. This means that the curve estimate is computed by measuring the distance to each of the observed explanatory variables separately. The reason for this simplification is that even in the case of using products of single-variate kernels, it is a complicated task to determine the appropriate scale parameters. Skipping the interaction terms reduces the number of scale parameters to be determined. For d-dimensional explanatory variables the weights in (4.2) become

$$W_{N,i}(x) = \frac{\prod_{j=1}^{d} K_{h_N}(x_j - x_{i,j})}{\frac{1}{N}\sum_{i=1}^{N}\prod_{j=1}^{d} K_{h_N}(x_j - x_{i,j})}. \tag{4.9}$$

To assess the variance of the curve estimate at point x Härdle (1990) propose the pointwise estimator given by

$$\hat{\sigma}^2(x) = \frac{1}{N}\sum_{i=1}^{N} W_{N,i}(x)(y_i - \hat{m}_{h_N}(x_i))^2. \tag{4.10}$$

This variance estimate can be used to determine pointwise confidence intervals for the curve estimate. The bias of the curve estimate, however, is not taken into account. This can be done using estimates of derivatives of $m(\cdot)$ and f, which, however, considerably increases the complexity of the calculations.

It is clear that the variance of the curve estimates depend on the number of observation entering the weighted average. However, also the variance

estimate (4.10) is heavily influenced by the number of observations. For instance, if a curve estimate is computed using a bandwidth, which implies that only one observation is used, the curve estimate will be equal to the observation of the dependent variable, and the variance of the estimate will equal the variance of this observation. The estimator (4.10) gives, however, zero variance, showing that (4.10) is not reliable in areas, where the number of observation is low.

Härdle (1990) shows that in the case of a one-dimensional stochastic explanatory variable, subject to non-restrictive conditions on the kernel and given that the variance of the response variable exists, then for

$$N \to \infty, \qquad h \to 0 \text{ and } Nh \to \infty$$

and in every point of continuity of $m(x)$, $f(x)$ and $\sigma^2(x)$ the estimate (4.2) converges in probability to the regression curve $m(x)$.

4.3.1 CHOICE OF KERNEL AND BANDWIDTH

The optimal choice of kernel and bandwidths can be considered as the one which minimizes the average squared error given by

$$d_A(\hat{m}_{h_N}, m) = \frac{1}{N}\sum_{i=1}^{N}(\hat{m}_{h_N}(x_i) - m(x_i))^2. \tag{4.11}$$

In this expression the true regression curve, $m(x)$, is assumed to be known. It can be shown, see Härdle (1990), that the parabolic kernel with bounded support, called Epanechnikov kernel, minimizes $d_A(\hat{m}_{h_N}, m)$ among all kernels, in which the bandwidth is optimally chosen. The Epanechnikov kernel is given by

$$K_h^{Epa}(u) = \frac{3}{4h}(1 - \frac{u^2}{h^2})I_{\{|u|\leq h\}}. \tag{4.12}$$

Other applicable kernels are, e.g. the rectangular, triangular or the Gaus-

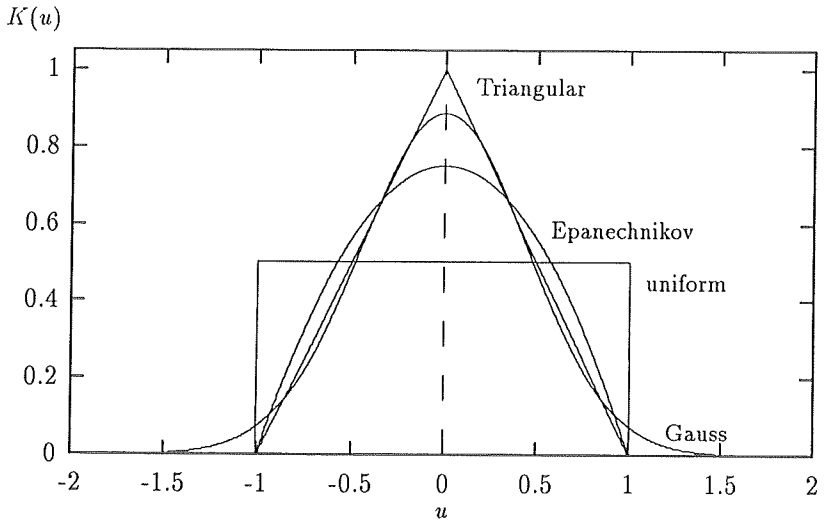


Figure 4.1: Different univariate kernel functions.

sian kernel given by

$$K_h^{Gau}(u) = \frac{1}{h\sqrt{2\pi}} \exp\left(-\frac{u^2}{2h^2}\right). \tag{4.13}$$

These kernels are depicted in Figure 4.1. The Epanechnikov kernel is shown for $h = 1$, and the Gaussian kernel for $h = 0.45$. The choice of the kernel is not critical, whereas it is more important to choose optimal bandwidths. Once the kernel is chosen, optimality with respect to bandwidth can be considered in terms of minimization of the average squared error given by (4.11).

In practice the true regression function is unknown, so instead the minimization of the prediction error estimate could be considered

$$p(\hat{m}_{h_N}) = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{m}_{h_N}(x_i))^2. \tag{4.14}$$

However, $p(\hat{m}_{h_N})$ is a biased estimate of $d_A(\hat{m}_{h_N}, m)$, since y_i enters the calculation of $\hat{m}_{h_N}(x_i)$. For instance, in the case where explanatory variables are fixed and not stochastic, considered in Section 4.5, the bandwidth that minimizes the prediction error, is the one that makes the estimate be an average alone of those observations occurring at the very same time of day.

If the observation y_i is left out of the calculation of $\hat{m}_{h_N}(x_i)$ the correlation between the observation y_i and the nonparametric curve estimate $\hat{m}_{h_N}(x_i)$ is removed. This is, of course, only true when the observations $\{y_i\}$ are uncorrelated. The corresponding measure of the prediction ability of the nonparametric curve estimate is referred to as the cross-validation (CV) function given by

$$CV(\hat{m}_{h_N}) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{m}^{\star}_{h_N,i}(x_i))^2, \tag{4.15}$$

in which the modified estimate is

$$\hat{m}^{\star}_{h_N,j}(t) = \frac{1}{N-1}\sum_{i=1,i\neq j}^{N} W_{h_N,i}(t)y_i. \tag{4.16}$$

The bandwidth that minimizes the CV-function can thus be used as an estimate of the optimal bandwidth with respect to minimization of $d_A(\hat{m}_{h_N}, m)$. Härdle & Marron (1985) prove, for conditions on the continuity of K, f and m, some further non-restrictive conditions, and given limits on the sequence of h_N for increasing N, that the minimizing value of (4.15) is asymptotically optimal with respect to the distance d_A. Seather (1992) investigate the performance of a number of popular methods for selecting bandwidth in kernel density estimation when applied on real data sets. One of these methods is the least squares cross-validation (LSCV) method corresponding to method outlined above for kernel regression. The conclusion is that for a limited set of data, the bandwidth proposed by LSCV in some cases may be far from a visually optimal bandwidth. The other proposals for selection of bandwidth also show a considerable variability around the bandwidth choices, which seem to be optimal by visual inspection.

In cases, where, for instance, the density of the explanatory variables is not constant or the gradients or m show large variations, it is appropriate to apply methods that allow the bandwidth to adapt locally to such aspects. The use of non-symmetric kernels is also relevant in the extreme regions of the observed explanatory variables to avoid bias effects from skewness in the observation distribution.

Härdle & Vieu (1992) discuss kernel regression smoothing of time series. They state the concept of α-mixing for dependent processes, which in practice is fulfilled if the correlation decreases with the distance in time. Subject to a number of conditions, in practice of little restriction, they show that for an bivariate times series $\{(x(t), y(t)\}$ fulfilling the α-mixing condition the bandwidth that minimizes the CV-function is asymptotically optimal.

Since a kernel estimate of a regression function (4.2), no matter what kernel is used, consists of nothing else but weighted averages, it is easily seen that it is possible for a predetermined bandwidth to make recursive calculations of the estimate. It is even possible to introduce a weighting in time hereby making the nonparametric curve estimate adaptive. The computations could be carried out in the following way: When a new observation occurs each of the estimates can be updated according to a scheme, which is determined beforehand from the kernel and the applied bandwidth Each new estimate of the selected points in the regression function will then turn out to be a weighting of the corresponding old estimate and the new observation.

4.4 ORTHOGONAL SERIES ESTIMATION

In the application of orthogonal series estimators it is assumed that the regression function can be represented as an infinite series of orthogonal functions

$$m(x) = \sum_{j=0}^{\infty} \beta_j \varphi_j(x). \tag{4.17}$$

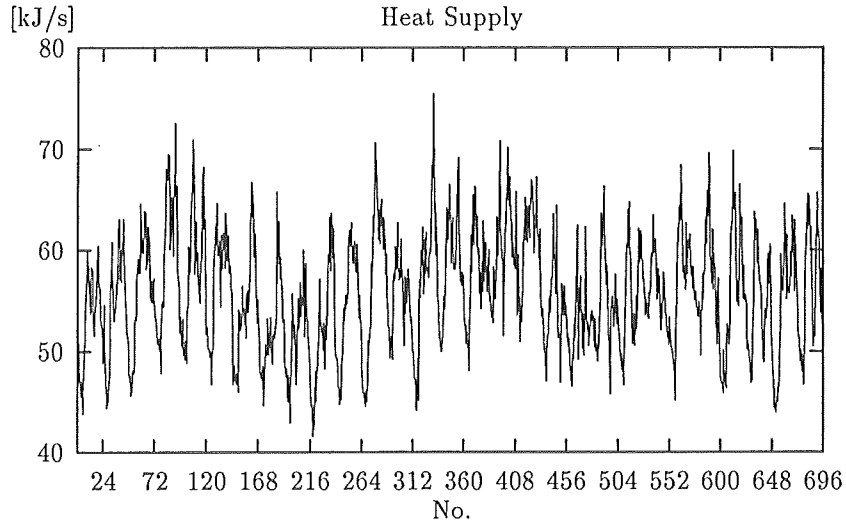


Figure 4.2: Supplied heat to 16 houses in Kulladal during February 1989.

This expression is valid for one-dimensional explanatory variables. $\{\varphi_j\}_{j=0}^{\infty}$ is the known basis of functions, and $\{\beta_j\}_{j=0}^{\infty}$ are the unknown coefficients in the infinite series.

It is reasonable to assume that the heat load can be adequately represented by a series of orthogonal functions. Moreover, when the observations are taken at a finite number of predictor values, the number of coefficients different from zero is bounded upwards by this particular number. Trigonometric functions and Legendre polynomials are two examples of orthogonal functions. Both of these are applied on the load data from Kulladal.

4.5 DIURNAL DEPENDENCE

For the investigation of nonparametric regression methods applied to load data from DH systems, measurements of the supplied heat to 16 terrace houses are used. The 16 houses are densely located, and the supplied

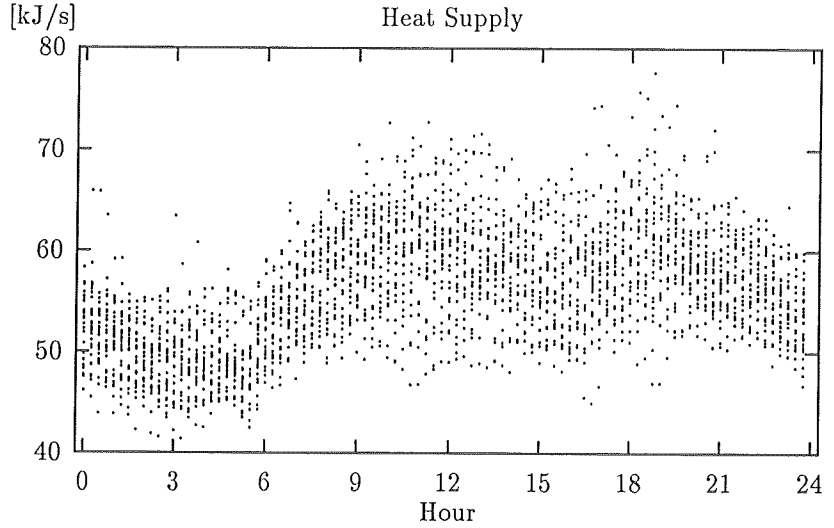


Figure 4.3: Supplied heat versus time of day and night.

power is computed from measurements of the water flow and the supply and return pipe temperatures just in front of the houses. The observations are equidistantly sampled on the interval [0,24] with 0.25 hour between the sample points. The transport delay between the consumers and the measuring instruments is not more than a couple of minutes. Since the sampling interval is 15 minutes, the transport delay is considered to be negligible.

In spite of the earlier mentioned remarks concerning the necessity of modelling building dynamics and influence of weather condition, the measurements from Kulladal are used directly in the calculations of the smoothed regression curves. An approach for describing the dynamics, using filtering of explanatory variables, is applied in Section 4.6.

In the present investigation it is assumed that the heat supply can be related to the time of day, and that there is no difference between the days for the considered period. Then the regression curve is a function only of the time

of day, and it is this functional relationship, which will be considered. In Figure 4.3 the measured supplied heat versus time of day is shown. A clear indication of a time dependent heat supply is seen.

As seen from Figure 4.2 the measurements of supplied heat subtracted a common diurnal profile seem to be correlated in time, and it is most likely the case that, when modelling the supplied heat as in (4.1) with $m(t)$ only being a function of time of day, the noise components will not be independent. Thus, the assumption of independence is not fulfilled due to not modelling the dynamics. The consequence of this is that the information about the regression function contained in data is minor compared to a set of data with the same number of independent observations.

Since the data consists of the same number of samples at each 15 minutes of the day, the curve estimates calculated in this section will show to be the same, whether they are based on the original data or on the averages at each sampling time of the day. Because of the considerable saving of computer time the curve estimates will thus all be based on the averages. Figure 4.4 shows these averages along with the median of the observations. The median is known to be robust towards possible outliers. Nevertheless, there is no remarkable difference between the averages and the median curve, and furthermore no outliers seem to be present in the original data. Therefore, the median is not considered any further.

4.5.1 KERNEL ESTIMATION

Figure 4.5 shows the Epanechnikov curve estimate calculated with a spectrum of bandwidths. It is clear, how more and more characteristics of the non-smoothed averages disappear, when the bandwidth is increased. Most of the characteristics are noise due to the stochastic nature of the heat load, but the drop at 5.15 in the morning seems to be real, seeing that the median has an even more pronounced drop. On the other hand this drop can not be given any reasonable explanation in terms of energy demand of the consumers, but instead from the configuration of the system in connection

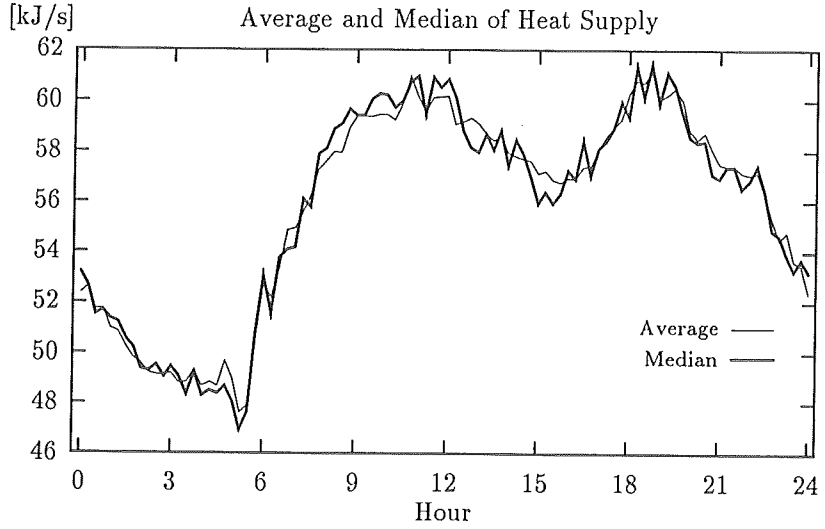


Figure 4.4: Average and median versus time of day.

with the consumption of others. The 16 houses lie on the same distribution line as a building with many flats. If it so that this buildings has a general night set-back and the step in temperature set point is at 5.15 it implies that the flow at the 16 drops momentarily due to the sudden increase in consumption. This will imply a decrease in heat supply to other consumers on the same supply line.

If the drop at 5.15 has to be visible and a reasonable smoothing of the noise at the same time is required, a value close to 0.75 seems to be a good compromise. The curve estimate with bandwidth 0.625 shown in Figure 4.6 thus removes most of the noise and still the drop in the morning can be seen.

Bringing the cross-validation function into action on the original data gives the result shown in Figure 4.8. Minimum of the CV-function (4.15) is obtained for a bandwidth close to 1.3. This value is somewhat bigger

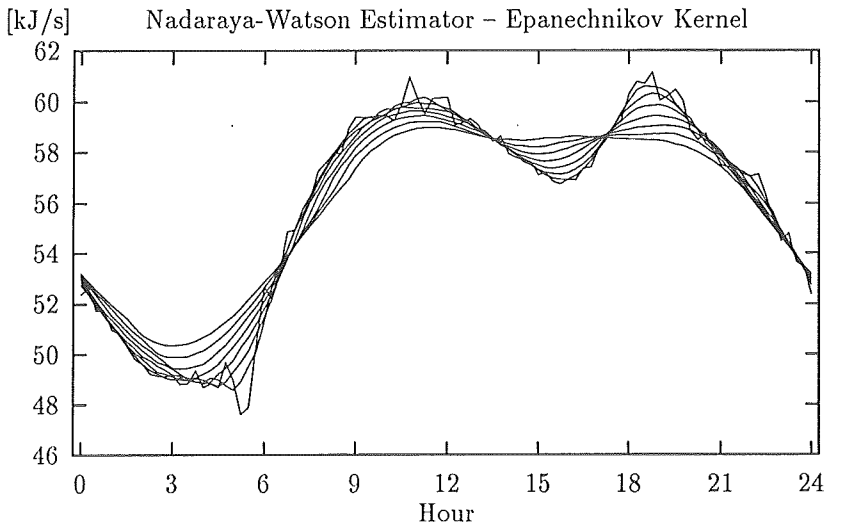


Figure 4.5: Smooth with Epanechnikov kernel and bandwidths 0.125, 0.625,..., 3.625.

than the visually chosen, and the drop at 5.15 will clearly disappear in a curve estimate with this bandwidth. This discrepancy possibly occurs as a consequence of the observations not being independent. For instance, if the additive noise components in the observations are autocorrelated in a way implying a low-pass filtering, the minimum of the CV-function will be obtained for a higher bandwidth than it would have been for independent noise.

These results indicate the need of nonparametric curve estimates in which the bandwidth is not only a function of the density of the predictor variables, but for instance also of the gradient of the regression curve. This matter is, however, not pursued any further in this investigation.

Considering the smoothed regression curves obtained with the Gaussian kernel on Figure 4.7, and by using the same arguments as above in choosing

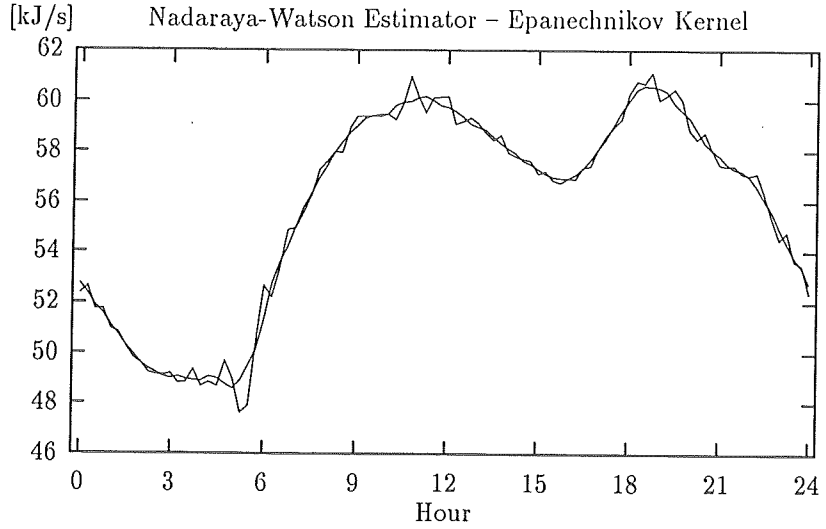


Figure 4.6: Smooth with Epanechnikov kernel and bandwidths 0.125 and 0.625.

the visually reasonable bandwidth, one comes to the conclusion, that a bandwidth close to 0.35 is the right one. However, the cross-validation function on Figure 4.9 indicates that a bandwidth between 0.60 and 0.65 is optimal. The same comments as given along with the Epanechnikov kernel apply here.

Visually it is difficult to observe any difference between the curve estimates obtained with the Epanechnikov and Gaussian kernel, though the optimum of the cross-validation functions is a little smaller for the Epanechnikov kernel. This agrees with the result in Härdle (1990) concerning optimality of kernel, even though the difference is rather small.

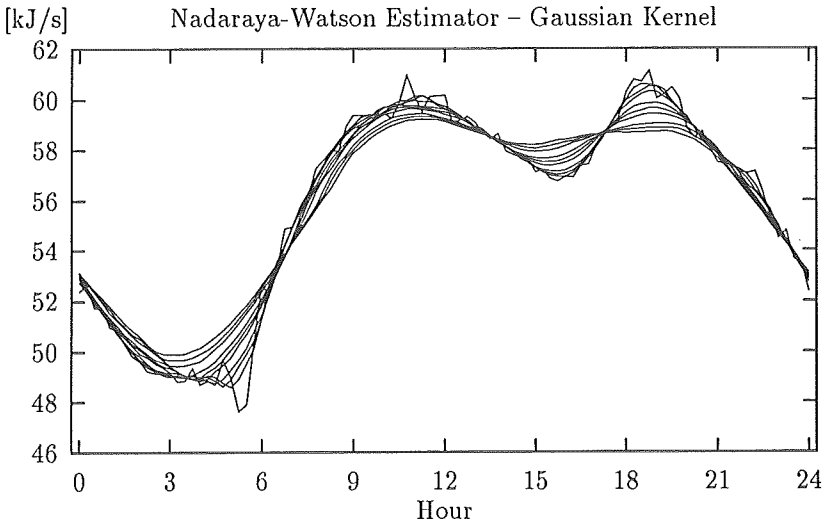


Figure 4.7: Smooth with Gaussian kernel and bandwidths 0.025, 0.125,..., 0.925.

4.5.2 ORTHOGONAL SERIES ESTIMATION

The following outline of the orthogonal series estimation is specific to the problem in consideration. The outline is dedicated this problem to avoid stating both the general expressions and the expressions particular to the discussed problem. Moreover, the method hereby probably turns out to be more comprehensible, and without too much puzzle it can be modified to other problems.

When the system of functions $\{\varphi_j\}_{j=0}^{\infty}$ constitutes an orthonormal basis on this interval, i.e.

$$\int_0^{24} \varphi_j(t)\varphi_k(t)dt = \delta_{jk} = \begin{cases} 0 & , \text{ if } j \neq k \\ & \\ 1 & , \text{ if } j = k \end{cases} , \tag{4.18}$$

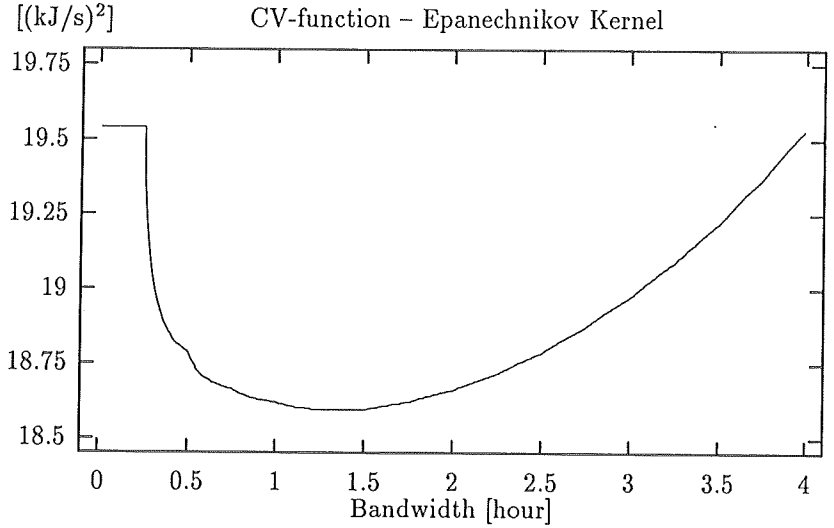


Figure 4.8: CV-function applied on the original data using the Epanechnikov kernel.

the coefficients in the series of orthogonal functions can be computed as

$$\beta_j = \int_0^{24} m(t)\varphi_j(t)dt. \tag{4.19}$$

In this expression the regression function $m(t)$ is unknown, but by using the observations and by writing the integral as a sum of integrals over disjoint intervals the following estimate of the coefficients is obtained

$$\hat{\beta}_j = \sum_{i=1}^{96} \bar{y}_i \int_{A_i} \varphi_j(t)dt, \tag{4.20}$$

where $\{A_i\}$ are quarterly intervals and

$$\bigcup_{i=1}^{96} A_i = [0, 24] \qquad \text{and} \qquad \forall i, j, i \neq j : A_i \bigcap A_j = \emptyset. \tag{4.21}$$

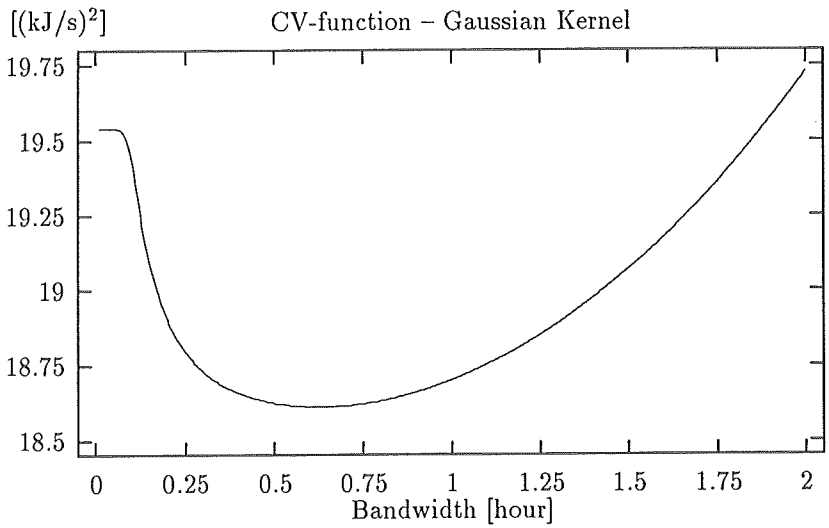


Figure 4.9: CV-function applied on the original data using the Gaussian kernel.

$\bar{y}_i$ is the average of the observations at the time of day t_i. Since the observations are averages of the supplied heat during the past 15 minutes, each interval is set to $A_i =]t_i - 0.125, t_i + 0.125]$ with $t_i = 0.25i$, $i \in \{1, 2, \ldots, 96\}$.

The approximation of the regression function by a finite number n of orthogonal functions is then

$$m_n(t) = \sum_{j=0}^{n} \hat{\beta}_j \varphi_j(t). \tag{4.22}$$

Introducing (4.20) into (4.22) gives

$$m_n(t) = \sum_{j=0}^{n} \left(\sum_{i=1}^{96} \bar{y}_i \int_{A_i} \varphi_j(x)dx \right) \varphi_j(t)$$

$$= \sum_{i=1}^{96} \bar{y}_i \left(\sum_{j=0}^{n} \varphi_j(t) \int_{A_i} \varphi_j(x)dx \right). \tag{4.23}$$

Comparing this approximation with the smoother in (4.2), it is observed that the orthogonal series estimator can be considered as a kernel estimator

$$\hat{m}_n(t) = \frac{1}{96} \sum_{i=1}^{96} W_{n,i}(t)\bar{y}_i. \tag{4.24}$$

with weights

$$W_{n,i}(t) = \frac{1}{96} \sum_{j=0}^{n} \varphi_j(t) \int_{A_i} \varphi_j(x)dx. \tag{4.25}$$

From the expression (4.20) of the estimates it is seen that it is possible to calculate the estimates recursively. The time update is given from the values of the integrals and the way in which the averages entering the estimates are calculated. The new estimates will then be a weighting of the corresponding old estimates and the new observation.

TRIGONOMETRIC FUNCTIONS

The system of trigonometric functions fulfilling the orthonormality in (4.18) is

$$\begin{aligned} \varphi_0(t) &= \tfrac{1}{\sqrt{24}} \\ \varphi_1(t) = \frac{1}{\sqrt{12}} \sin 2\pi \frac{t}{24} \qquad & \qquad \varphi_2(t) = \frac{1}{\sqrt{12}} \cos 2\pi \frac{t}{24} \\ \vdots \qquad & \qquad \vdots \\ \varphi_{n-1}(t) = \frac{1}{\sqrt{12}} \sin 2\pi \frac{nt}{48} \qquad & \qquad \varphi_n(t) = \frac{1}{\sqrt{12}} \cos 2\pi \frac{nt}{48} \end{aligned} \tag{4.26}$$

The order of trigonometric function estimate is $n_F = \frac{n}{2}$, and it thus contains $2n_F + 1$ orthonormal functions.

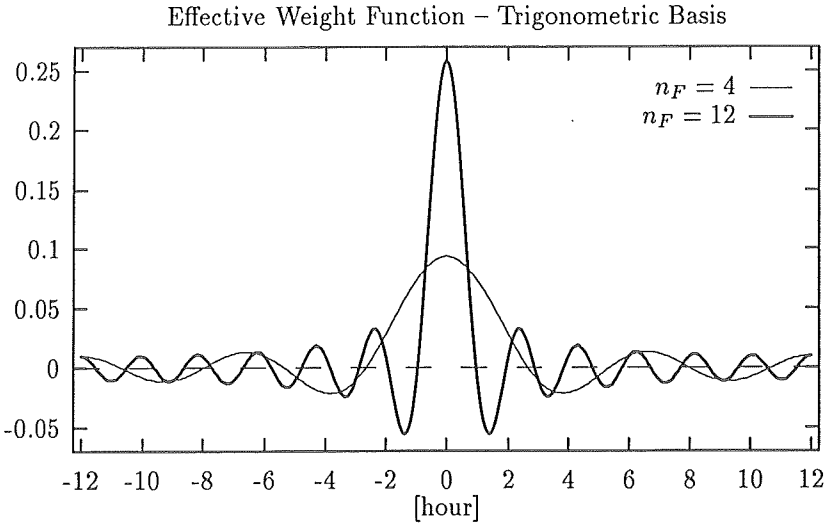


Figure 4.10: The effective weight functions for the trigonometric function approximation of order 4 and 12. The values on the abscissa is the distance between the time of the curve estimate and the time of the observation used in the curve estimate.

The effective weight functions as given in (4.25) computed for the trigonometric functions with $n_F = 4, 12$ are shown in Figure 4.10. It is seen, that all of the observations enter the calculation of the estimate at each time of the day, but the weight is decreasing with the distance between the time of the observation and the time of the estimated response. It is also clear that observations further away than one hour in some sense are averaged out due to the fluctuating course around zero of the effective weight function.

The Figures 4.11-4.12 show the trigonometric function estimates of orders $n_F = 4$ and 12. The estimate of order $n_F = 4$ follows in broad outline the variations of the load during the day, and for order $n_F = 12$ the estimate moreover adopts the greater details of the averages as for instance the drop at 5.15 am. Furthermore, the estimate seems to pick up some of the more

noisy parts, so a reasonable choice of the order would be somewhere in between 6 and 10.

LEGENDRE POLYNOMIALS

The Legendre polynomials that constitute an orthonormal basis on [0,24], can be calculated from following expression

$$P_m(t) = P_m^s\left(\frac{t-12}{12}\right)\sqrt{\frac{2m+1}{24}}, \tag{4.27}$$

where the standardized Legendre polynomials are given from the recurrence relation

$$mP_m^s(x) = (2m-1)xP_{m-1}^s(x) - (m-1)P_{m-2}^s(x), \tag{4.28}$$

with

$$P_0^s(x) = 1 \quad \text{and} \quad P_1^s(x) = x. \tag{4.29}$$

The curve estimates obtained with the Legendre polynomial approximation of orders 8 and 24 are shown in the Figures 4.13-4.14. The number of orthogonal functions entering these approximations are the same as those entering the two trigonometric approximations shown.

The results are very similar for the two different types of orthonormal functions, yet the trigonometric approximation seems to adopt a little better to the observations. In addition the Legendre polynomial approximation has some difficulties in connecting the two ends of the curve estimate at midnight. This is due to the choice of interval on which, the Legendre polynomials fulfill the orthogonality property. If the interval [-12,12] had been chosen the discontinuity problem had occurred at 12 o'clock. The discontinuity in the estimate is a major drawback to the use of Legendre polynomials in the present context.

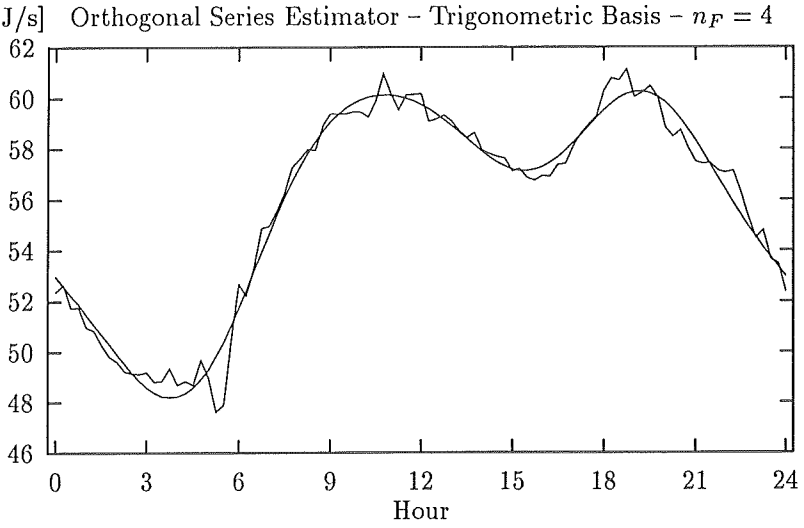


Figure 4.11: Trigonometric function approximation of order 4.

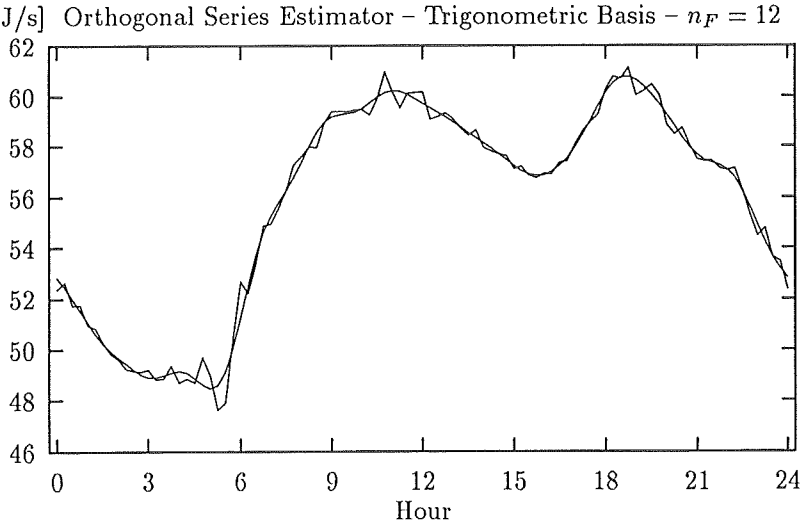


Figure 4.12: Trigonometric function approximation of order 12.

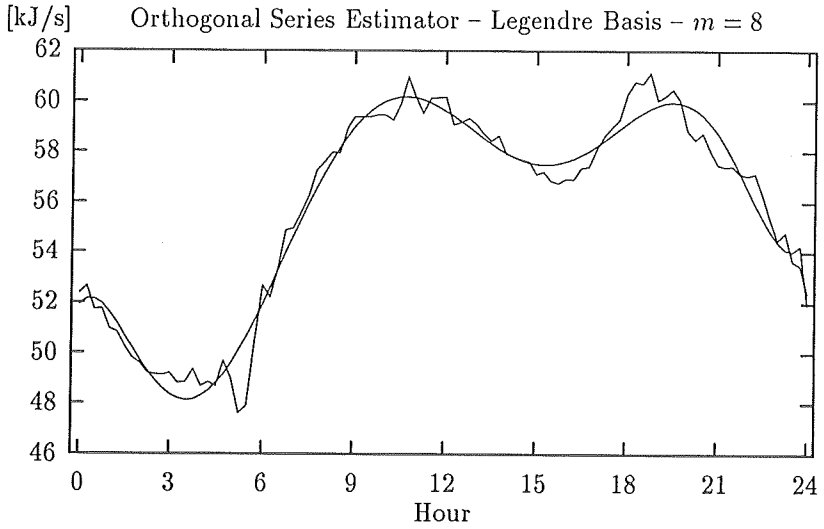


Figure 4.13: Legendre polynomial approximation of order 8.

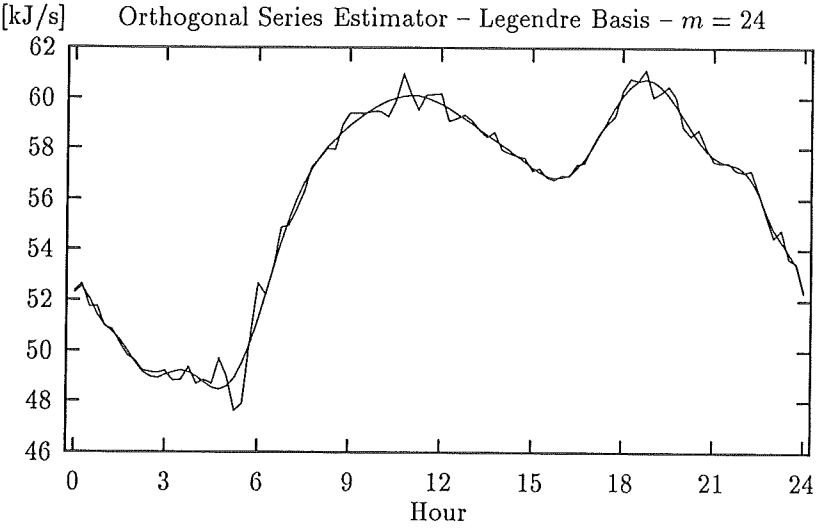


Figure 4.14: Legendre polynomial approximation of order 24.

4.6 DEPENDENCE ON TIME OF DAY, AMBIENT AIR TEMPERATURE AND SUPPLY TEMPERATURE

The investigation in this section serves the purpose of exploring the dependence of the heat load variations on the two most influential explanatory variables in DH systems, viz. ambient air temperature and supply temperature. It is clear that this dependence in both cases is dynamic meaning that a change in either of them has an influence on the heat load which evolves into the future. Nevertheless, in the first place a regression curve is determined for the description of momentary heat load as a function of these two explanatory variables. As the heat load is expected to follow a considerable systematic diurnal variation the variable being the time of day is also used as an explanatory variable. This makes the dimension of the explanatory vector equal to three. The differences between days is not taken into account in the current study. This is avoided to limit the number of explanatory variables.

The data used for this investigation were collected in the DH system in Esbjerg, Denmark during a period from August 14th till December 10th, 1989. This a period without any non-typical days meaning that the only difference between days is due to a typical weekly pattern, and this is considered to be of minor influence in the present context. These data is a subset of the data used in Chapter 5 for estimation of a number of dynamic models. The same data is shown in Figure 5.1, where the observations used here are numbered 1008 to 3843. In Figure 4.15 the heat supply versus ambient air temperature and supply temperature is plotted. The comprehension of this plot is improved by also looking at Figure 5.1. It is seen that there is a clear dependence between ambient air temperature and supply temperature. This is due to the way the supply temperature is controlled in relation to ambient air temperature. Furthermore, it is seen that the variation of data increases with falling ambient air temperature.

For the estimation of the regression surface the kernel method is applied using the Epanechnikov kernel (4.12). The product kernel is used, i.e. the

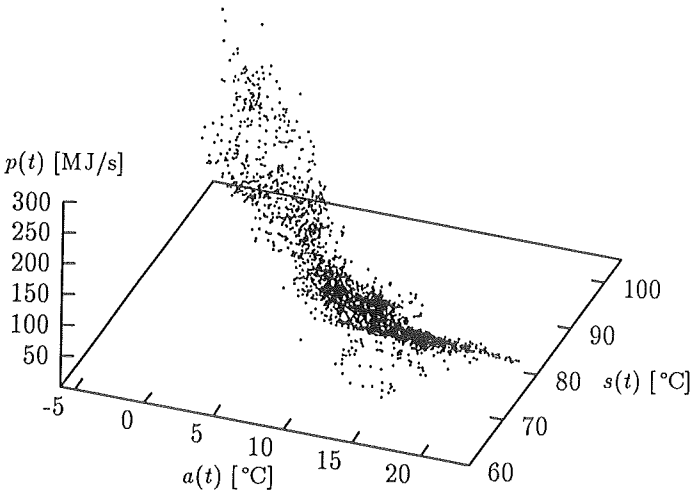


Figure 4.15: Heat load $p(t)$ versus ambient air temperature $a(t)$ and supply temperature $s(t)$ from August 14th to December 10th in Esbjerg.

curve estimate at time t, at ambient air temperature a and for supply temperature s is estimated as

$$\hat{p}_{h_t,h_a,h_s}(t,a,s) = \frac{\sum_{i=1009}^{3843} K_{h_t}(t-t(i))K_{h_a}(a-a(i))K_{h_s}(s-s(i))p(i)}{\sum_{i=1009}^{3843} K_{h_t}(t-t(i))K_{h_a}(a-a(i))K_{h_s}(s-s(i))}. \quad (4.30)$$

The optimal scale parameters for each of the kernels are assessed by evaluating the CV-function in a three-dimensional grid. The minimum CV-function was found in $(h_t, h_a, h_s) = (1.90, .56, .34)$ with a function value of 125.3. With these scale parameters the regression surface was estimated in a full grid consisting of $24 \times 50 \times 50$ points. The time-of-day parameter was applied for 24 hourly values corresponding to each hour of the day

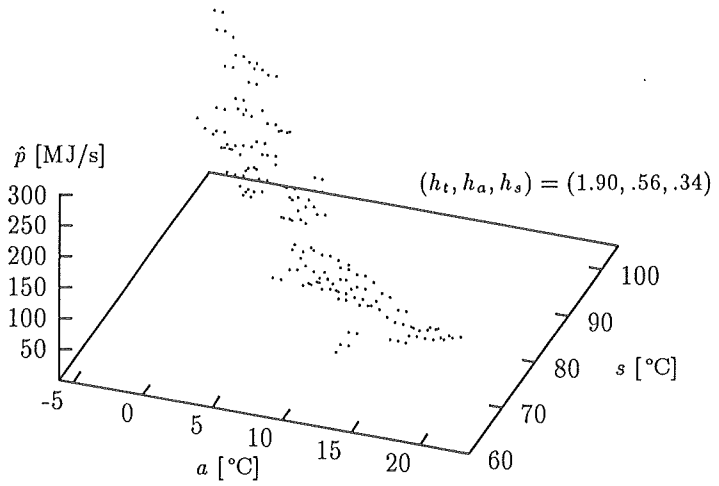


Figure 4.16: Epanechnikov kernel estimate of the three-dimensional dependence on time of day, ambient air temperature and supply temperature shown at the time of day $t_d = 7$.

and the applied values of ambient air temperature and supply temperature ranged from the minimum to the maximum observation of the data set in 50 equally spaced values. Figure 4.16 shows the surface estimates for t equal to 7 for those values in the grid where the density estimate was different from zero. As seen from the figure this resulted in a sparsely evaluated grid, and the grid estimates do not seem to lie on a smooth surface. Furthermore, in a considerable part of the shown points the estimate was actually based on a single observation. This implies that the estimate of the pointwise variance, as given in (4.10), becomes zero in these points. That is, the surface estimate corresponds to a single observation, and the variance estimate is not reliable. This result suggests that the bandwidths found in the minimum of the CV-function are too small, at least they are too small to give any surface interpretation.

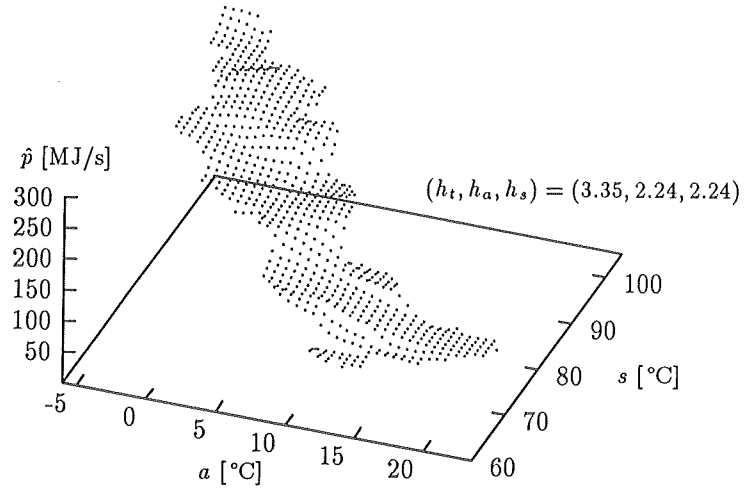


Figure 4.17: Epanechnikov kernel estimate of the dependence on time of day, ambient air temperature and supply temperature shown at $t_d = 7$.

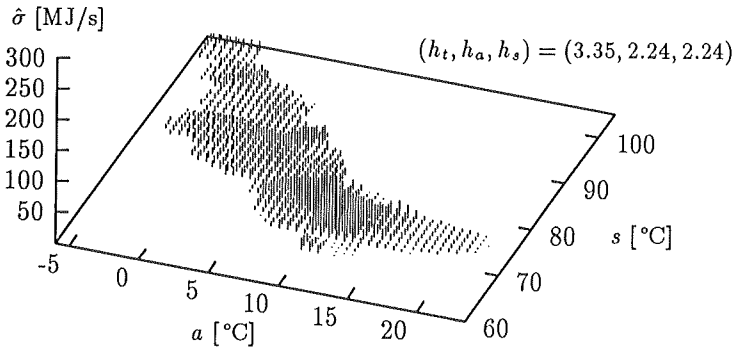


Figure 4.18: Pointwise estimate of standard deviation of the surface estimates.

In Figure 4.17 the surface estimate calculated in the same grid of points but with bandwidths $(h_t, h_a, h_s) = (3.35, 2.24, 2.24)$ is shown. These values are found by calculation of the surface estimate for varying bandwidths, and the estimate giving the best is determined by visual interpretation. For these values of the scale parameters the CV-function value is equal to 148.3. Figure 4.18 shows an impulse plot of the estimated pointwise standard deviation in the very same points. In these figures estimates are only shown in points for which the density estimate is bigger than 0.01 and the variance estimate is different from zero. For this choice of smoothing parameters the number of grid points has increased considerably and the interpretation of a surface is clear now. Although all the smoothing parameters are increased compared the optimal values found from cross-validation the surface does not seem to be over-smoothed.

For the highest values of a the surface seems to be a plateau on which the heat load is independent of actual ambient air temperature. From this plot it is not possible to see if the load also is independent of supply temperature. For decreasing values of a below 15 °C the heat load increases smoothly. At the same time there seems to be a positive dependence on supply temperature which is of more significance than the dependence on ambient air temperature. Possibly this is caused by the degree of smoothing being higher in the direction of ambient air temperature than in the direction of supply temperature. Actually the surface seems to be more humpy than what is reasonable from a physical point of view so probably the bandwidth for the supply temperature should be even bigger.

4.6.1 FILTERING OF SUPPLY AND AMBIENT AIR TEMPERATURE

It can be discussed whether it is reasonable to regress the heat load on momentary values of ambient air temperature and supply temperature. As a matter of fact it is not reasonable at all since, as outlined in Chapter 2, the influence of these variables is dynamic, and therefore the approach of

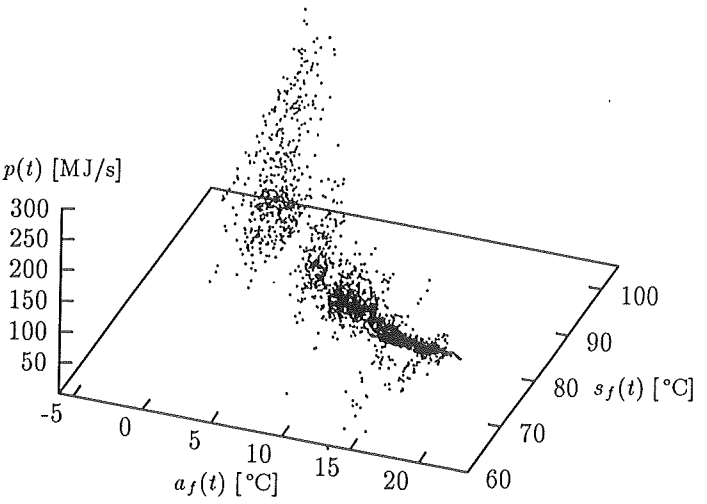


Figure 4.19: Heat load versus filtered ambient air temperature and filtered supply temperature from August 14th to December 10th in Esbjerg.

estimating a non-dynamic dependence relation surface is out of place.

Chapter 5 describes the result of estimating linear as well as nonlinear transfer function models. For a linear transfer function model, the model can be interpreted as a linear combination of filtered explanatory variables, see e.g. Ljung (1987). The approach here is to leave the linearity of the model, and use nonparametric regression to describe how the heat load depend on the filtered explanatory variables. In doing this it is expected that it is uncovered whether the linear dynamic model is insufficient and, if so, what the nonlinear dependence looks like. However, it is clear that the result will not be a complete dynamic model since the transfer function from the prediction errors is not taken into account.

The nonparametric regression is carried out in exactly the same way as above with the only change that ambient air temperature and supply tem-

perature are filtered through respective transfer functions. That is, the three explanatory variables are time of day, t_d, $a_f(t)$ and $s_f(t)$, where the two last-mentioned are given by transfer functions with unit dc-gain, viz.

$$a_f(t) = 0.22\frac{1 - 0.97q^{-1}}{1 - 1.32q^{-1} + 0.33q^{-2}}a(t). \tag{4.31}$$

and

$$s_f(t) = 2.22\frac{1 - 0.91q^{-1}}{1 - 0.80q^{-1}}s(t) \tag{4.32}$$

The poles and zeroes in these transfer functions are found in a linear transfer function model estimated on the very same set of data, see (5.34) in Chapter 5, and the coefficients in front of transfer functions are determined so the filters have unit gain. Figure 4.19 shows the heat supply versus the filtered variables. Compared to the non-filtered observations the spread of the ambient air temperature is reduced due to the low-pass filtering in (4.31), whereas the filtered supply temperature scatters over an increased range.

This is due to the placement of zero and pole in (4.32) which implies overshoot. It looks as if the filtered observations are more likely to lie on a surface than the raw data. Moreover, the filtered observations does not give the interpretation of a linear relation between ambient air temperature and heat load.

By calculating the CV-function in a three dimensional grid the optimum value was 36.0 for scale parameters $(h_t, h_a, h_s) = (1.90, .45, 1.57)$. The fact that the optimal CV-function is so much lower for the filtered variables indicates that these are more suitable for predicting the heat load.

The surface estimate for the optimum scale parameters is shown in Figure 4.20 for $t_d = 7$ with estimated pointwise standard deviation to match in Figure 4.21. Estimates are only shown for those surface points for which the density estimate was bigger than 0.01 and the variance estimate was different from zero. It seems to be the case that the smoothing parameters are too small to provide a reasonable number of point estimates. For this reason the estimation was carried out with higher smoothing parameters until

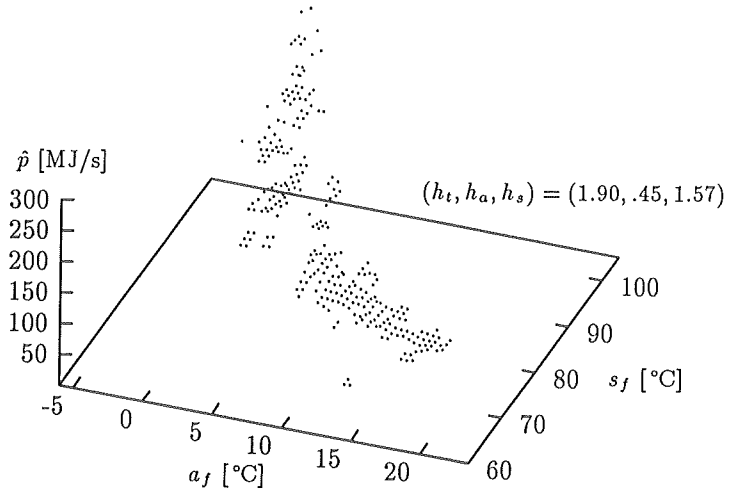


Figure 4.20: Estimate of the dependence on time of day, filtered ambient air temperature and filtered supply temperature shown at $t_d = 7$.

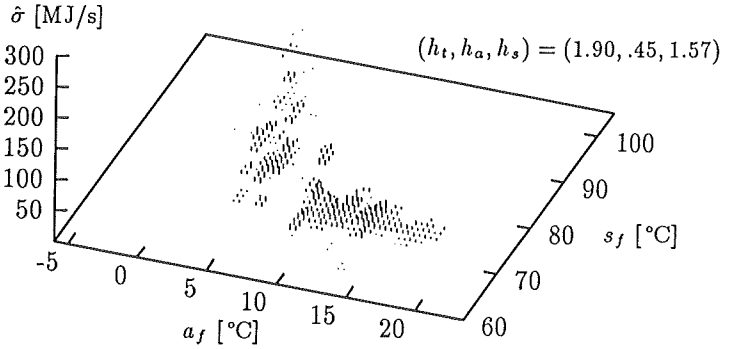


Figure 4.21: Pointwise estimate of standard deviation of the surface estimates.

a visually acceptable result was, viz. for $(h_t, h_a, h_s) = (3.35, 2.24, 2.68)$. For this bandwidth set the CV-function was 58.0. The resulting regression relationship and the pointwise standard deviations are shown for $t_d = 7$ in Figures 4.22-4.23, respectively. In addition, the surface estimate is shown from a different angle in Figure 4.24. Corresponding plots of curve estimate and pointwise standard deviations are shown in Figures 4.25-4.26 for $t_d = 19$.

In both cases the estimated regression curve indicates a relationship between ambient air temperature and heat load which is clearly nonlinear. That is, for ambient air temperatures above 16 °C the heat load seems to lie on a plateau meaning that the load does not depend on the value of these two explanatory variables. For decreasing ambient air temperature there seems to be a smooth transition to a linear dependence between load and filtered ambient air temperature. Below a temperature about 6 °C the linear dependence seems to have a different slope than above this temperature.

For ambient air temperatures below 16 °C there also seems to be a positive correlation between load and supply temperature. It is difficult to conclude more than the existence of a positive correlation between supply temperature and heat load. The surface shows some considerable changes in gradients, particularly about 75 °C. This is possibly due to the choice of bandwidth being a little too small implying the surface estimate to adopt too much of peculiarities of data.

4.7 PREDICTION USING NONPARAMETRIC REGRESSION

Predictions of the load in energy systems can be used as input to optimization of production and distribution of the required energy. For the purpose of optimization the predictions are usually made on-line. For each new observation the model describing the load is updated, after which it is used to make new predictions. Härdle (1990) mentions the possibility of using nonparametric regression for prediction purposes. The advantage of this approach is that no restriction of a fixed parametric model is im-

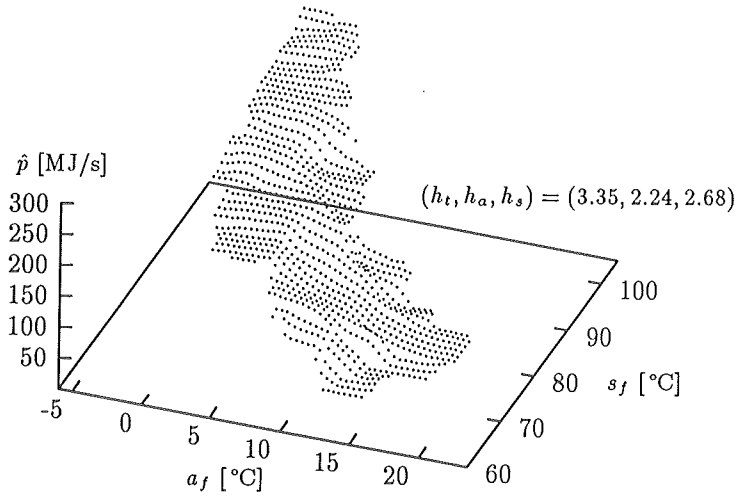


Figure 4.22: Estimate of the dependence on time of day, filtered ambient air temperature and filtered supply temperature shown at $t_d = 7$.

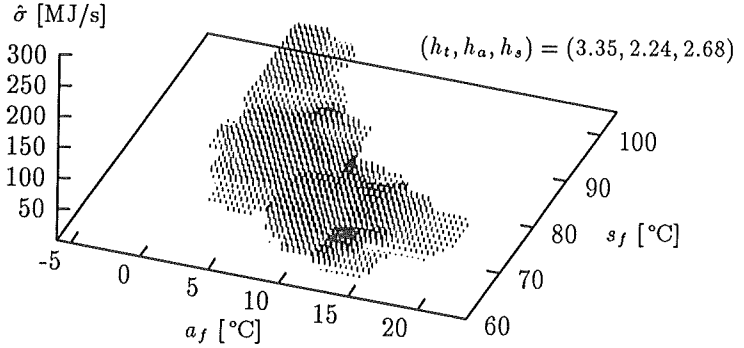


Figure 4.23: Pointwise estimate of standard deviation of the surface estimates.

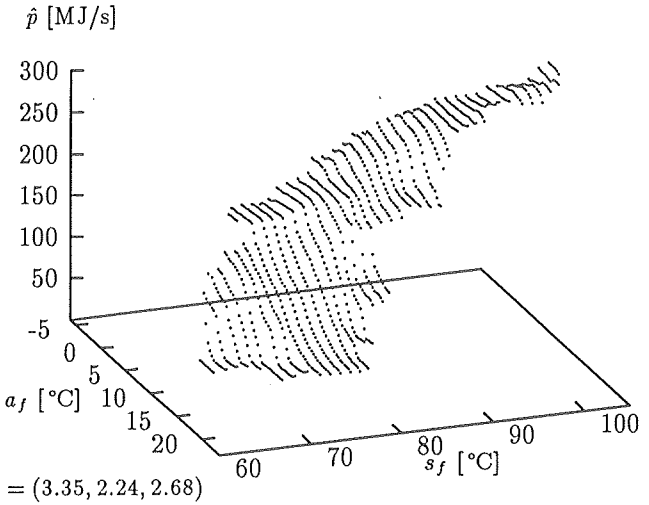


Figure 4.24: Estimate of the dependence on time of day, filtered ambient air temperature and filtered supply temperature shown at $t_d = 7$.

posed on the mapping from observations to prediction. The prediction is simply calculated as a weighted average of previously collected data, where the weighting in use is determined by the choice of method and the degree of smoothing. Accordingly any of the system features, which is actually described by data, may also be found in the prediction. This cannot be assured, when a parametric model, of restricted complexity compared to the system, is used. The drawback of applying nonparametric regression for prediction is that it is computationally inefficient to calculate the prediction as a weighted average of a large number of observations. However, the previous proposal for recursive calculation of nonparametric regression curves is also applicable to the prediction issue.

Based on the characteristics of the load in DH systems it seems to be a advantageous that the profile component of the model is estimated in the same step as the component describing the dynamics and the weather

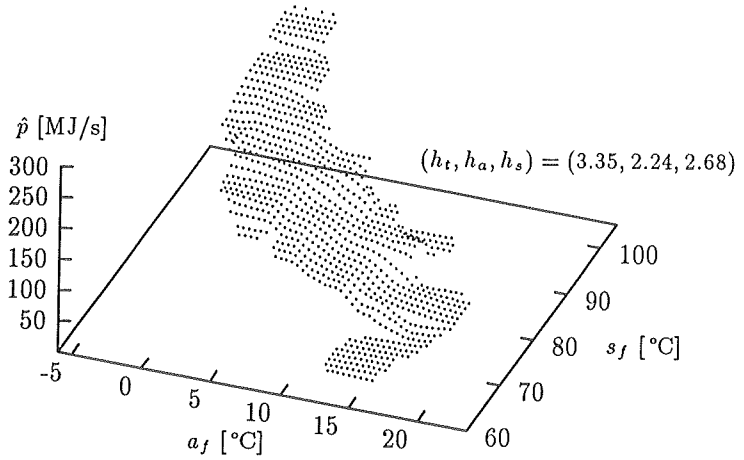


Figure 4.25: Estimate of the dependence on time of day, filtered ambient air temperature and filtered supply temperature shown at $t_d = 19$.

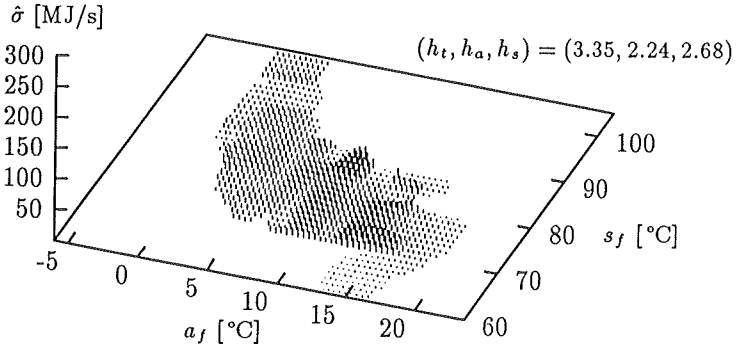


Figure 4.26: Pointwise estimate of standard deviation of the surface estimates.

dependence. In this situation the profile must be parametric. Then one of the orthogonal series estimators could be used as part of a composite model, but the components of the profile will then no longer be orthogonal.

For long-range prediction purposes results have indicated that it can be beneficial to let the profile describe as much of the cyclic variation as possible, see (Holst & Ekelund, 1987) and (Holst & Jonsson, 1984) for predictions of electricity load. This could be obtained by using the original data in the calculation of a profile and thereafter describing the residuals from the profile calculation by an ARMAX-model. In this case there are no restrictions on which kind of nonparametric regression method could be applied. Alternatively, minimization of long term prediction errors when using a composite model would lead to a model, in which the long-term prediction is a suitable weighting of a profile and an ARMAX-model.

4.8 CONCLUSION

In this chapter the application of different nonparametric regression methods on load data from DH systems has been investigated. Using data from a DH system where the distribution system can be neglected kernel estimates with Epanechnikov and Gaussian kernel as well as orthogonal series estimates using trigonometric functions and Legendre polynomials have been applied for describing the load as function of time of day. Furthermore, the kernel method has been used to obtain a surface estimate of the heat loads dependence on time of day, ambient air temperature and supply temperature.

Based on the first part of the investigation it can be concluded that all the considered smoothing methods are capable of adopting the diurnal variations of the heat load in the particular system. It is only a question of using an appropriate bandwidth in the kernel method or a suitable number of terms in the orthogonal series, respectively, to obtain an estimate with the desired flexibility and smoothness. There is no reason to believe that this should not apply to the diurnal variation in any DH system as the load

variations in general will follow a smooth curve.

For the kernel method the choice of kernel is not critical. The choice of bandwidth is of much more influence on the resulting estimate. To determine the optimal bandwidth the leave-one-out cross-validation method has been applied. This method, however, demonstrated a poor performance which possibly can be ascribed to the observations being correlated in time. It seemed to be a better approach to use visual inspection if only to get an estimate with the desired smoothness characteristic. Alternatively, improved methods for determination of bandwidths are required.

When using orthogonal series for estimation of a diurnal variation trigonometric functions seems to be suitable choice. The primary reason for this is that the regression curve bears the wrap-around property, i.e. the regression curve connects where the explanatory variable, being the time of day and night, changes at midnight. When using trigonometric functions with periods of $24, 12, 8, 6, 4, 3, \ldots$ the curve estimate will automatically have the same quality. Although the Legendre polynomials can be applied with as many degrees of freedom as wanted the curve estimate is not guaranteed to connect all the way around.

In the second part of the investigation the kernel method has been applied for exploring the dependence of heat load on ambient air temperature and supply temperature, respectively, in a setting where also the diurnal dependence is described. By using appropriate filtering of these variables the dynamic influence has, at least to a greater extent, been accounted for. This has implied the possibility of estimating a surface, which for the particular set of data clearly is observed to be nonlinear in ambient air temperature. A positive correlation between supply temperature and heat load is also disclosed.

The investigation in this chapter has demonstrated the use of nonparametric regression in exploring relationships in dynamic systems. This has involved the use to a large extent of plotting of curve and surface estimates. It is demonstrated that nonparametric regression, compared to raw data

plots, significantly enhance the exposure of dependencies in data. Due to this quality it is recommendable to use nonparametric regression in the identification phase of model building. Other applications of nonparametric regression, which are not investigated in this chapter, are prediction, outlier detection and determination of missing observations.

.....

All programming in the calculation of nonparametric regression results was performed in Fortran on a HP-9000/750 using double precision arithmetic.

.....

5

Dynamic Heat Load Modelling

In the present chapter the issue will be to construct a dynamic model for the load in a DH system. In the specification of the problem the delimitation of the DH system will be as given in Chapter 2. This means that the ambient weather is external to the system, and that the load model has the climate variables listed in Chapter 2 as potential explanatory variables. The supply temperature at heat supplying unit(s) are external control variables, and the power used by the pumps in the system is determining for the dynamic response and the transport delays in the system. Finally, the load definition given in Chapter 2 is applied.

Modelling of heat load in DH systems has in the recent years been subject of interest mostly for optimization purposes. Sejling (1987) estimated transfer function models using heat load observations from the DH system in Esbjerg, Denmark. These models were improved by introducing diurnal profiles implemented by trigonometric functions, and it is proposed to use recursive least squares with exponential forgetting for adaptive estimation of j-step predictors. Wiklund (1989) considered models of the same kind as well as state space models of diurnal and weekly profiles. Likewise Benonysson (1991) has applied similar models on sub-station loads in the DH transmission system in Ishøj, Denmark. Malmström (1991) has also used the approach of estimating the parameters of j-step predictors with RLS and exponential forgetting. Recently Waldemark (1993) has applied feed-forward neural network models for description of daily mean values of the load, and use of a neural net for prediction of the load 1,6,12 and 24

hours ahead is demonstrated.

The considerations in Chapter 2 relating to the dependence of load on, e.g. ambient air temperature, show that the linear model class is insufficient for describing the relation between the load and external variables. This is confirmed by the application of nonparametric regression in Chapter 4 to the dependence on ambient air temperature, supply temperature and time of day.

There are two approaches for obtaining a model, which describes the nonlinear relations. One approach is that of constructing an explicit modelling of the nonlinearities in the system. The model investigations in this chapter are examples of this approach. Another approach is that of using a model, which actually is insufficient, for instance, a linear model, and applying adaptive estimation methods to let the insufficient model approximate the system locally. Chapter 6 describes the results of applying adaptive estimation on heat load models.

A discussion of the advantages and drawbacks of the adaptive approach will be given here, by means of which a motivation for the investigation of the nonlinear modelling approach appears.

ADVANTAGES AND DRAWBACKS OF THE ADAPTIVITY APPROACH FOR MODELLING NONLINEAR SYSTEMS

The approach of modelling the system by adaptive estimation of a linear model is, in some respect, appealing due to the estimated model then being simply a data-determined local linear approximation to the nonlinear system. It is not likely that the dynamic relations in the system contain nonlinearities with abrupt changes, like thresholds. For instance, is it the case that there is a smooth transition for the dynamic relations between ambient weather and heat load, when passing from a winter situation to a summer situation. It is, however, possible that abrupt changes in the system configuration, e.g. connection of large groups of consumers, im-

plies abrupt changes dynamic relations. It is reasonable, if the changes occurring by changes in system configuration is excluded, to assume that a linear model will be able to provide a description of the load in a DH system, which is satisfactory in subranges of the total variation range of the external variables. The extension of the subranges and the degree of model fit depend, of course, on the degree of nonlinearity of the system. An adaptively determined model can thus be assumed to be a good approximation for the subranges in which the external variables recently have occurred.

However, there also exist some drawbacks of the adaptivity approach:

1. UNCERTAIN ESTIMATES. To obtain a locally valid model estimate only the recently sampled data should influence the estimates. In an adaptive estimation, e.g. RLS with exponential forgetting, this can be obtained by down-weighting the influence of observations as the distance to the observation in time increases. One consequence of such an approach is, however, that the effective number of data used in the estimation does not increase in the same way as for a recursive method, which is not adaptive. For instance, the effective number of observations for RLS with exponential forgetting never exceeds a certain limit. The limitation of the number of observations implies that the uncertainty of the parameter estimates will be larger compared to a non-adaptive estimation method.

2. DELAYED MODEL ESTIMATE. The center point of the effectively weighted data is some distance behind present, and therefore the model estimate will always be a delayed representation relative to the current conditions. Thus, the model estimate does not represent the system subject to the current operating conditions, but instead subject to what on average has occurred in the most recent past.

3. BAD PERFORMANCE FOR ABRUPT LEVEL CHANGES IN EXTERNAL VARIABLES. Even if the delay of the model estimate is ignored, and the linear model estimate is assumed to approximate the system well, subject to current conditions, sudden and considerable changes can occur just after present, which implies that the estimated model will

not perform well when used, for instance, for prediction purposes. In this case it might happen that the predictor becomes unstable, calling for the use of techniques for securing stability.

These drawbacks would all vanish if it was possible to determine a model being valid over the full range of variation of the external variables. Then the adaptivity would be superfluous as a tool for making local linear approximations. The adaptivity would, however, still be necessary as remedy for having the model estimate follow changes in system configuration.

5.1 MODEL BUILDING

When building a model of a system or a process it is of significant value to use physical insight and knowledge about relations among processes taking place in the system and processes influencing the system. A DH system, however, can be considered as a coupling of many individual sub-systems, each one being complex distributed dynamic systems described by partial differential equations and with built-in transport lags in the heat transfer. Therefore, a model describing the load will require knowledge about the system configuration, about the dynamics of the system components and about the hot water consumption and strategies of the consumers for control of indoor temperature. This level of detail in knowledge about a DH system is hardly ever available, and if it were, the model would end up being much too complex to be employed for modelling of heat load, and, moreover, it would be a most complicated task to estimate the parameters of such a complex model.

Instead of using the physical insight explicitly, the approach in this chapter will be to use knowledge about variables, which are known to have an influence on the heat load in DH systems, and to use statistical methods for determination of model orders.

In Chapter 2 it is discussed which external variables have an influence on the load in a DH system. Earlier results have shown (Sejling, 1987) that the

ambient air temperature, the supply temperature and the time and type of day are the by far most important external variables. For the purpose of elaborating on nonlinear dependences the investigation in this chapter will concentrate on forming a model using only these variables to explain the variations of the heat load. This is done in the conviction that use of the other external variables most likely will imply an improved description of the heat load.

Results will be given for four different model classes, viz.

1. General linear transfer function models.
2. Transfer function models for second order expansions of explanatory variables.
3. Smooth threshold transfer function models.
4. Neural network models.

5.2 DATA

The investigation is carried out using observations from the DH system in Esbjerg, Denmark. The data consists of two separate periods of observations, hereafter referred to as Data A and Data B, respectively. The observations used are hourly averages of:

- Heat supply.
- Temperature of the supply water from the CHP plant in Esbjerg.
- Ambient air temperature at the CHP plant in Esbjerg.

The observations in Data A were sampled from July 6th until December 10th, 1989, and in Data B from December 28th, 1989 until April 11th, 1990. In these two periods there are no holidays apart from in the first few days of Data B, and the observations from these days are not used for estimation or validation. For convenience in labelling the hours in the week

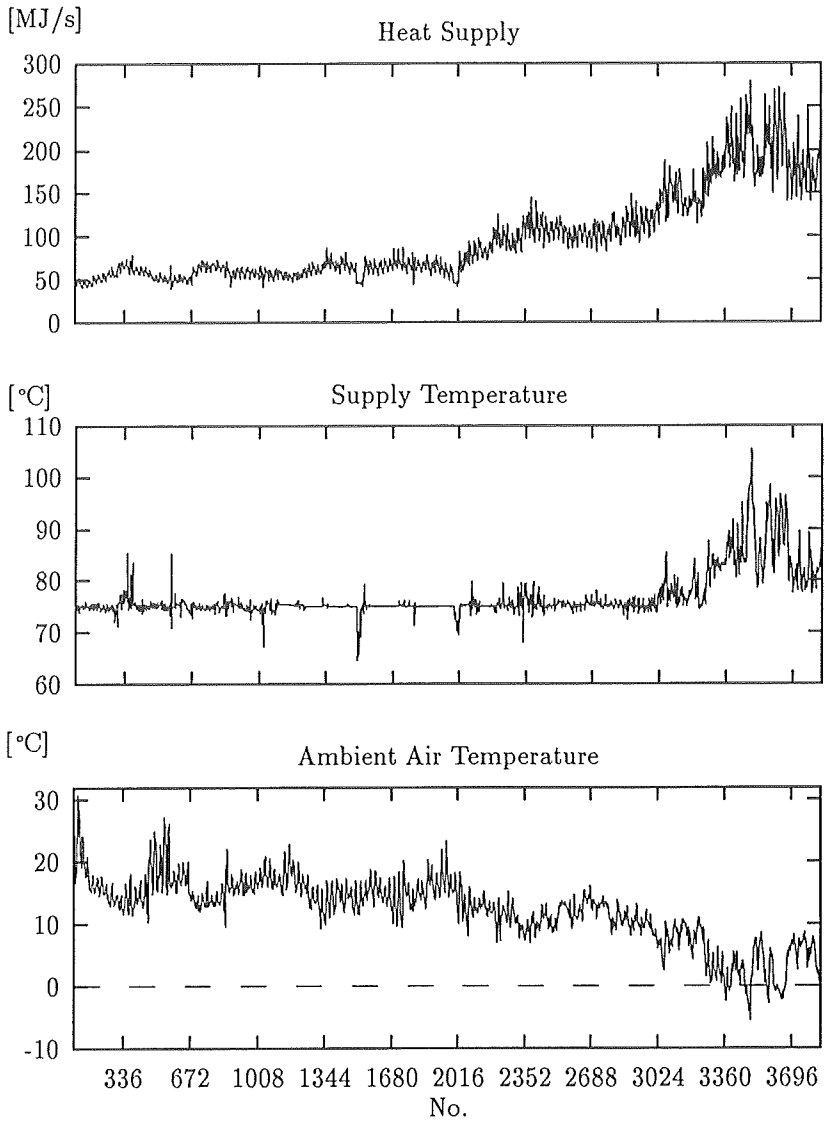


Figure 5.1: Heat supply, supply temperature and ambient air temperature in Esbjerg in Data A.

the observations are numbered starting with no. 1 in the first hour after midnight on Monday, July 3rd. Thus, the first observation in Data A has no. 89 and the last no. 3843, and Data B has observations in the interval 4283-6791.

Figures 5.1-5.2 show the observed heat load, supply temperature and ambient air temperature for the two data sets. Throughout the investigation Data A is used for estimation of model parameters, and Data B is used for validation of the models. That is, the estimated models are applied on a different set of data to test the general applicability of the model estimate. Data A is chosen for the estimation as this is the set of data, which has the largest variation range of the ambient air temperature.

Looking at the curve of the heat supply it is possible see a diurnal cycle, which also is occurring in periods without diurnal cycles in the ambient air temperature. It is remarkable how small the amplitude in summer is compared to that in winter. Since the load in summer almost surely is unaffected by the ambient air temperature, the cycles must be explained by hot tap-water and industrial consumption alone. In winter the ambient air temperature is expected to have a profound influence on the load. Thus, diurnal cycles in the ambient air temperature implies larger amplitudes in the diurnal cycles of the load, and furthermore, the use of night set-back, though it is redistributing the load between day and night, is amplifying the amplitude. Consequently, the relation between ambient air temperature and heat load depends on the actual level of the ambient air temperature, and the amplitude of the cycles, not explained explicitly by the ambient air temperature, probably also depends on the level of the ambient air temperature.

In Table 5.1 the averages and standard deviations of the load, the first order difference, the diurnal difference, the weekly difference as well as of multiple differences of the load for Data A are listed. The averages and standard deviations are calculated for two separate periods of almost equal length with the purpose of being able to make a quantifiable comparison of the two periods, and to be able to draw conclusions separately for the two

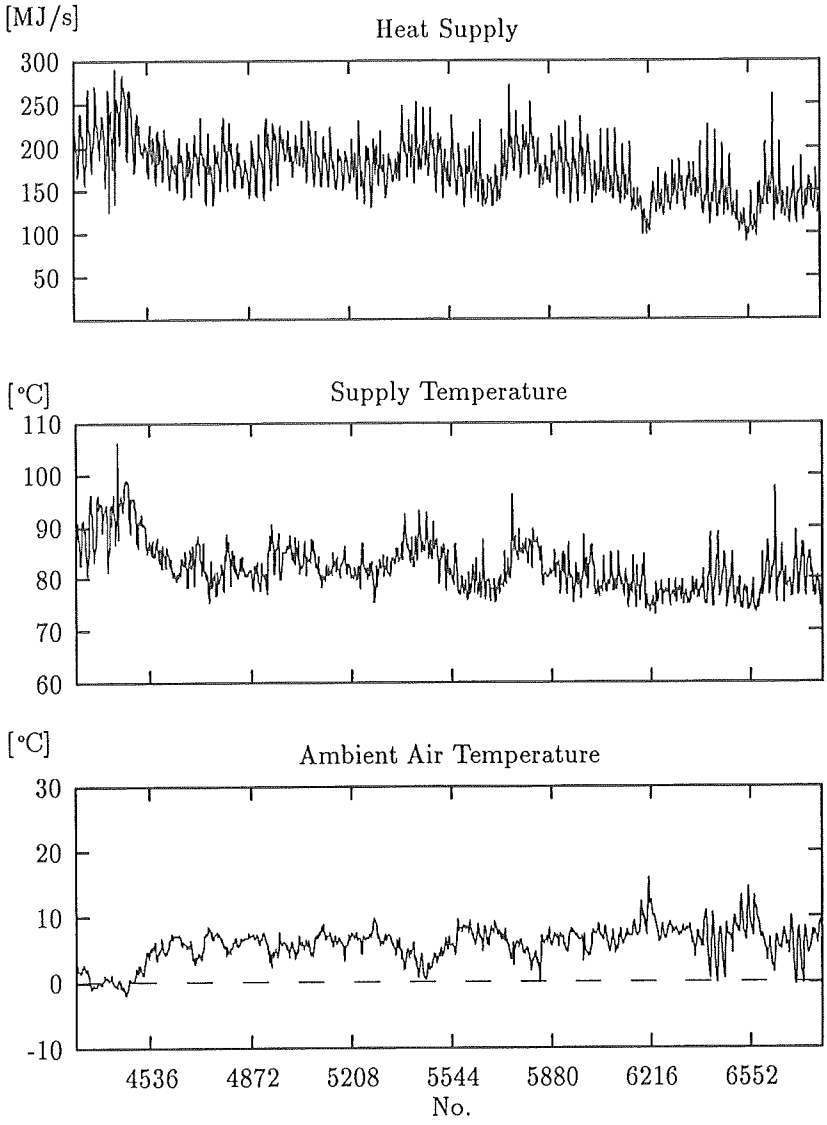


Figure 5.2: Heat supply, supply temperature and ambient air temperature in Esbjerg in Data B.

Data	Average		Std Dev	
	505-2172	2173-3843	505-2172	2173-3843
$p(t)$	61.87	141.59	9.54	44.29
$\nabla_1 p(t)$	0.02	0.05	2.77	7.24
$\nabla_{24} p(t)$	0.41	1.48	6.75	18.81
$\nabla_{168} p(t)$	1.51	10.18	11.44	26.28
$\nabla_{24}\nabla_{168} p(t)$	0.58	-0.79	9.10	22.63
$\nabla_1\nabla_{24} p(t)$	0.01	0.00	2.96	7.01
$\nabla_1\nabla_{24}\nabla_{168} p(t)$	0.02	0.02	3.95	8.98

Table 5.1: Descriptive statistics of the load data from Esbjerg ($\nabla_k p(t) = p(t) - p(t-k)$).

periods.

As seen from Figures 5.1-5.2 and Table 5.1 the characteristics of the load in these two periods are quite different. In the first period the level and the variation of the load are almost constant indicating that the correlation between ambient air temperature and heat supply is not as profound in summer as it is in winter time. In the second period the level and variation of the load seem to increase with falling ambient air temperature.

The statistics of the differenced series are shown in Table 5.1 in order to expose the different instationarities. It is clear that the first order difference significantly reduces the standard deviation. This value can also be interpreted as the standard deviation of the 1-hour ahead prediction error using a random walk, i.e.

$$p(t) = p(t-1) + e(t) \qquad \text{and} \qquad \hat{p}(t \mid t-1) = p(t-1), \tag{5.1}$$

where $\hat{p}(t \mid t-j)$ denotes the prediction of $p(t)$ using observations until $t-j$.

Likewise, the 24-hour difference reduces the standard deviation in both periods. This means that diurnal correlation is so strong that the 24-hour difference results in a variance reduction. On the other hand, it is seen that the 168-hour difference in all cases results in an increase of the standard deviation.

The statistics in Table 5.1 lead to the conclusion that one-hour differencing, 24-hour differencing or both are possible transformations of the observations before models are estimated.

5.3 ESTIMATION

All models considered in this chapter can be written on the form:

$$y(t) = g_\theta(y^{t-1}, u^t) + e(t) \tag{5.2}$$

with

$$y^{t-1} \equiv (y(t-1), y(t-2), \ldots) \tag{5.3}$$

and

$$u^t \equiv (u_1(t), u_1(t-1), \ldots, u_{n_u}(t), u_{n_u}(t-1), \ldots)\,. \tag{5.4}$$

That is, $y(t)$ is a function $g_\theta(\cdot)$, possibly nonlinear, of (y^{t-1}, u^t) superposed a zero mean i.i.d. random variable, $e(t)$, which is independent of (y^{t-1}, u^t). n_u is the number of external variables in the model.

The parameters θ in the model (5.2) is estimated by minimization of the sum of squared prediction errors (prediction error variance), see e.g. Klimko & Nelson (1978). This means that the estimate $\hat{\theta}_N$ is found as the parameters minimizing the criterion

$$V(\theta) = \frac{1}{N}\sum_{t=1}^{N}\frac{1}{2}\left(y(t) - \hat{y}_\theta(t)\right)^2 = \frac{1}{N}\sum_{t=1}^{N}\frac{1}{2}\varepsilon_\theta^2(t). \tag{5.5}$$

The prediction $\hat{y}_\theta(t)$ is obtained conditioning on the observed $\{y(t)\}$ until time $t-1$ and $\{u_i(t)\}$ until time t. That is, it is given as the expected

value conditioned on these observations and on the value of the parameter vector

$$\hat{y}_\theta(t) \equiv E\left[y(t) \mid \theta, y^{t-1}, u^t\right]. \tag{5.6}$$

The covariance matrix for the parameter estimates obtained by this estimation method,

$$\mathrm{Cov}\left\{\hat{\theta}_N\right\} = E\left[\hat{\theta}_N - \theta_0\right]\left[\hat{\theta}_N - \theta_0\right]^{\mathrm{T}}, \tag{5.7}$$

includes as the diagonal elements the variance of each of the estimates. θ_0 is the underlying true parameters in the model. The expected value in this expression is evaluated subject to the assumption that the probability density function of the total number of dependent variables entering the criterion, denoted y^N, is $f_y\left(\theta_0, \sigma_e^2; y^N\right)$.

A lower limit on the covariance matrix for any unbiased estimator, i.e. an estimator for which $E[\hat{\theta}_N] = \theta_0$, is given by the *Cramér-Rao inequality*

$$\mathrm{Cov}\left\{\hat{\theta}_N\right\} \geq M_N^{-1}, \tag{5.8}$$

where M_N is the *Fisher information matrix* given by

$$\begin{aligned} M_N &= E\left[\frac{\partial}{\partial\theta}\log f_y\left(\theta, \sigma_e^2; y^N\right)\right]\left[\frac{\partial}{\partial\theta}\log f_y\left(\theta, \sigma_e^2; y^N\right)\right]^{\mathrm{T}} \\ &= -E\left[\frac{\partial^2 \log f_y\left(\theta, \sigma_e^2; y^N\right)}{\partial\theta^2}\right]. \end{aligned} \tag{5.9}$$

For the model (5.2) with the innovations, $e(t) \equiv \varepsilon_{\theta_0}(t)$, being independent random variables with probability density function $f_e(x)$, it can be shown (Ljung, 1987) that the expectation in (5.9) can be approximated from data by

$$M_N^* = \frac{1}{\kappa_0}\left[\sum_{t=1}^{N} \psi_\theta(t)\psi_\theta{}^{\mathrm{T}}(t)\right] \Big|_{\theta = \theta_0}, \tag{5.10}$$

where

$$\frac{1}{\kappa_0} \equiv \int_{-\infty}^{\infty} \frac{[f_e'(x)]^2}{f_e(x)} dx \tag{5.11}$$

and

$$\psi_\theta(t) = \frac{d}{d\theta}\hat{y}_\theta(t) = -\frac{d}{d\theta}\varepsilon_\theta(t). \tag{5.12}$$

Furthermore, it can be shown that the parameter κ_0 equals the innovation variance σ_e^2 if the innovations are normal distributed.

For the LS estimator, subject to assumptions, which are taken to be fulfilled in the cases treated in this chapter, the parameter estimates are asymptotically central and normal

$$\sqrt{N}\left(\hat{\theta}_N - \theta_0\right) \in AsN(0, \boldsymbol{P}_\theta) \tag{5.13}$$

with

$$\boldsymbol{P}_\theta = \sigma_e^2 \left[\lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} \psi_\theta(t)\psi_\theta{}^{\mathrm{T}}(t)\right]^{-1} |_{\theta=\theta_0}\,. \tag{5.14}$$

For a finite number of data points the estimate of the covariance matrix for $\hat{\theta}_N$ is given by

$$\mathrm{Cov}\left\{\hat{\theta}_N\right\} = \tilde{\sigma}_e^2 \left[\sum_{t=1}^{N} \psi_\theta(t)\psi_\theta{}^{\mathrm{T}}(t)\right]^{-1} |_{\theta=\hat{\theta}_N} \tag{5.15}$$

with

$$\tilde{\sigma}_e^2 = \frac{1}{N}\sum_{t=1}^{N} \varepsilon_{\hat{\theta}_N}^2(t). \tag{5.16}$$

Note that ˆ is used for LS parameter estimates, whereas ˜ corresponds to ML estimates.

Comparing the covariance estimate in (5.15) to the Cramér-Rao lower limit in (5.10) shows that the lower limit actually is reached - the estimator is efficient - when the innovations are normal distributed. Non-normal distributed innovations imply $\kappa_0 < \sigma_e^2$ and hereby lack of efficiency, but in this case the following estimator is efficient (Ljung, 1987)

$$\tilde{\theta}_N = \arg\min_\theta \frac{1}{N}\sum_{t=1}^{N} -\log f_e\left(\varepsilon_\theta(t)\right). \tag{5.17}$$

This is the maximum likelihood (ML) estimator, which in the special case of normal distributed innovations simplifies to the LS estimator.

Subject to the assumption of normal II distributed innovations the likelihood function is

$$f_y(\theta, \sigma_e^2; y^N) = \prod_{t=1}^{N} \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left[-\frac{1}{2}\frac{\varepsilon_e^2(t)}{\sigma_e^2}\right], \tag{5.18}$$

and the log-likelihood becomes

$$\log f_y(\theta, \sigma_e^2; y^N) = -\frac{N}{2}\log 2\pi - \frac{N}{2}\log\sigma_e^2 - N\frac{V(\theta)}{\sigma_e^2}. \tag{5.19}$$

The ML estimator of the innovation variance, given in (5.16), is found as the solution to the derivative of the log-likelihood with respect to σ_e^2. Inserting this in (5.18) simplifies the expression

$$\log f_y(\theta, \tilde{\sigma}_e^2; y^N) = -\frac{N}{2}\log V(\theta) + \text{constant}. \tag{5.20}$$

Differentiating twice with respect to θ, and by using that

$$\frac{\partial}{\partial\theta} V(\theta) \mid_{\theta = \hat{\theta}_N} = 0, \tag{5.21}$$

lead to the following expression for the second order derivative of the log-likelihood evaluated in the LS estimate

$$\begin{aligned} \frac{\partial^2}{\partial\theta^2}\log f_y(\theta, \tilde{\sigma}_e^2; y^N) \mid_{\theta = \hat{\theta}_N} &= -\frac{N}{2} V(\hat{\theta})^{-1} \frac{\partial^2}{\partial\theta^2} V(\theta) \mid_{\theta = \hat{\theta}_N} \\ &= -\frac{N}{\tilde{\sigma}_e^2}\frac{\partial^2}{\partial\theta^2} V(\theta) \mid_{\theta = \hat{\theta}_N}. \end{aligned} \tag{5.22}$$

In this investigation the estimation is carried out by numerical minimization of the criterion (5.5). The procedure for minimization is described in Melgaard & Madsen (1991). The minimization routine provides the numerical approximation to the Hessian, $\hat{\boldsymbol{H}}(\hat{\theta}_N)$, obtained in the minimum of the criterion,

$$\hat{\boldsymbol{H}}(\hat{\theta}_N) = \frac{\partial^2 V(\theta)}{\partial\theta^2} \mid_{\theta = \hat{\theta}_N}. \tag{5.23}$$

Inserting (5.22) in (5.9), and by interchanging the expected value by the value obtained in the parameter estimate, gives a lower limit on the covariance matrix, which for the assumption of normal distributed innovations, corresponds to an estimate of the covariance matrix for the estimate of θ. That is, the estimate of the covariance matrix for the parameter estimate is given by

$$\widehat{\mathrm{Cov}}\left\{\hat{\theta}_N\right\} = \frac{\tilde{\sigma}_e^2}{N}\left[\hat{H}(\hat{\theta}_N)\right]^{-1}. \tag{5.24}$$

5.4 PERFORMANCE ASSESSMENT

To evaluate the performance of the investigated models two measures are used. The first one is the square root of the sum of squared prediction errors normalized by the number of prediction errors entering the sum, i.e.

$$SS = \sqrt{\frac{1}{t_2 - t_1 + 1}\sum_{t=t_1}^{t_2}\left(y(t) - \hat{y}_\theta(t)\right)^2}. \tag{5.25}$$

The interval range of SS does not necessarily coincide with the start and end of the summation in the minimized criterion.

In the measure given by (5.25) all prediction errors are given equal weight. Thus SS expresses the average standard deviation of the prediction errors in the period of evaluation. This measure is appropriate for comparison between different systems and/or situations if the variance of the load is independent of the level of the load. If, on the other hand, the load variance is proportional to the level, the following relative measure of the prediction ability allows for comparison between different situations, see e.g. Holst (1977),

$$SS^{\%} = \sqrt{\frac{1}{t_2 - t_1 + 1}\sum_{t=t_1}^{t_2}\left(\frac{y(t) - \hat{y}_\theta(t)}{y(t)}100\right)^2}. \tag{5.26}$$

In this measure each prediction error is compared to the corresponding load value and multiplied by 100. Hereby a relative measure in percent

is obtained, making the comparison of prediction ability of the load in different systems possible.

As both criteria (5.25) and (5.26) will steadily decrease when the number of parameters are increased, minimization of either of the with respect to the number of parameters is not a recommendable way of deciding upon an appropriate model order.

A more appropriate criterion for determining the model order is the Bayes Information Criterion (BIC), derived by Schwarz (1978), which apart from a constant is given by

$$BIC = N \log \tilde{\sigma}_e^2 + n_p \log N \tag{5.27}$$

N is the number of observations used in the estimation, and n_p is the number of parameters in the model. $\tilde{\sigma}_e^2$ is the ML estimator, (5.16), of the innovation variance for normal distributed innovations. An estimate of the appropriate model can then be found by minimizing BIC with respect to model order and model parameterization.

To evaluate the performance of a model fit it is appropriate to use an unbiased estimate of the innovation variance given by

$$\hat{\sigma}_e^2 = \frac{1}{N - n_p} \sum_{t=1}^{N} \varepsilon_{\hat{\theta}_N}^2(t). \tag{5.28}$$

5.5 MODELLING USING TRANSFER FUNCTIONS

In this section the results of the estimation of transfer function models are outlined. The estimation is carried out using Data A.

In the estimation the influence of possible transients of the chosen initial values is removed or reduced to insignificance by not using the prediction errors until observation no. 1008 in the criterion to be minimized. That is, the observations until no. 1008 are used only for to let the transients from initial values vanish.

For all of the transfer function models the predictions are obtained as the sum of components of the form

$$x_f(t) = \frac{B(q^{-1})}{F(q^{-1})}x(t), \tag{5.29}$$

where $x(t)$ can be either an external variable or the prediction error $y(t) - \hat{y}_\theta(t)$. B and F are both finite polynomials and F is monic.

In the practical implementation $x_f(t)$ is obtained recursively starting at time t_0 (as early in Data A as possible for the actual model) by using the recursion

$$x_f(t) = (1 - F(q^{-1}))x_f(t) + B(q^{-1})x(t). \tag{5.30}$$

This means that the filtered value at time t can be expressed, with explicit dependence on initial values, as

$$x_f(t) = \sum_{i=0}^{t-t_0+n_B} \pi_i x(t-i) + \sum_{i=1}^{n_F} \psi_i(t-t_0)x_f(t_0-i), \tag{5.31}$$

where n_B and n_F are the orders of B and F, respectively. The initial values are in the implementation chosen as $x_f(t) = x(t)B(0)/F(0)$, $t \in [t_0 - n_F; t_0 - 1]$.

By not using the prediction errors until $t > 1008$ in the estimation criterion, it is assumed that the weights $\{\psi_i(t-t_0)\}_{i=1}^{n_F}$ have become so small that the influence of initial values has vanished, and therefore the applied initialization is considered to be sufficiently good. Furthermore, all observations in Data A until no. 2016 represent the system in a typical summer situation implying that it actually is reasonable to leave out some of these observations to avoid too much influence of summer observations.

The weight sequences $\{\pi_i\}_{i=0}^{t-t_0+n_B}$ and $\{\psi_i(t-t_0)\}_{i=1}^{n_F}$ depend on the parameters and orders of B and F. In addition $\{\psi_i(t-t_0)\}_{i=1}^{n_F}$ depend on the time passed since start of the recursive calculation, $t - t_0$.

Writing down explicit expression for the weight sequences it turns out that if all the solutions to the equation $F(z) = 0$ (poles) have modulus less than

one (are inside the unit disc) then each of the weights $\psi_i(t-t_0)$ approaches zero for $t - t_0 \to \infty$. Moreover, the weights will be dominated by the pole being most close to the unit circle.

5.5.1 GENERAL TRANSFER FUNCTION MODELS

In this section only linear transfer function models are considered. The models in this class can be written on the form

$$A(q^{-1})y(t) = \sum_{i=1}^{n_u} b_0^i \frac{B^i(q^{-1})}{F^i(q^{-1})} q^{-k_i} u_i(t) + \frac{C(q^{-1})}{D(q^{-1})} e(t) \tag{5.32}$$

with $\{u_i(t)\}$ being external processes and $\{e(t)\}$ white noise with variance σ_e^2. This class is in Ljung (1987) referred to as *the general family of model structures* including a number of the commonly used model structures as special cases. For instance, the finite impulse response (FIR) model is obtained with all polynomials except the B^i ones being one, and the ARMAX model corresponds to case where the F^i and D polynomials are one.

Each of the external processes can either be generated independently of $\{y(t)\}$, or they may be processes that can be controlled.

For a model, actually being capable of describing the system, the innovation sequence does not contain any information about the relations in the system, neither dynamic nor stationary. Otherwise, this information could itself be described by a model, which therefore should be included in the original model structure. Thus the condition on the innovation sequence, for a sufficient model, is that it is a zero mean, independent and stationary process.

Each of the polynomials A, B^i, F^i, C and D are polynomials in the backshift operator. In the implementation they can all be specified as products of a number of lower order polynomials depending on which parameteriza-

tion is wanted. I.e. each polynomial in (5.32) is of the following form

$$G(q^{-1}) \equiv \prod_{j=1}^{n_G} G_j(q^{-1}) \equiv \prod_{j=1}^{n_G} 1 + g_{j,1}q^{-1} + \cdots + g_{j,n_{G_j}}q^{-n_{G_j}}. \qquad (5.33)$$

It is obvious, according to the results in Chapter 2 and Chapter 4, that a transfer function model being linear in supply and ambient air temperature will not be sufficient in describing the dynamic dependence on these variables. However, the purpose of showing the results with this kind of model is to allow for comparison with other models.

After estimation of a number of the most appropriate structures of the transfer functions from external variables and noise process, the model estimate below was found

$$\begin{aligned}
\nabla_{24}p(t) &= 3.57\,\frac{1-0.91q^{-1}}{1-0.80q^{-1}}\,\nabla_{24}s(t) \\
&- 1.62\,\frac{1-0.97q^{-1}}{1-1.32q^{-1}+0.33q^{-2}}\,\nabla_{24}a(t) + \\
&\left[-2.87\sin\left(\frac{2\pi t}{24}\right)-0.34\cos\left(\frac{2\pi t}{24}\right)+4.14\sin\left(\frac{2\pi t}{12}\right)+2.76\cos\left(\frac{2\pi t}{12}\right)\right]\nabla_{24}I_{\{\mathrm{A}\}}(t) \\
&+ \frac{1+0.05q^{-1}-0.91q^{-24}}{1-1.10q^{-1}+0.18q^{-2}}\,e(t). \qquad (5.34)
\end{aligned}$$

The variables in the model are

$p(t)$: heat load.

$s(t)$: supply temperature.

$a(t)$: ambient air temperature.

$I_{\{\mathrm{A}\}}(t)$ is an indicator function defined as one if the current hour is on a day differing from typical work days and zero otherwise. ∇_{24} is the diurnal difference operator defined by $\nabla_{24}x(t) = x(t) - x(t-24)$. This means that $\nabla_{24}I_{\{\mathrm{A}\}}(t)$ is equal to one on the first non-work day following a work day,

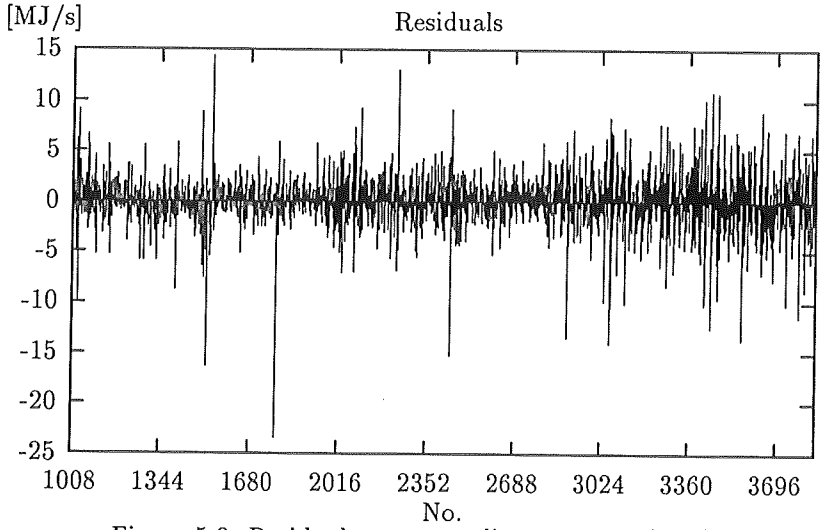


Figure 5.3: Residuals corresponding to model (5.34).

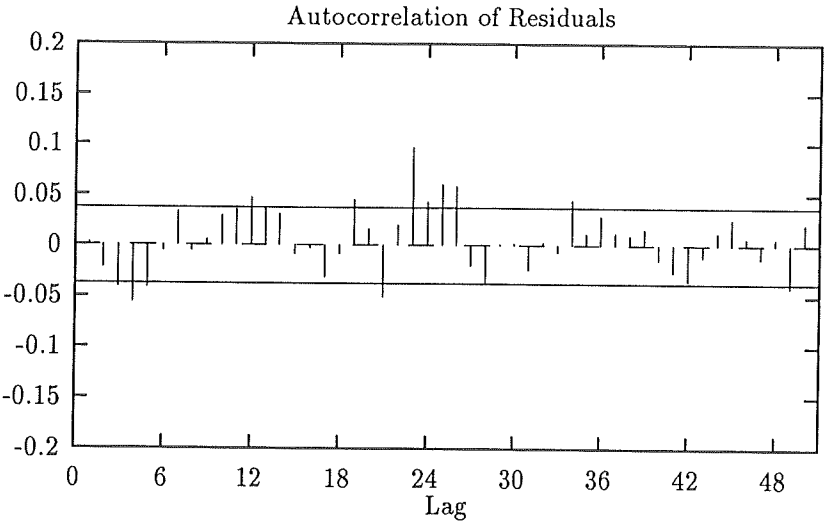


Figure 5.4: Autocorrelation of residuals corresponding to model (5.34).

Period	SS	$SS^{\%}$
1009-3843	2.40	2.73
1009-1953	1.98	3.71
1954-2898	2.09	2.39
2899-3843	3.01	1.72
$\hat{\sigma}_e^2 = 5.80$	$BIC = 5089$	

Table 5.2: Performance of model (5.34).

and equal to minus one on the first work day following a non-work day. For all other days is it zero.

All the parameter estimates of this model were found to be significant on 5 % level using the Hessian approximation as an estimate of the parameter covariance matrix. The estimated standard deviations of the parameter estimates can be found in Table 5.3.

This model includes an additive diurnal profile, composed of sine and cosine functions with periods 24 and 12 hours, for weekend days and days classified as non-working days. A diurnal profile common for all days is, in this particular model, redundant due to the diurnal differencing of both heat load and external variables.

Figure 5.3 shows the residuals, i.e. the difference between the load observations and the model predictions. The sample autocorrelation function of the residuals, shown on Figure 5.4, exposes that the model is not fully able to describe the diurnal variation of the heat load. One possible explanation to this is that the amplitude and form of the systematic diurnal variation is changing with the time of year and possibly also with the level of the ambient air temperature. The autocorrelation also shows that the model is not sufficient for describing a variation with a period of twelve hours. This

half-diurnal period, which obviously is not sufficiently described by the model, is probably due to the typical heat consumption having two peaks during the day. The first peak appears in the morning, when people take showers etc., and the other peak, being lower than the first one, appears in the early evening, when people cook dinner. These two peaks, with an intermediate distance of about twelve hours, must be the explanation to the half-diurnal correlation in the residuals. This leads to the conclusion that the diurnal profile is not sufficiently good in modelling the the systematic diurnal heat consumption, probably because this part of the load is not constant over the year. The autocorrelation at lag 23 is the most significant one. This can probably be ascribed to the shifting from summer time to standard time one night in the autumn. It is not clear how this shifting actually is compensated for in the collection of data the collection of data. Indeed, there is one observation for each of the 24 hours for all days in data, but is not known when and how the clock was changed in the computer that took care of the data collection. Furthermore, even if the shifting of the computer clock was done in the most appropriate way, first of all at the right time at night, it can not be excluded that the diurnal variation has changed. For instance, the part of the diurnal variation, coming from the diurnal variation in ambient weather condition, has shifted in relation to the hour of day, and this will imply a 23 hours correlation for the observations on the day following the night of shifting.

In Table 5.2 is listed the performance assessments of model (5.34). The class of models estimated in this section does not include models being able to describe neither time varying seasonal correlation nor time varying diurnal profiles. Therefore, it is not to be expected that (5.34) gives a complete description of these elements. It is probable that an adaptive estimation of the linear model would result in a better fit to data. The reason is that the adaptive estimation would allow the linear model to give an approximative description of the nonlinearities. In Chapter 6 examples of recursive estimation, with a forgetting factor, of models being linear in the parameters are shown.

5.5.1.1 ESTIMATION ON THREE DISJOINT SUBSETS OF DATA A

According to the system description in Chapter 2 and as elucidated in the introduction of this chapter, a model in the general family of model structures, being linear in supply temperature and ambient air temperature, will not be able to give a sufficient description of the heat load variations. In this section this point will be explored by using three disjoint sets of Data A in the estimation of the parameters in the model with the same structure as the previously estimated one, viz.

$$\begin{aligned}\nabla_{24}p(t) &= b_0^1 \frac{1+b_1^1 q^{-1}}{1+f_1^1 q^{-1}} \nabla_{24}s(t) + b_0^2 \frac{1+b_1^2 q^{-1}}{1+f_1^2 q^{-1}+f_2^2 q^{-2}} \nabla_{24}a(t) + \\ &\left[b_0^3 \sin\left(\frac{2\pi t}{24}\right)+b_0^4 \cos\left(\frac{2\pi t}{24}\right)+b_0^5 \sin\left(\frac{2\pi t}{12}\right)+b_0^6 \cos\left(\frac{2\pi t}{12}\right)\right] \nabla_{24} I_{\{\mathrm{A}\}}(t) \\ &+ \frac{1+c_1 q^{-1}+c_{24} q^{-24}}{1+d_1 q^{-1}+d_2 q^{-2}} e(t). \end{aligned} \tag{5.35}$$

A natural division of the data period appears from the course of variation of the ambient air temperature. In the interval [1009;2016] the ambient air temperature is fluctuating around a level of 16 °C. Thereafter the temperature seems to drop to a lower level and keeps on changing following a downward trend until around no. 3200, where it drops again and fluctuates around a level of 4 °C. Due to these observed differences, and taking into account a reasonable distribution of the number of observations in each disjoint set, the parameters in model (5.35) were estimated with the calculation period being either [1009;2016], [2107;3104] or [3105;3843]. The parameter estimates and the minimum criterion value for the three periods are shown in Table 5.3, along with those from the full period.

The lacking sufficiency of the model (5.35) becomes obvious when calculating the residual variance for the overall fit of the three model estimates. This gives $\tilde{\sigma}_e^2 = 4.83$ $(= \sum\sum \varepsilon_\theta^2/(n-3N_p))$, which indeed is much smaller than when the same model was fitted to Data A in full length. That is, a much better fit is obtained with three local model estimates.

Statistic	Criterion period			
	1009-3843	1009-2016	2017-3104	3105-3843
$\hat{b}_0^1$	3.57 (.05)	2.12 (.07)	3.31 (.09)	4.34 (.1)
$\hat{b}_1^1$	-0.91 (.02)	-0.88 (.02)	-0.89 (.09)	-0.93 (.02)
$\hat{f}_1^1$	-0.80 (.02)	-0.96 (.01)	-0.76 (.09)	-0.75 (.03)
$\hat{b}_0^2$	-1.62 (.07)	-0.81 (.1)	-1.42 (.1)	-2.52 (.2)
$\hat{b}_1^2$	-0.97 (.007)	-0.94 (.02)	-0.95 (.01)	-0.98 (.006)
$\hat{f}_1^2$	-1.32 (.04)	-0.97 (.1)	-1.37 (.08)	-1.30 (.06)
$\hat{f}_2^2$	0.33 (.04)	-0.02 (.1)	0.38 (.08)	0.31 (.06)
$\hat{b}_0^3$	-2.87 (.08)	-2.24 (.2)	-2.80 (.4)	-3.61 (.9)
$\hat{b}_0^4$	-0.34 (.07)	-0.42 (.2)	0.01 (.5)	0.10 (.6)
$\hat{b}_0^5$	4.14 (.2)	2.72 (.3)	4.39 (.3)	5.58 (.6)
$\hat{b}_0^6$	2.76 (.1)	1.58 (.2)	2.27 (.3)	3.91 (.5)
$\hat{c}_1$	0.05 (.008)	0.03 (.01)	0.06 (.01)	0.05 (.01)
$\hat{c}_{24}$	-0.91 (.007)	-0.92 (.01)	-0.91 (.01)	-0.91 (.01)
$\hat{d}_1$	-1.10 (.02)	-1.20 (.03)	-1.03 (.04)	-1.04 (.04)
$\hat{d}_2$	0.18 (.02)	0.31 (.03)	0.12 (.03)	0.14 (.04)
$\hat{\sigma}_e^2$	5.80	2.28	5.09	7.93
$\sum \varepsilon_\theta^2$	16365	2268	5460	5738

Table 5.3: Parameter estimates in model (5.35) for the different periods.

There is a number of notable differences for the three model estimates. As a first and important point the stationary gain from ambient air temperature is -4.94, -8.51 and -8.33 MJ/°Cs for the three periods, respectively (to calculate these values estimates to six decimal places are used). This is a significant difference, which, indeed, is in correspondence with the expected seasonal variation.

Turning to the dynamic relations the dominating pole of the transfer function from ambient air temperature is 0.992, 0.987 and 0.990 for the three periods, respectively. The fast reaction, on the other hand, differs in as much as the short term pole is almost zero (-0.038) for the first criterion period, whereas it is 0.406 and 0.317 for the two last periods, respectively. A likely interpretation of these results is that the dominating pole corresponds to an overall heat capacity of the DH system including distribution network and consumers' buildings, for which reason this pole can be assumed to be system dependent alone and therefore constant over the year. The other pole accounts for the faster heat load reactions to changes in ambient air temperature, and the stationary gain describes to which degree the ambient air temperature eventually influences the demand of heat. Indeed, it is in accordance with the expectation that these characteristics depend on the actual level of ambient air temperature. These differences in the estimated dynamic relation between ambient air temperature and heat load are illustrated by the step responses in Figure 5.5.

For the estimated response to a 1 °C step in supply temperature (Figure 5.6) it is again possible to observe a distinct difference between the summer and non-summer situation. In summer it seems that increasing the supply temperature eventually implies a considerable increase of necessary heat load, and moreover, the time constant of the transient is remarkably large. The stationary gain can be calculated to 6.97, 1.49 and 1.26 MJ/°Cs for the three periods. However, as the supply temperature is almost constant in the first subset, it is most likely that this quality of the model estimate for the first subset of the data is unreliable due to lacking excitation.

For the non-summer situation it is seen that the instantaneous increase in

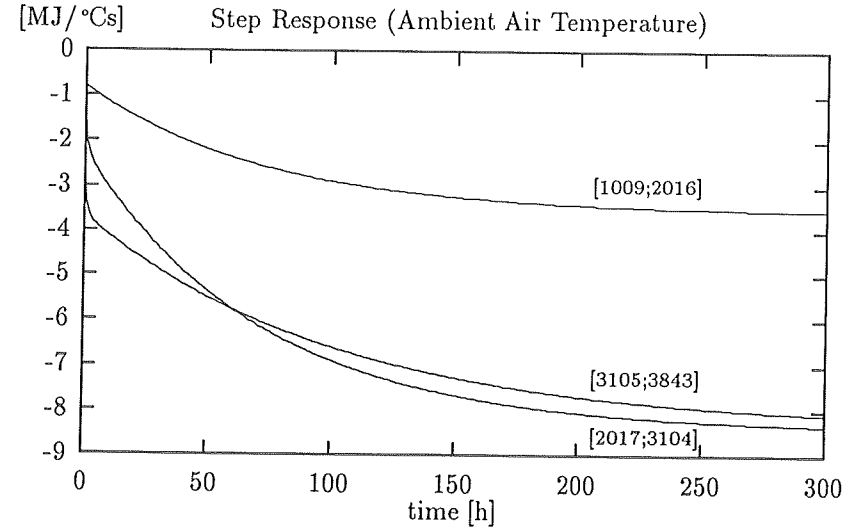


Figure 5.5: Response to a 1 °C step in ambient air temperature using the parameter estimates in model (5.35) for the three separate computation intervals.

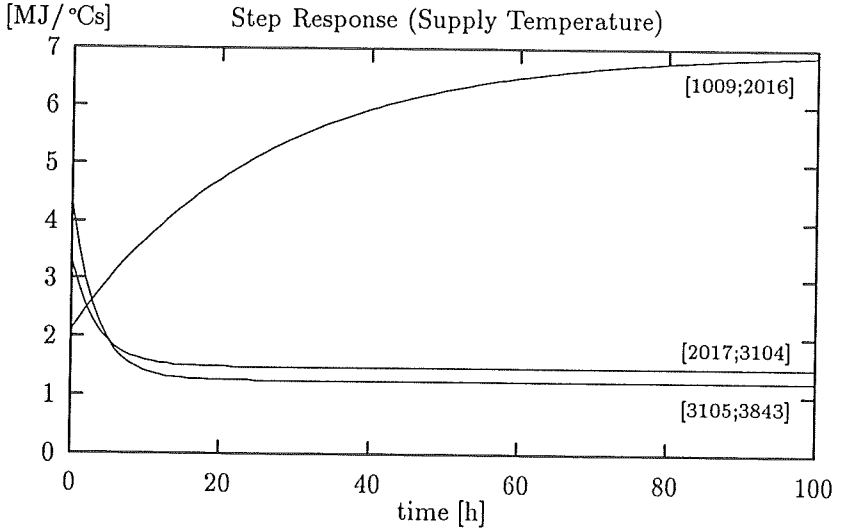


Figure 5.6: Response to a 1 °C step in supply temperature using the parameter estimates in model (5.35) for the three separate computation intervals.

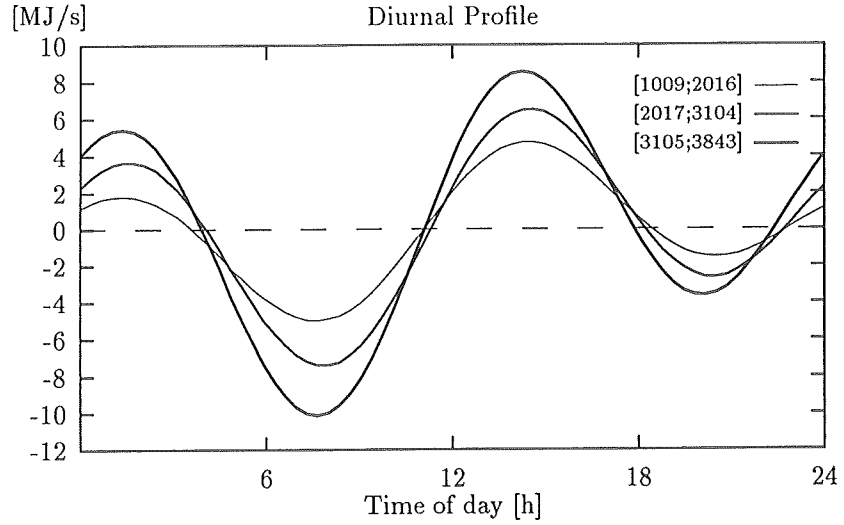


Figure 5.7: Diurnal trigonometric weekend profile in model (5.35) for the three separate criterion intervals.

heat load, following a 1 °C step in supply temperature, is higher than in the summer situation – the estimated instantaneous increase is presumably proportional to the average mass flow during the criterion period – but the additional heat load due to increase in supply temperature falls off quickly, and after about ten hours it has reached a stationary level between one and two MJ/°Cs. This stationary level can be interpreted as the stationary relation between heat loss and temperature level.

Figure 5.7 shows the trigonometric profile for the three periods. It seems that the phase content of the profile is constant (the curves cross each other close to 0 MJ/s), whereas the amplitude is different for the three periods.

The last component of the model to analyze with respect to differences for different criterion periods is the innovation filter. The parameters of the numerator are not changing significantly, whereas the poles for the first

period are (0.778;0.428) and (0.890;0.143),(0.885;0.151) for the two last periods, respectively. Again the difference is clear between the summer and the non-summer situation.

The results in this section has given a clear indication that the linear model, (5.35), is insufficient in describing the dependence on supply temperature and ambient air temperature. Likewise, the diurnal profile seems to show a variation in amplitude for the different periods, which is not described by the linear model. In the following sections nonlinear proposals for an improved description of the dependence on the external variables will be given.

5.5.2 SECOND ORDER NONLINEAR MODELS

The linearity of the transfer function model is evidently a restriction in the models' capability of describing load variations and dependence on external variables. Both stationary and dynamic relations obviously depend on operation point and weather condition.

In this section it is tried to improve the performance of the dynamic model (5.34) by introducing additional transfer functions with second order components. As will appear, this extension implies, e.g. that stationary dependency relations between the supply temperature and the heat load varies linearly with the inverse mass flow, being low-pass filtered. Likewise for the ambient air temperature the stationary dependence varies linearly with the low-pass filtered ambient air temperature. This kind of models belongs to the class of second order Volterra series expansions. In Priestley (1988) and Tong (1990) the general class of Volterra series expansions, as well as special nonlinear model structures, are described.

The response of the system to changes in the supply temperature must clearly depend on the mass flow, since the time delays through the system, in steady state, is proportional to the inverse of the mass flow. Hence, $\dot{m}_f^{-1}(t-1)$ is introduced in a supplementary transfer function from sup-

ply temperature to allow for an approximation of the mass flow dependent transfer function from the supply temperature to the heat load. This approximation of the model implies a linear relation between the stationary gain and the low-pass filtered inverse mass flow.

As the result of estimation of a number of candidate models with second order transfer functions the following model gave the lowest value of the information criterion (BIC)

$$
\begin{aligned}
\nabla_{24} p(t) \;=\; & \overset{(.04)}{6.30}\,\frac{1-\overset{(.01)}{1.05}\,q^{-1}}{1-\overset{(.02)}{0.83}\,q^{-1}}\,\nabla_{24} s(t) - \\
& \overset{(.02)}{9.32}\,\frac{1-\overset{(.03)}{1.22}\,q^{-1}}{1-\overset{(.02)}{0.87}\,q^{-1}}\,\dot{m}_f^{-1}(t-1)\,\nabla_{24} s(t) - \\
& \overset{(.08)}{2.88}\,\frac{1-\overset{(.002)}{0.98}\,q^{-1}}{1-\overset{(.02)}{1.29}\,q^{-1}+\overset{(.02)}{0.30}\,q^{-2}}\,\nabla_{24} a(t) + \\
& \overset{(.01)}{0.12}\,\frac{1+\overset{(.03)}{0.24}\,q^{-1}}{1-\overset{(.03)}{0.08}\,q^{-1}}\,a_f(t)\,\nabla_{24} a(t) + \\
& \left[-\overset{(.02)}{2.71}\sin\left(\frac{2\pi t}{24}\right)-\overset{(.02)}{0.18}\cos\left(\frac{2\pi t}{24}\right)+\overset{(.02)}{4.08}\sin\left(\frac{2\pi t}{12}\right)+\overset{(.02)}{2.54}\cos\left(\frac{2\pi t}{12}\right)\right]\nabla_{24} I_{\{\mathrm{A}\}}(t) \\
& + \frac{1+\overset{(.006)}{0.05}\,q^{-1}-\overset{(.005)}{0.91}\,q^{-24}}{1-\overset{(.02)}{1.08}\,q^{-1}+\overset{(.02)}{0.17}\,q^{-2}}\,e(t).
\end{aligned}
\tag{5.36}
$$

The filtered variables in this model are determined by exponential smoothing from

$$
\dot{m}_f^{-1}(t) = \frac{1-\hat{\alpha}_1}{1-\hat{\alpha}_1 q^{-1}}\,\frac{s(t)-r(t)}{p(t)}, \qquad \hat{\alpha}_1 = \overset{(.002)}{0.989}, \tag{5.37}
$$

Period	SS	$SS^{\%}$
1009-3843	2.22	2.30
1009-1953	1.59	2.84
1954-2898	2.04	2.26
2899-3843	2.84	1.64
$\hat{\sigma}_e^2 = 4.94$	$BIC = 4695$	

Table 5.4: Performance of model (5.36).

where $r(t)$ is the return temperature, and

$$a_f(t) = \frac{1-\hat{\alpha}_2}{1-\hat{\alpha}_2 q^{-1}} a(t), \quad \hat{\alpha}_2 \overset{(.001)}{=} 0.994 \,. \tag{5.38}$$

$\dot{m}_f(t)$ expresses, apart from a scale parameter, the low-pass filtered mass flow through the heat plant.

The parameters of the filters (5.37) and (5.38) were estimated simultaneously with the other parameters of the model. The filters implement a simple exponential smoothing with unit gain of the inverse mass flow and ambient air temperature, respectively.

Likewise, a transfer function from ambient air temperature with $a_f(t)$-dependent gain is introduced to obtain an approximation to the dependence of heat load on ambient air temperature. That is, the additional transfer function has the implication that the dynamic dependence on ambient air temperature is a function of the low-pass filtered ambient air temperature itself. For instance, this model can be given the interpretation that the stationary gain of the relation between heat load and ambient air temperature depends linearly on the level of the ambient air temperature (the low-pass filtered ambient air temperature).

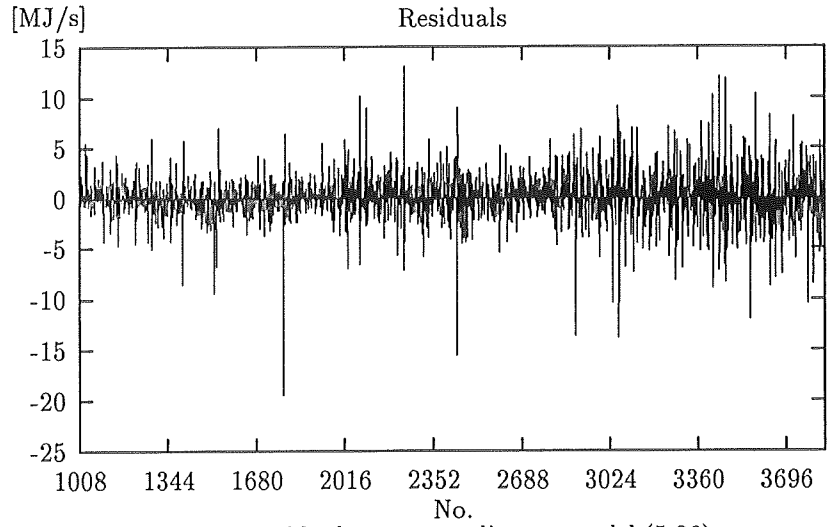


Figure 5.8: Residuals corresponding to model (5.36).

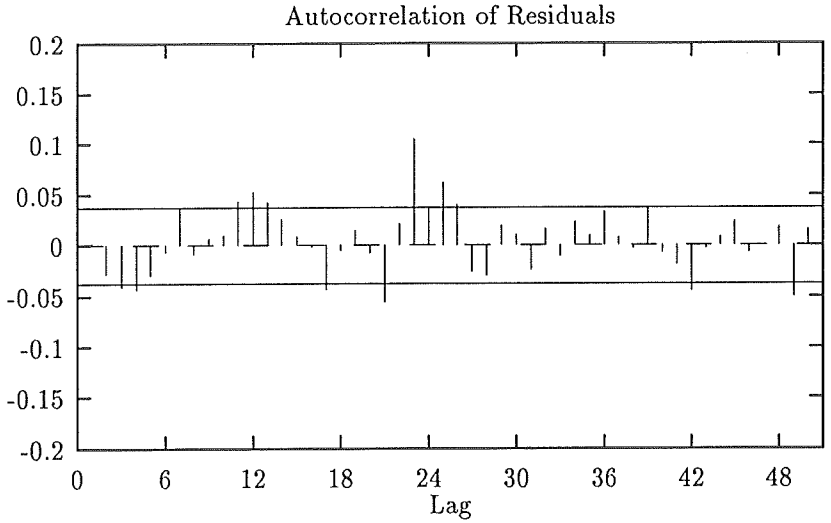


Figure 5.9: Autocorrelation of residuals corresponding to model (5.36).

A remarkable characteristic of the two separate transfer functions from supply temperature is the non-minimum phase zero-placement in both cases. However, it is the zero-placement for the composite transfer function that matters. To make an assessment of the zero-placement in the transfer function, which is composed of both transfer functions from supply temperature, it is assumed that the low-pass filtered mass flow can be considered as being independent of the supply temperature. Hereby the zero-placement of the composite transfer function can be given as a function of the low-pass filtered mass flow. It then shows that for all values of the low-pass filtered mass flow in an interval given by the range for which the low-pass filtered mass flow actually is seen in Data A, the zeroes are inside the unit circle.

Comparing Table 5.2 and Table 5.4 the improvement is obvious. Both the estimated residual variance and the information criterion, BIC, are reduced considerably. However, it is remarkable how the lowering of SS and $SS_{\%}$ is not equally pronounced in the different time periods. The improvement is clearest in the first period and smallest in the mid-period. From Figure 5.8 is also clear that in the first period the size of the large residuals is reduced considerably compared to Figure 5.3, whereas later on in the residual sequence the differences are not so obvious.

The estimated autocorrelation function of the residuals, shown in Figure 5.9, discloses that the seasonal correlation in the heat supply is not better described by model (5.36) than by model (5.34). The systematic course of the autocorrelation estimates is still evident, and the estimates around lag 24 has not become smaller.

5.5.3 SMOOTH THRESHOLD TRANSFER FUNCTION MODELS

As it was the case in the preceding section the objective of the investigation in this section is to obtain a description of the nonlinear influence of the external variables. The model class, here investigated, is the *smooth*

threshold autoregressive (STAR) model applied on the transfer function model, see e.g. Tong (1990).

The basic idea of using thresholds in general transfer function models is that hereby regimes are introduced in which one particular set of model components of the model is valid. Applying the principle of thresholds in model (5.32) gives the following general formulation of the model

$$A^{J_t}(q^{-1})y(t) = \sum_{i=1}^{n_u} b_0^{i^{J_t}} \frac{B^{i^{J_t}}(q^{-1})}{F^{i^{J_t}}(q^{-1})} q^{-k_i^{J_t}} u_i(t) + \frac{C^{J_t}(q^{-1})}{D^{J_t}(q^{-1})} e(t). \qquad (5.39)$$

This means that the total model actually consists of a number of models, all with the structure of a linear transfer function model, where in every time step only one of the models is in force. The value of J_t corresponds to the regime being in force at time t determining what model structure and parameters are valid at this moment.

The flexibility of this model structure is unlimited, i.e. any number of regimes can be used, whereby a nonlinear system can be described by a model being composed of local linear approximations each one being valid in their local regime. Furthermore, there is no limit on what process or state value actually determines the regime currently being in force.

However, the increased flexibility is obtained at the expense of an increased number of parameters. Therefore, the need of physical insight is essential in the search for an improved system description, the structure of which being without a superfluous number of regimes and with a suitable choice of regime determining processes.

For DH systems the situation of having hard bounds on the regimes seems to be non-physical. It certainly is the case that the effect of the hard bounds can be made as small as wanted just by using a sufficiently large numbers of regimes. However, smoothing out the crossings of the regime borders might enable the same model performance with fewer regimes.

Note that when estimating a model on three disjoint sets, as in Section

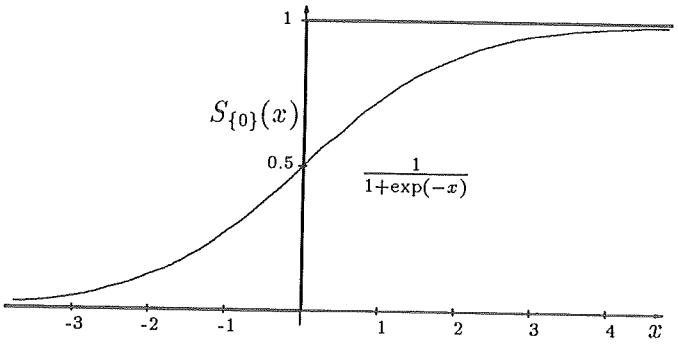


Figure 5.10: The unit step function and the standard logistic function.

5.5.1.1, it corresponds to a threshold model, where time is the regime determining variable.

Recognizing that an alternative formulation of (5.39) exists, where the regimes are obtained by means of the step function

$$S_{\{\alpha\}}(x) = \begin{cases} 0 & \text{if} \quad x \leq \alpha \\ 1 & \text{if} \quad x > \alpha \end{cases}, \tag{5.40}$$

it is seen that a smoothing can be made by interchanging $S_{\{\alpha\}}(x)$ by a continuous and smooth function, for instance, the logistic distribution function

$$F_{\alpha,\beta}(x) = \frac{1}{1 + \exp\left(\frac{-(x-\alpha)}{\beta}\right)}. \tag{5.41}$$

By approaching β to zero the logistic function approximates the step function. Thus, the model still holds the capability of describing abrupt changes.

Turning to the actual problem of modelling the nonlinearities of heat load dependence on supply temperature and ambient air temperature, the logistic distribution function is introduced in the transfer functions from these processes.

Among a number of different models with smooth threshold gain dependence the following resulted in the lowest value of BIC. The standard deviations of the parameters estimates are marked in parenthesis.

$$\begin{aligned}
\nabla_{24}p(t) \;=\; & \overset{(.1)}{5.42}\;\frac{1-\overset{(.02)}{1.01}\,q^{-1}}{1-\overset{(.03)}{0.81}\,q^{-1}}\,\nabla_{24}s(t) - \\
& \frac{\overset{(.1)}{3.40}}{1+\exp(\overset{(.1)}{4.08}-\overset{(.1)}{15.57}\,\dot{m}_f^{-1}(t-1))}\;\frac{1-\overset{(.04)}{1.21}\,q^{-1}}{1-\overset{(.05)}{0.85}\,q^{-1}}\,\nabla_{24}s(t) + \\
& \left(\frac{\overset{(.1)}{1.19}}{1+\exp(\overset{(.1)}{19.14}-\overset{(.04)}{1.52}\,a_f(t))}-\overset{(.1)}{1.98}\right)\;\frac{1-\overset{(.008)}{0.97}\,q^{-1}}{1-\overset{(.06)}{1.36}\,q^{-1}+\overset{(.06)}{0.37}\,q^{-2}}\,\nabla_{24}a(t) + \\
& \left[-\overset{(.2)}{2.38}\sin\left(\frac{2\pi t}{24}\right)-\overset{(.2)}{0.27}\cos\left(\frac{2\pi t}{24}\right)+\overset{(.1)}{4.02}\sin\left(\frac{2\pi t}{12}\right)+\overset{(.1)}{2.60}\cos\left(\frac{2\pi t}{12}\right)\right]\nabla_{24}I_{\{\mathrm{A}\}}(t) \\
& + \frac{1+\overset{(.008)}{0.05}\,q^{-1}-\overset{(.009)}{0.92}\,q^{-24}}{1-\overset{(.02)}{1.08}\,q^{-1}+\overset{(.02)}{0.17}\,q^{-2}}\,e(t). \qquad (5.42)
\end{aligned}$$

The filtered variables in this model are given by

$$\dot{m}_f^{-1}(t) = \frac{1-\hat{\alpha}_1}{1-\hat{\alpha}_1 q^{-1}}\,\frac{s(t)-r(t)}{p(t)}, \quad \hat{\alpha}_1 = \overset{(.006)}{0.974} \qquad (5.43)$$

and

$$a_f(t) = \frac{1-\hat{\alpha}_2}{1-\hat{\alpha}_2 q^{-1}}\,a(t), \quad \hat{\alpha}_2 = \overset{(.009)}{0.951}\,. \qquad (5.44)$$

The transfer function from supply temperature is seen to consist of two separate transfer functions, one of which has a gain being dependent on the low pass filtered scaled inverse mass flow, lagged one hour. Hereby, the composite transfer function at time t includes parameters being function of $\dot{m}_f^{-1}(t-1)$. $F_{\alpha,\beta}(\dot{m}_f^{-1}(t-1))$ is introduced to allow for an approximation to a mass flow dependent transfer function from the supply temperature to the heat load. As for the preceding nonlinear model each of the transfer

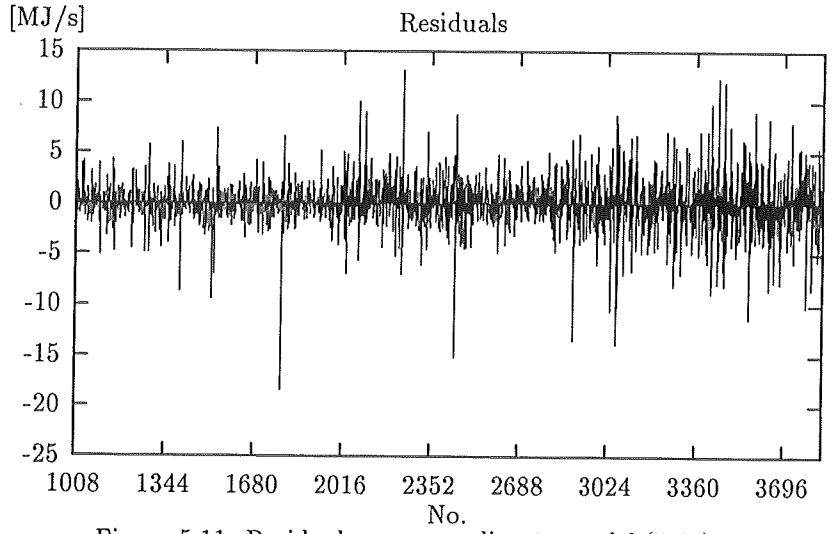


Figure 5.11: Residuals corresponding to model (5.42).

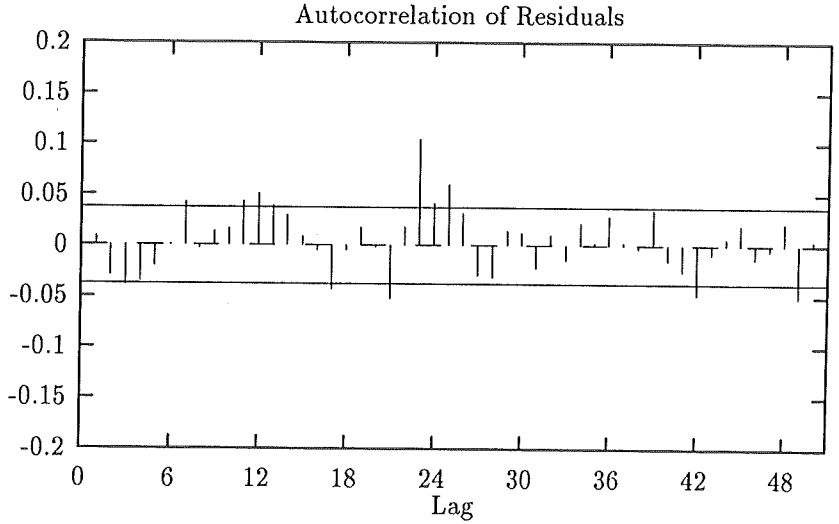


Figure 5.12: Autocorrelation of residuals corresponding to model (5.42).

Period	SS	$SS^{\%}$
1009-3843	2.20	2.27
1009-1953	1.56	2.78
1954-2898	2.03	2.24
2899-3843	2.83	1.64
$\hat{\sigma}_e^2 = 4.89$	$BIC = 4672$	

Table 5.5: Performance of model (5.42).

functions is non-minimum phase. However, for values of the mass flow, which from data are found as realistic the composite transfer function turns out to be minimum phase.

In the transfer function from ambient air temperature the gain is a function of the filtered ambient air temperature, through the smooth threshold. This implies that in stationarity the relation between ambient air temperature and heat load is not linear but s-shaped like the logistic function. The dynamic response is, of course, also different from that in model (5.34). This extension is introduced to allow for an approximation of the dependence of the transfer function from ambient air temperature to heat load on the actual level of the ambient air temperature.

The performance of the smooth threshold model, Table 5.5, is a little better than for the second order model. In spite of the increased number of parameters the BIC is lower for the smooth threshold model. However, looking at the residuals and the empiric autocorrelation of these it is difficult to observe any change in pattern and size.

5.5.4 ANALYSIS OF MODELLING RESULTS

In this section the results for the three model types considered previously are further analyzed.

First the DC-gain for the three models, $H^s(e^{i\omega})\,|_{\omega=0}$, in the transfer from supply temperature are compared. For the linear model the DC-gain is constant, $H_s(1) = 1.61$ MJ/°Cs, whereas for the second order model there is a linear dependence between gain and $\dot{m}_f^{-1}$, i.e.

$$H_s(1) = 15.4\dot{m}_f^{-1} - 1.89 \text{ MJ/°Cs}. \tag{5.45}$$

For the smooth threshold model the DC-gain is

$$H_s(1) = \frac{4.86}{1 + \exp(4.08 - 15.6\dot{m}_f^{-1})} - 0.389 \text{ MJ/°Cs}. \tag{5.46}$$

These relations are shown on Figure 5.13. For the observations in Data A $\dot{m}_f^{-1}$ is typically 0.5 °Cs/MJ in the summer period and between 0.2 and 0.3 °Cs/MJ in the coldest period. Hereby it is seen that both the threshold model and the second order model bear the, possibly erroneous, quality of postulating a large DC-gain for the summer observations. This was also observed for the differentiated linear model estimate.

Physically it is more likely that the DC-gain is almost independent of both $\dot{m}_f^{-1}$ and time of year. If any, the likely correlation is that the DC-gain increases with falling ambient air temperature and increasing supply temperature due to the heat losses from the distribution system being positively correlated with the difference between the temperature level in distribution system and the temperature of the material surrounding the pipe lines. On the other hand, in colder periods are the pumps typically supplied with more power than in warmer periods, since the heat supply increases with flow rate and the pressure drops generally increase with the heat consumption in the system. This implies an increased supply of heat from the pumps in the system, which may be an explanation to the lower DC-gain for lower ambient air temperatures.

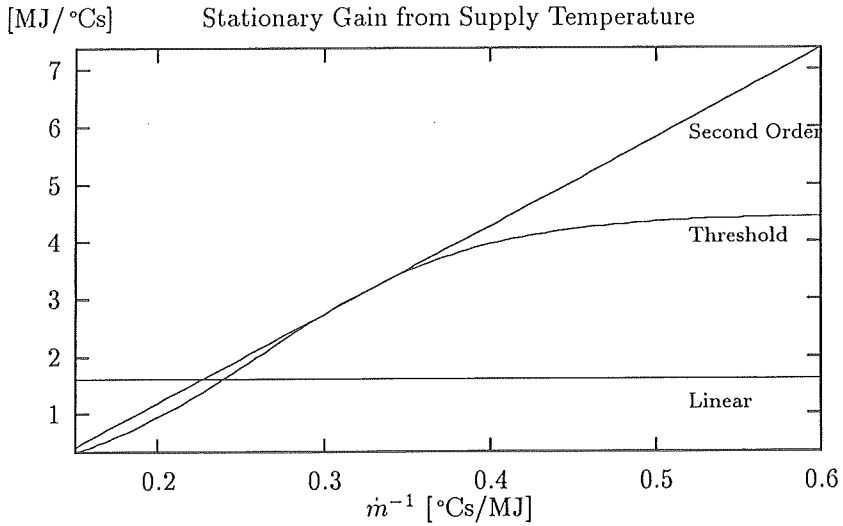


Figure 5.13: Stationary Gain in transfer function from supply temperature.

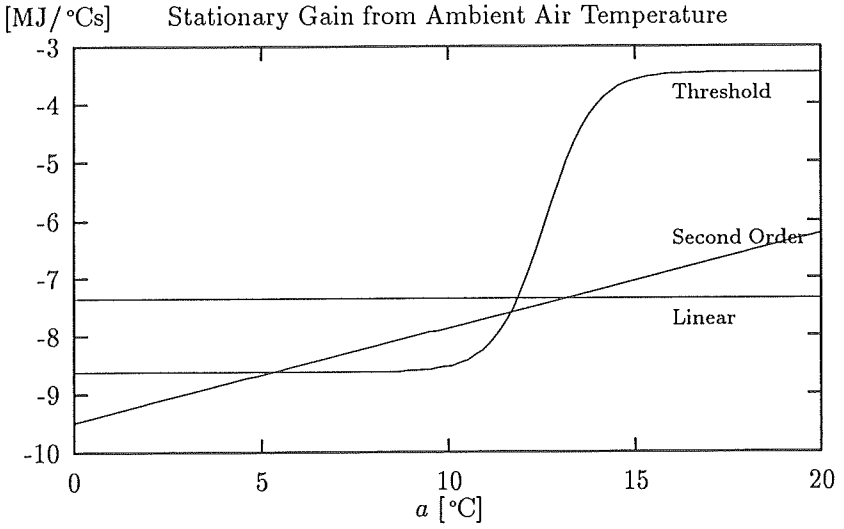


Figure 5.14: Stationary Gain in transfer function from ambient air temperature.

In the transfer functions from ambient air temperature to load the stationary gain is found as the fraction between the stationary increase in load and an infinitesimal increase in ambient air temperature, when contributions from all other model components are set to zero. Thus, for the second order model the stationary gain becomes

$$H_a(1) = 0.163a - 9.48 \text{ MJ/°Cs}, \tag{5.47}$$

while for the smooth threshold model it is

$$H_a(1) = \frac{5.17}{1 + \exp(19.1 - 1.52a)} - 8.62 \text{ MJ/°Cs}. \tag{5.48}$$

For the linear model the gain is $H_a(1) = -7.35$ MJ/°Cs. These relation between stationary gain and ambient air temperature, shown in Figure 5.14, seem reasonable. For the second order model the relations between gain and ambient air temperature is linear with a slope of the right sign and with a crossing of the constant gain in the linear model at about 13 °C. The threshold model shows a two-level gain with a smooth transition for ambient air temperature from 10 °C to 15 °C. The placement of the two levels is reasonable, but the transition between the two levels seems to take place at an ambient air temperature placed a little lower than expected. Presumably is there a demand of heating, when the ambient air temperature is below the typically required indoor temperature (around 20 °C), pointing towards an increased correlation between ambient air temperature and heat load for temperatures just below 20 °C. On the other, are there heat contributions from other sources, viz. light, machines and people themselves, which probably imply a lowering of the ambient air temperature point, at which supply of heat from radiators sets in.

For the nonlinear models the dynamic relation between supply temperature and heat load is outlined by showing a pseudo step response. That is, the model response to a step in supply temperature is calculated assuming that $\dot{m}_f^{-1}$ is uninfluenced by the supply temperature change. This is, of course, not in accordance with the true system reaction and not in accordance with the model either. The reason the model response is calculated in this simplified way is that the flow changes following the supply temperature

step is not known. The pseudo step response can nevertheless still be used in comparison of the dynamic relations of the models.

The real system reaction to changes in supply temperature is that no momentary change in flow takes place, but transiently and stationarily an increase will imply a decrease in the flow (assuming the return temperature is left uninfluenced). The reason for this is that the flow is determined by the consumers (and the control of pumps) dependent on the temperature of the district heating water. Therefore, since the change of plant supply temperature does not arrive at the consumers immediately, this change of supply temperature will first influence the consumers' flow demand sequentially as the temperature change spread out in the system.

The step response to a 1 °C increase in supply temperature is shown for three constant levels of the flow and for the two nonlinear models on Figure 5.15 and Figure 5.16, respectively. For both models the unrealistic stationary gain for the high level of inverse flow is recognized. However, the momentary load increase corresponds to the expectations for both models, and this also goes for the lower levels of inverse flow. That is, an increase in supply temperature is followed by an increase in required heat production, which thereafter falls off to a constant level being somewhat above zero.

The model response to a 1 °C step in ambient air temperature is found as the model prescribes, i.e. when calculating the response using the relevant transfer function the low-pass filtered temperature step is applied. For this reason the step response depends on the level before and after the step as well as the size of the step. Figures 5.17-5.18 show for the two nonlinear models, respectively, the response to a 1 °C increase in ambient air temperature at three different levels. First of all it is seen that the difference in between the responses for different levels of ambient air temperature is bigger for the threshold model than for the second order model. This corresponds to the previously shown gain variations. Likewise, for the step at low ambient air temperature level the response is almost the same for the two models, whereas for higher ambient air temperature levels the response seems to be too big in the second order model.

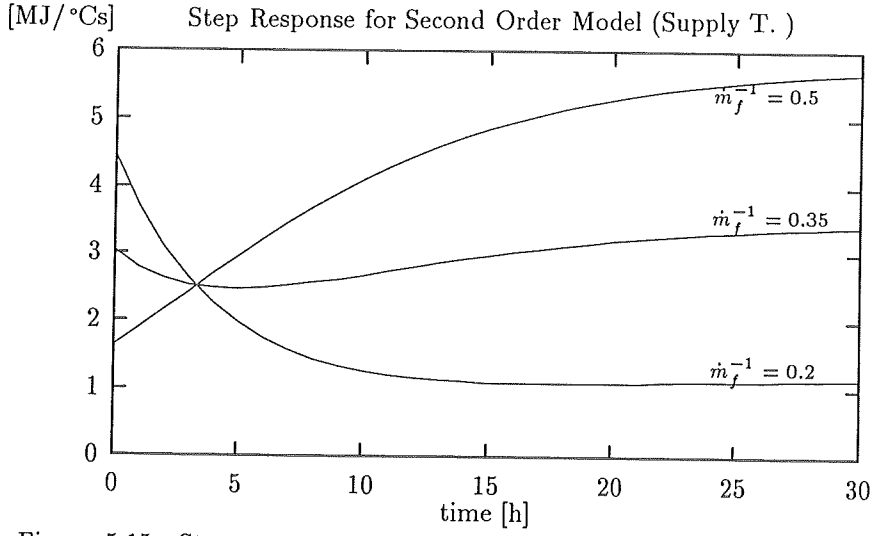


Figure 5.15: Step response to a 1°C increase of supply temperature in second order model for different flow.

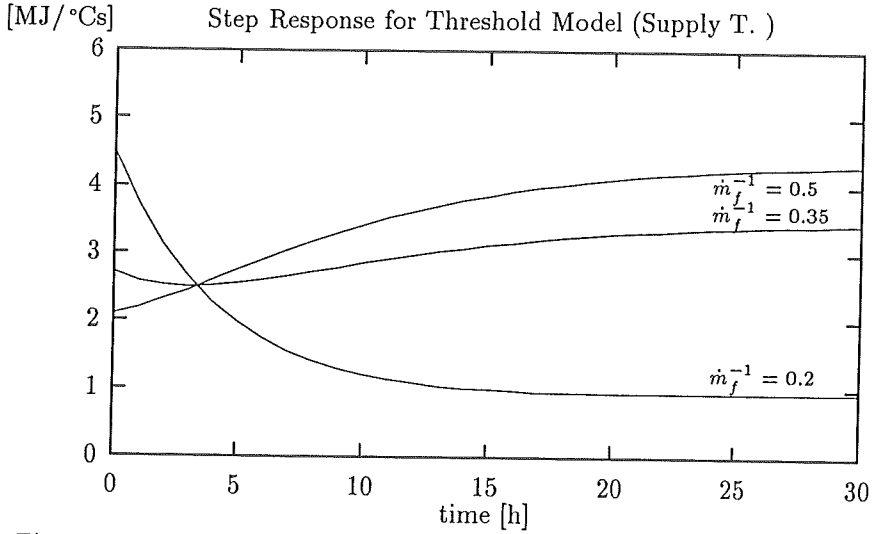


Figure 5.16: Step response to a 1°C increase of supply temperature in smooth threshold model for different flow.

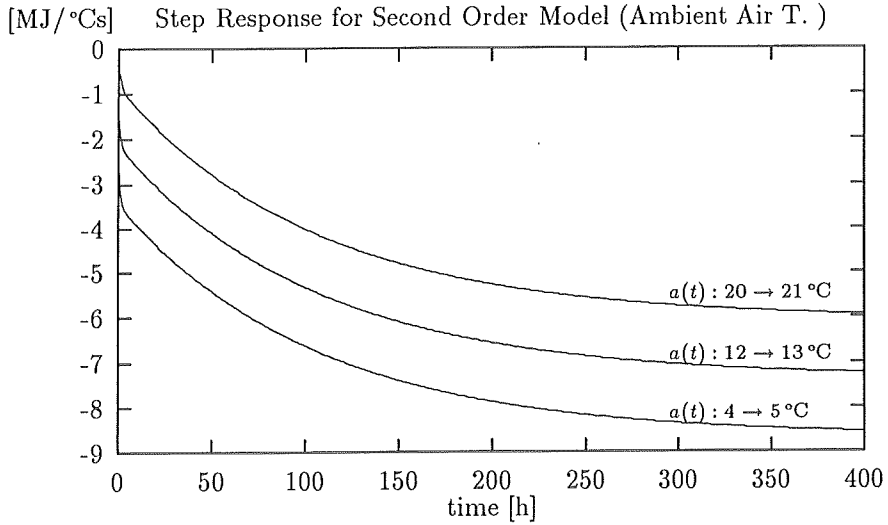


Figure 5.17: Step response to a 1 °C increase ambient air temperature in second order model for different ambient air temperature levels.

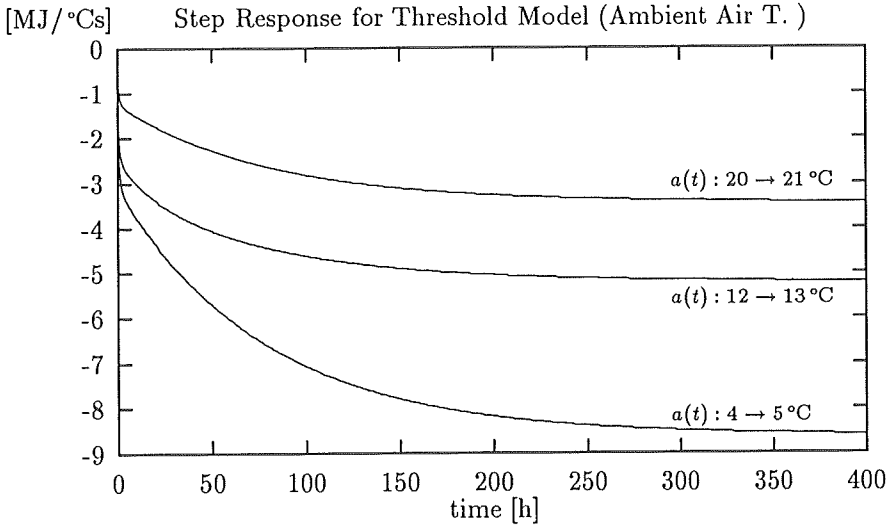


Figure 5.18: Step response to a 1 °C increase of ambient air temperature in smooth threshold model for different ambient air temperature levels.

The analysis of this section shows that both of the considered extensions of the linear transfer function model result in qualities of the model estimate, which are reasonable and principally in agreement with actual physically derived relations. The unrealistic properties of the model estimates can all be traced back to lacking excitation of data.

5.6 MODELLING USING NEURAL NETWORK

In this section artificial neural networks are considered as a class of model structures holding the capability of describing the existing nonlinear relations in the heat load dependence on supply temperature and ambient air temperature. It has been proved that multi-layer networks hold the capability of approximating any continuous function (Hornik, Stinchcombe, & White, 1989).

This investigation considers the application of one particular choice of neural network in gaining a sufficient nonlinear predictor of the heat demand. This class of neural networks includes one, which previously has been successfully applied to, e.g. the sunspot series, see Weigend, Huberman, & Rummelhart (1990).

In the following the neural network model, called the *perceptron*, used in this section is stated using the terms of the neural network literature, see Másson & Wang (1990).

ARCHITECTURE

The general structure of the network is shown on Figure 5.19, where the neurons are shown as circular nodes. In this investigation only models with two layers of neurons are considered. The motivation for the decision of leaving out the models with only one layer of neurons is an intuitive belief that a two-layer model has a higher degree of flexibility in modelling

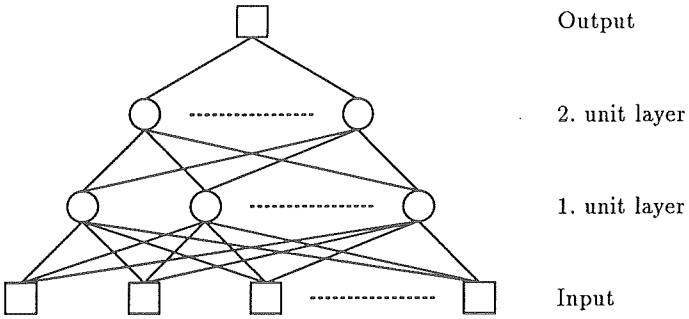


Figure 5.19: Architecture.

nonlinearities than a one-layer model, for the same number of parameters.

The network model has n_I input, $u_i(t), i \in [1; n_I]$, and n_1 and n_2 neurons in the two intermediate layers, respectively. The network has one output corresponding to the dependent variable, the variations of which is the object of the modelling.

In the general structure each input is connected to each neuron in the first layer, each neuron in the first layer is connected to each neuron in the second layer and each neuron in the second layer is connected to the output. Each connection has an adjustable weight, $w_{..}$. The output and each neuron has an adjustable bias, $b_{..}$.

The output of neuron j in the first layer is

$$o_j^{(1)}(t) = g\left(\sum_{i=1}^{n_I} w_{ij}^{(1)} u_i(t) - b_j^{(1)}\right), \qquad j \in [1; n_1], \tag{5.49}$$

where $g(\cdot)$ is the activation rule of the neuron, see below. The output of

neuron k in the second layer is

$$o_k^{(2)}(t) = g\left(\sum_{j=1}^{n_1} w_{jk}^{(2)} o_j^{(1)}(t) - b_k^{(2)}\right), \qquad k \in [1; n_2], \tag{5.50}$$

and the output of the network is

$$o(t) = \sum_{k=1}^{n_2} w_k^{(o)} o_k^{(2)}(t) - b^{(o)}. \tag{5.51}$$

The total number of parameters (weights and biases) is $N_p = (n_I + 1)n_1 + (n_1 + 2)n_2 + 1$.

Using a neural network as a model describing functional relations is in character a nonparametric approach (Ripley, 1992). The reason is that a neural network can be applied without using any physical knowledge about the system. The neural network can simply be applied as a nonlinear approximation to the existing system relations in the same way as a traditional nonparametric method can be applied. Furthermore, it is impossible to give any physical interpretation of the individual weights and biases of a neural network model.

ACTIVATION RULE

The nonlinearity of the neural network is introduced in the activation rule of the neurons. The activation rule of the neuron is the nonlinear transformation from input to neuron output. In this investigation the sigmoid, performing a smooth mapping $(-\infty; \infty) \rightarrow (0; 1)$, given by

$$g_S(x) = \frac{1}{1 + \exp(-ax)}, \tag{5.52}$$

is used. This is in fact the logistic function shown on Figure 5.10 with $a = 1$. Since, as noted by (Weigend et al., 1990), the slope a of the sigmoid can be absorbed into weights and biases without loss of generality, the sigmoid is used with $a = 1$.

DYNAMICS OF COMPUTATION

The dynamics or order of computation in the network carried out for each observation is: Apply the input values to obtain the output of the first layer neurons. Then use these as input to the second layer neurons and use the output of these in the bias corrected linear combination to give the network output. This type of network is referred to as a *feed forward* network in the perceptron class.

If the identity ($g(x) = x$) is used as activation rule, the neural network degenerates to a linear filter. Thus, the network will be able to perform as a predictor obtained from an ARX-model, since for an ARX-model the one-step predictor is linear in both model parameters and regressor variables. The neural network will, however, contain superfluous parameters due to the intermediate layers rendering redundant in a linear filter.

Determining the weights and biases in a neural network is, in the neural network terminology, referred to as *training* of the net. Using a data set, which represents the desired performance of the net, the weights and biases of the net are adjusted in the training process to obtain a net performance matching that represented by the training set. The process of determining the weights and biases is referred to as parameter estimation, the weights and biases being the parameters of the model.

It is assumed that the dependent variable can be described by the following model

$$y(t) = o(u(t)) + e(t). \tag{5.53}$$

The neural network output, $o(u(t))$, is written, as it actually is, being dependent solely on the input vector $u(t)$. The neural network is assumed to give a sufficient description of the system behavior with the vector $u(t)$ applied as input to the net, and, thus, $e(t)$ is the innovation allowing for discrepancy between network output and dependent variable observation. Hence, $\{e(t)\}$ is , as previously, assumed to be white noise with variance σ_e^2.

Denoting the parameters of the model θ, the prediction of the network is given as

$$\hat{y}_\theta(t) = o_\theta(u(t)). \tag{5.54}$$

The estimate of the neural network parameters, minimizing the prediction error variance, is thus found by minimization of the criterion function in (5.5).

The objective of this investigation is to obtain a modelling of the nonlinearities in the variations of the heat load. In the section on general transfer function modelling the best performing model, rewritten to match the formulation in (5.53), would have the form

$$\begin{aligned} H_1(q^{-1})\nabla_{24}p(t) &= H_2(q^{-1})\nabla_{24}s(t) + H_3(q^{-1})\nabla_{24}a(t) \\ &+H_1(q^{-1})\nabla_{24}\nu_1(t) + H_1(q^{-1})\nabla_{24}\left(\nu_2(t)I_{\{\mathrm{A}\}}(t)\right) + e(t). \end{aligned} \tag{5.55}$$

$\nu_1(t)$ and $\nu_2(t)$ are trigonometric diurnal profiles, the first one actually being superfluous due to the diurnal differencing.

Each of the transfer functions H_i are infinite impulse response filters. This means that theoretically an infinite number of inputs and parameters would be necessary in a parameter linear formulation, if each one of the weights of the impulse responses were to be estimated as one parameter. Likewise, presumably an infinite number of inputs would be necessary in a neural network formulation. However, each of the transfer functions has been formed as rational functions, i.e. they can be written

$$H_i(q^{-1}) = \frac{B_i(q^{-1})}{A_i(q^{-1})} \tag{5.56}$$

with A and B being finite polynomials.

This means that multiplying with $\prod_{i=1}^{3} A_i(q^{-1})$ makes the dependence on lagged dependent and external variables finite, but hereby the model error end up being the non-white MA process $\prod_{i=1}^{3} A_i(q^{-1})e(t)$. Hence, using the transfer function model structure as a guideline for choice of neural network structure, either a large number of inputs or feedback of the model error is called for.

The choice in this investigation is to make a rude cut-off of the lagged dependent and external variables without using feedback of model error. The reason is:

1. Increasing the number of inputs by one increases the number parameters by n_1, wherefore it is of crucial importance to keep the number of inputs low so the model will be a suitable generalization of the system without fitting special features of the training set. Over-parameterization should be avoided.
2. Using feedback of model error will increase the deviation of the criterion from being quadratic in the parameters, implying additional complications in the numerical minimization.

The chosen number of inputs is 15. With the dependent variable being $p(t)$ the inputs are

$$\left\{ \begin{array}{lll} p(t-1) & s(t-24) & 1-I_{\{\mathrm{A}\}}(t) \\ p(t-2) & a(t) & \sin(2\pi t/24) \\ p(t-24) & a(t-1) & \cos(2\pi t/24) \\ s(t) & a(t-24) & \sin(2\pi t/12) \\ s(t-1) & I_{\{\mathrm{A}\}}(t) & \cos(2\pi t/12) \end{array} \right\}$$

For this selection of inputs, estimation were carried for $n_1 \in [2;8]$ and $n_2 \in [2;3]$. The resulting values of $SS(1009;3843)$ and BIC are shown in Table 5.6.

In the estimation of the neural network models, the model estimate was continuously applied on the validation set (Data B) to test its prediction performance. The reason for doing this that in the field of training neural network models, it is recommended to apply the model, in the course of the estimation, on a test set, and stop the training when the performance on the set deteriorates. The idea of this approach is that the neural network structure should be so wide that it for a certain set of weights and biases reaches the optimal prediction performance. For neural networks there are no way of determining an optimal net structure. Consequently, to reach the optimal performance the net structure has to be too wide implying

$\hat{\sigma}_e^2$ / BIC							
	n_1						
n_2	2	3	4	5	6	7	8
2	5.88	4.56	4.19	3.92	3.93	3.97	3.90
	5307	4713	4597	4532	4665	4801	4890
3	5.91	4.62	4.24	3.87			
	5348	4784	4672	4547			

Table 5.6: Results for different numbers of neurons in the two layers.

that it also holds the capability of fitting to special characteristics in the estimation set. To avoid this to happen the training/estimation of the model ought to stopped when the performance on the test set deteriorates. There is, of course, no guarantee that there does not exist a set of weights and biases for which a better performance on both the estimation set and the test set is obtained. It actually depends on the particular trajectories, which the parameter follows in the progress of the estimation. This, in turn, means that the choice of initial parameter values may be determining for the final model estimate.

The fact that the test set is used for determining when to stop the minimization implies that there is a crucial dependence on the chosen validation set. The validation set must be carefully chosen to contain the general relations, being the object for the model formulation. Furthermore, the variation ranges of the processes entering the model (dependent and explanatory variables) must be sufficiently wide for the eventual use of the model.

Figures 5.20-5.21 show the course of prediction error standard deviation in the progress of the estimation for both the estimation set and the validation

Period	SS	$SS^{\%}$
1009-3843	1.95	2.02
1009-1953	1.37	2.48
1954-2898	1.82	1.98
2899-3843	2.48	1.47
$\hat{\sigma}_e^2 = 3.92$	$BIC = 4532$	

Table 5.7: Performance of a model with $n_1 = 5$ and $n_2 = 2$.

set. For $(n_1, n_2) = (2, 2)$ it is seen that the estimated standard deviation of the prediction error for the set does not increase. According to the discussion above this means that the structure of the neural network model is not too wide; it is most likely too restrictive. However, for $(n_1, n_2) = (8, 2)$ the prediction error for the test set shows a minimum around iteration no. 470 after which it slightly increases. This is a clear indication that the model is over-parameterized.

Table 5.6 shows that the fit of the neural network model clearly is improved when the number of units in the first layer is increased until $n_1 = 5$, but for higher values of n_1 the variance estimate seems to remain on a constant level and BIC increases. It obviously does not pay off to increase n_2 from 2 to 3. Thus, the lowest value of BIC were found for $(n_1, n_2) = (5; 2)$. For this configuration the comparison with the preceding results can be made from Table 5.7. A considerable improvement is found for all three periods.

Comparing the residuals, shown on Figure 5.22, with those from the other models it is seen that the majority of the residuals are reduced in size. However, a number of the residuals have increased. This indicates a difference in the modelling capabilities of the different approaches.

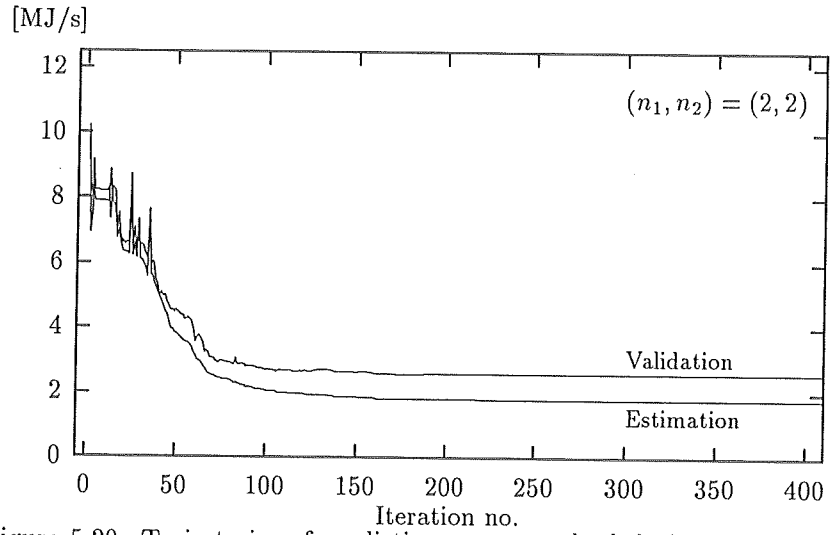


Figure 5.20: Trajectories of prediction error standard deviation for both estimation set and validation set in the course of estimation, $(n_1, n_2) = (2, 2)$.

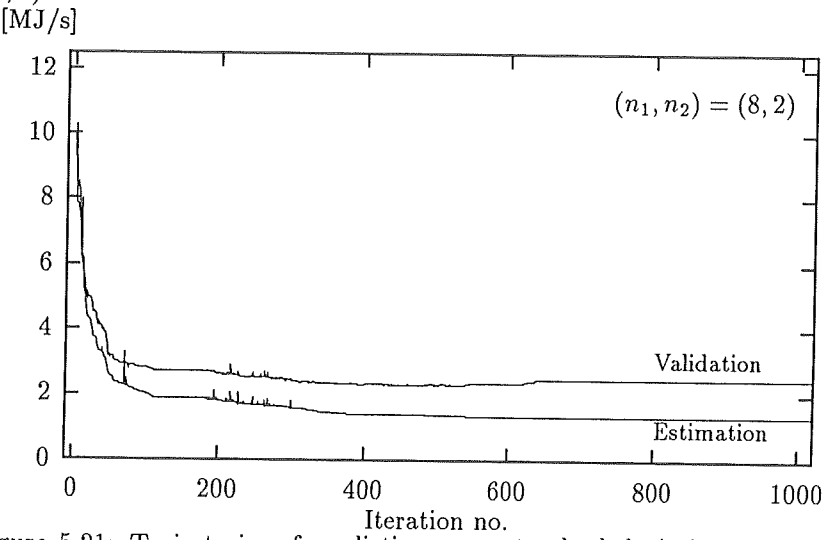


Figure 5.21: Trajectories of prediction error standard deviation for both estimation set and validation set in the course of estimation, $(n_1, n_2) = (8, 2)$.

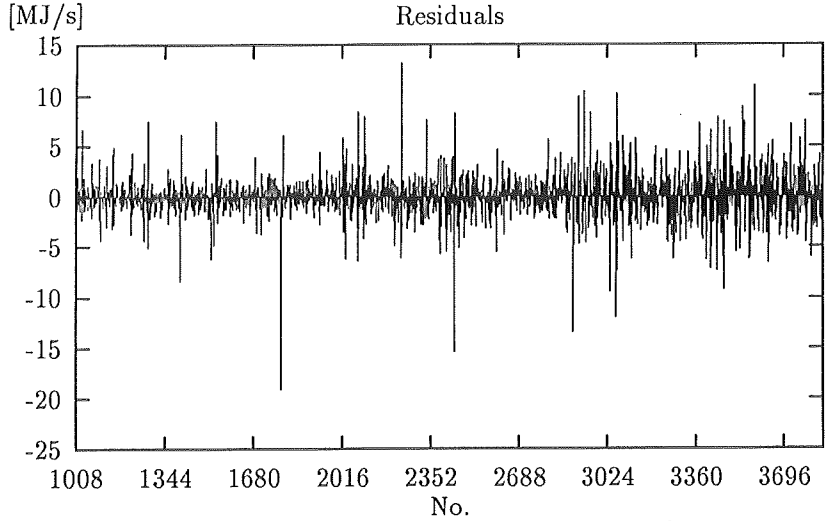


Figure 5.22: Residuals for the model with $n_1 = 5$ and $n_2 = 2$.

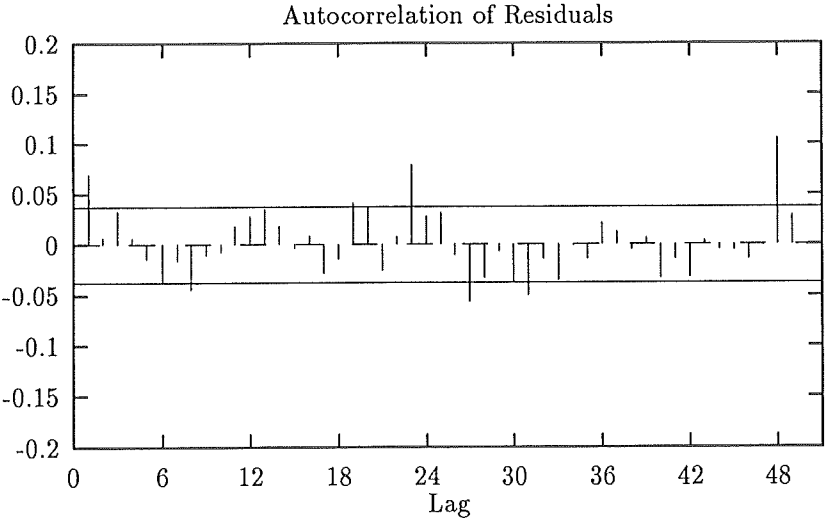


Figure 5.23: Autocorrelation of residuals for the model with $n_1 = 5$ and $n_2 = 2$.

The largest of the residuals recur for all model estimates hinting the possibility of errors in the observations. Examining the observations and the residuals in detail reveals that the large modelling errors most likely are caused by extremes in the load variation. These extremes in turn are the consequence of extraordinary operating points or large changes in supply variables, which, due to the complexity of the system, cause even the considered nonlinear models to be insufficient. It is a question whether the observations in such extreme situations should be considered as erroneous (outliers). However, seeing that the large residuals are not reduced considerably when either the nonlinear models or the neural network models are applied it seems to be appropriate to apply a robust estimation method. This is a matter for further investigation.

The autocorrelation of the residuals in Figure 5.23 shows that the correlation is not fully described. This calls for the employment of models with feed back of prediction error, viz. a recurrent network model, to obtain a description of this correlation.

5.7 MODEL VALIDATION

Table 5.8 summarizes the modelling results and shows the standard deviation of the prediction errors when the estimated models were applied on Data B. The important result is that, although the neural network model is superior in fitting Data A, the second order model and the smooth threshold model are better when tested on Data B. The linear transfer function model is still inferior to the others, but the difference from the neural network is not so evident. Furthermore, the threshold model is performing worse than the second order model even though it seems to give a better and more reasonable description of Data A.

Figure 5.24 gives a visual illustration of the contents of Table 5.8. In addition is shown the result for increasing number of neurons in the first layer. This indicates that, different from the recommendation of the values of BIC, $(n_1, n_2) = (4, 2)$ is a better choice of structure. Anyhow, the

Model	Estimation (Data A)		Validation (Data B)	# Parameters
	$\hat{\sigma}_e^2$	BIC	$\tilde{\sigma}_e^2$	
Linear	5.80	5104	11.29	15
Second Order	4.94	4695	9.59	23
Smooth Threshold	4.89	4672	9.89	25
Neural Network	3.92	4532	10.99	95

Table 5.8: Results from estimation and validation.

difference is small, and it can be concluded that the best attainable performance of a neural network with two layers of neurons, without feedback of prediction errors and with the chosen input pattern is reached.

The explanation to the clear difference between the neural network performance for the estimation set and the validation set is the large number of degrees of freedom in the model. This large flexibility allows the neural network to fit to special to characteristics of the estimation set, and so its capability of generalization (fitting to other data sets) is reduced. One way of avoiding this is by continuously testing the network on a different data set when it is estimated, and then stop the estimation when its performance on the test set deteriorates. Another proposal is to remove some of the connections in the network to obtain an appropriate modelling flexibility. For instance, the covariance estimate of the parameter estimates in (5.24) could be used to propose which parameters should be set to zero. This is, however, not done in the present investigation.

For the smooth threshold model it is hardly reasonable to postulate that it contains too many parameters to be a load model being valid in all seasons of the year. The reason for the worse performance on Data B is

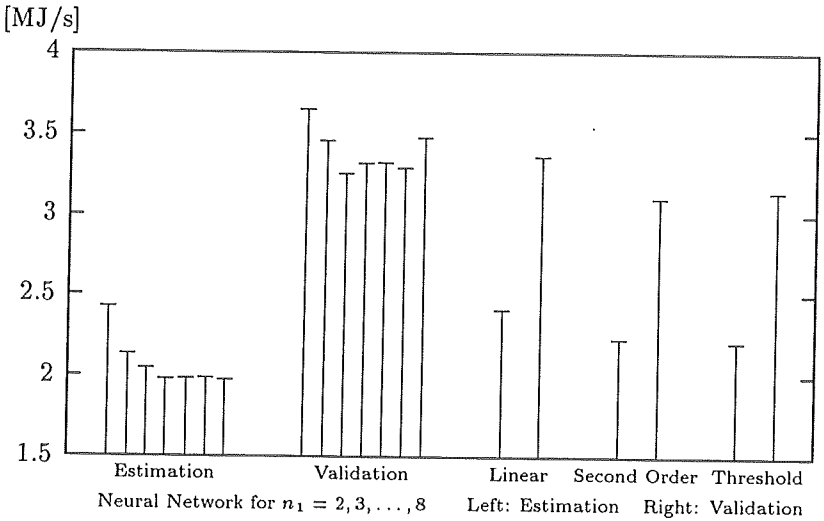


Figure 5.24: Bar diagram of $\hat{\sigma}_e$ for the estimation set and $\tilde{\sigma}_e$ for the validation set.

probably that Data A is too short a period with too little variation in data to represent the system under all operating conditions.

5.8 OPTIMAL PREDICTION

For the models considered in the preceding chapter, all of which can be written on the form

$$y(t) = g_\theta\left(y^{t-1}, u^t, x^t\right) + e(t) \tag{5.57}$$

predictions j step ahead, $\hat{y}(t+j \mid t)$, can be obtained as the expected value of $y(t+j)$ conditional on observations up to time t, i.e.,

$$\hat{y}(t+j \mid t) = E\left[y(t+j) \mid y^t, u^{t+j}, x^t\right]. \tag{5.58}$$

In this section the difference between control variables (u) and external

variables (x), which cannot be controlled, is emphasized for the reason that the way they are treated, when the model is used for prediction, is conceptually different. The difference is that the external variables x have to be predicted whereas the control variables can be determined in accordance with a control law making the dependent variable follow a desired course.

The predictor (5.58) has the quality of being minimum mean square error (Madsen, 1989). That is, when using the conditional expectation as predictor the expected prediction error is minimal in mean square sense.

It is necessary for this quality of the conditional mean to be valid generally (for all j) that the applied model actually is a true description of the system, at least in the space of variables in which the system is operated. When the model is an approximation to the system (possibly determined from data) the approximation giving the optimal j-step prediction is only optimal for that particular value of j.

Using the model explicitly the prediction can be found as

$$\hat{y}(t+j \mid t) = g_\theta\left(\hat{y}^{t+j-1|t}, u^{t+j}, \hat{x}^{t+j|t}\right) \tag{5.59}$$

with

$$\hat{y}^{t+j-1|t} \equiv (\hat{y}(t+j-1 \mid t), \hat{y}(t+j-2 \mid t), \ldots) \tag{5.60}$$

and

$$\hat{x}^{t+j|t} \equiv (\hat{x}_1(t+j \mid t), \hat{x}_1(t+j-1 \mid t), \ldots, \hat{x}_{n_x}(t+j \mid t), \ldots)\,. \tag{5.61}$$

n_x is the number of non-susceptible external variables.

To use the model for j-step predictions the calculation of the preceding predictions $\hat{y}(t+l \mid t)$, $l = 1, \ldots, j-1$ is required. Correspondingly, the predictions of the external variables x_i, necessary for the calculation of the sequence of predictions of the dependent variable, have to be available. These can, for instance, be found by separate prediction methods. Furthermore, the prediction of y is calculated as a function of chosen future values

of the control variables. That is, this approach is based on the future control variables being independent values. When the control is determined by feedback from the dependent variables y the control rule, describing the relation between $u(t)$ and previous values of the dependent variable and other external variables, this relation is introduced into the predictor to the prediction at time $t + j$ conditioned on observations of the dependent variable and all the control variables up to time t.

The prediction by evaluation of expected values is straightforward for the linear transfer function models. The evaluation of the expected values for the nonlinear models is more demanding as it requires knowledge about the higher order moments of the stochastic variables, which the prediction is conditioned on. For instance, a prediction based on evaluation of expected values for predictions of explanatory variables in the second order nonlinear model requires that the means and variances of these predictions are known.

5.9 EXPLICIT PREDICTORS

The implication of a model being an approximation instead of a true and complete description is significant. Due to the model being an approximation to the system the parameter values in the approximation can be determined to be optimal with respect to one particular purpose/criterion. For instance, when the model is applied for prediction j steps ahead the parameters will have to be particular to j to be optimal. Furthermore, the optimal parameters will depend on the operation point of the system. The consequence of this is that the parameters will have to be determined so the performance of the approximation is optimal with respect to the application of the approximation.

In the Chapter 6 different methods, for which the approximation of a model is allowed to adapt to the application of the model, i.e. prediction horizon j and operation point, is outlined and tested. This matter is treated in Holst (1977).

5.10 ATYPICAL DIURNAL TYPES

One important issue in modelling heat load is the handling of differences coming alone from differences in diurnal types. The sources of the diurnal differences are varying habitual hot tap-water consumption as well as varying control of heating.

First of all there are likely to be load differences among the days of the week due to differing tap-water consumption and heating in public services, offices and industry. This should be recognized as a typical weekly pattern in the heat load variation. Apart from the typical days there also exist days which are diverging with respect to the pattern of the typical diurnal variation. Holidays, non-working days and non-school days are examples of days being aberrant from the otherwise typical diurnal pattern for a given day in week. Such days can be predicted in advance. However, also unpredictable atypical days also occur, e.g., strike days.

Previously in this chapter dynamic heat load models have been estimated, where the only discrimination is between usual work days and days in the weekend (no atypical days were present in data). It does not seem to be feasible to extend either of these models to be able to describe the heat load on all types of atypical days. This would require too large a number of parameters in the models.

When a model, that does not take into account the particular diurnal pattern of a given atypical day, is estimated and used for prediction, precautions must be taken in three respects.

1. In the estimation of the model, observations from atypical days should not be used.

2. When predicting on an atypical day, the prediction should be accompanied by a modification, which accounts for the difference between the load had the day not been atypical and the load for the given type of atypical day. For instance, for Maundy Thursday the discrepancy

between model predictions and actual observations for previous years can be used as an estimate of the modification of the model prediction for the current year, possibly corrected for weather influence.

3. For prediction on days, where observations from atypical days enter the predictor in the regressor space, the actual observation should not be used. Instead observations from the same day, had it not been aberrant, should enter the predictor. For instance, model predictions could be used instead of the actual observations.

Holst & Ekelund (1987) propose a formalized method that implement these precautions. Ericsson & Haglund (1991) use a fully connected perceptron network with 64 input and 35 neurons in one hidden layer for a dedicated modelling of the load on atypical days.

5.11 CONCLUSION

In this chapter a linear transfer function model, two nonlinear extensions of this model and a two neuron layer feed forward neural network model have been estimated on load data from the DH system in Esbjerg.

When estimating the linear model on disjoint intervals of data it becomes obvious that the linear model is insufficient in describing the dynamic dependence on supply temperature and ambient air temperature. All the nonlinear models, including the neural network model, gives a significantly better fit to data, and when testing the models on a validation set this still is the case.

For the nonlinear extensions to the transfer function model the problem is how to obtain the physically true and best description. In the models proposed in this chapter the transfer function from supply temperature has been extended by introducing a dependence on a low-pass filtered inverse mass flow. Likewise the transfer function from ambient air temperature is made dependent on a low-pass filtered ambient air temperature. These ex-

tensions are both giving an improved description. In general, the nonlinear relations found seem reasonable from a physical point of view. However, probably because of lacking excitation in data, the dependence on inverse flow seems to be inconsistent with the expected physical dependence for summer situations with low mass flow.

The estimated neural network model with five and two units in the neuron layers respectively is clearly superior to the nonlinear extensions of the transfer function models in fitting data. The reason for this is first of all the large flexibility of the model due to the large number of parameters (95). Moreover, the smooth nonlinear transformation of the neurons with a lower and an upper saturation level seems to be consistent with physical relations existing in DH systems. Finally, the fully connected neural network model allows for modelling of any interdependence between different inputs.

The problem of paramount importance in neural network modelling is how to determine the appropriate structure and connecting of inputs and neurons. When the neural network model, which was clearly superior in fitting the estimation set, was tested on the validation set the result was poor compared to the nonlinear transfer function models. This hints that the flexibility of the neural network model is too wide allowing the model to fit to special features of the estimation set, which ought not be contained in a generalized representative description of the system.

Therefore the conclusion in this chapter must be that the nonlinear extensions of the transfer function models are the most appropriate for modelling the load.

.....

The estimation results in this chapter were obtained using a numerical minimization routine developed by Melgaard & Madsen (1991). The software for implementation of models were programmed in Fortran, and the calculations were carried out in double precision arithmetic on a HP-9000/750.

.....

6

Adaptive Methods for Prediction of Heat Load

In the preceding chapter different dynamic models for the load in DH systems describing the dependence on the ambient air temperature, supply temperature and time and type of day have been investigated. The purpose was to attain models capable of giving a sufficient description of the dynamic relations. This chapter is devoted to the issue of predicting the heat load using adaptive methods for parameter or state estimation.

Adaptive estimation of linear models has in the recent years been one of the preferred practices for heat load prediction in DH systems. Sejling (1987) introduced a linear model with trigonometric profiles and transfer functions from supply temperature, ambient air temperature and wind velocity. j-step predictor formulations were derived and estimated by the use of simple exponential forgetting. This prediction approach is now implemented in a computer program running at a number DH systems in Denmark; see Madsen, Palsson, Sejling, & Søgaard (1992b) for a description of the program and the results obtained by using it.

Wiklund (1989) applies a similar approach on load data in Umeå, Sweden, and compares with corresponding off-line estimates. He also investigates the state space formulation of exponential smoothing of diurnal and weekly profiles updated by Kalman filter recursions. This technique is also tested in the present chapter.

6.1 ADAPTIVE ESTIMATION OF LINEAR MODELS

RLS estimation requires that the model can be written in a formulation being linear in the model parameters, viz.

$$y(t) = \boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{\theta} + e(t). \tag{6.1}$$

$\boldsymbol{x}(t)$ is a vector of regressors, and $\boldsymbol{\theta}$ are the parameters of the model. $e(t)$ is the white noise innovation at time t.

One version of the RLS algorithm with exponential forgetting for estimation of the parameter vector $\boldsymbol{\theta}$ is (Ljung & Söderström, 1983)

RLS .....

$$\varepsilon_{\hat{\theta}(t-1)}(t) = y(t) - \boldsymbol{x}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1) \tag{6.2}$$

$$\boldsymbol{P}_\lambda(t) = \frac{1}{\lambda}\left(\boldsymbol{P}_\lambda(t-1) - \frac{\boldsymbol{P}_\lambda(t-1)\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{P}_\lambda(t-1)}{\lambda + \boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{P}_\lambda(t-1)\boldsymbol{x}(t)}\right). \tag{6.3}$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \boldsymbol{P}_\lambda(t)\boldsymbol{x}(t)\varepsilon_{\hat{\theta}(t-1)}(t) \tag{6.4}$$

.....

$\boldsymbol{P}(t)$ is a scaled covariance matrix, and λ is the weighting factor making the estimation adaptive.

Parkum (1992) and Parkum, Poulsen, & Holst (1992) propose an extension to the simple exponential forgetting algorithm consisting of the use of a vector of forgetting factors with the same dimension as the parameter space. The idea of this algorithm is to make an eigenvector/-value determination of the covariance matrix in each step, then measure the amount of information in the regressor in each eigenvector direction and subsequently apply a forgetting factor in each eigenvector direction that keep the accumulated information in the covariance matrix on a constant level. Application to simulated data shows promising results.

An important issue in estimation is the situation where the complexity of the model is restricted compared to the system. The estimation of a model is traditionally carried out assuming that the model, with some true parameters, is capable of giving a correct system description, and that discrepancies between observations and model predictions are due to additive white noise. One consequence of this is that, for a sufficiently large number of observations used in the estimation, the uncertainty of the estimates will be negligible. However, usually models are approximations to physically complicated systems, which implies that the model, no matter what values the parameters have, will give an inaccurate and insufficient description of the system dynamics. This deficiency will not vanish for an increasing number of observations as will the uncertainty arising from the assumed noise corruption of the observations.

It is certainly the case that a model describing the load in a DH system will be approximative. This means that the parameter estimates will be influenced by the fact that part of the dynamic relations are left unmodelled, and consequently the estimates will be influenced by this restriction in model complexity.

Henningsen & Sejling (1990) propose a method for quantification of uncertainty in recursive estimation of models with restricted complexity. The method originates from a Bayesian approach to quantification of the model uncertainty, see Goodwin & Salgado (1989) and Salgado, de Souza, & Goodwin (1990). For a dynamic system having input $u(t)$ and output $y(t)$ it is assumed that the model describing the relation between the input and output can be written (see also Figure 6.1)

$$y(t) = \{G_0(\theta, q^{-1}) + G_\Delta(q^{-1})\}u(t) + e(t), \tag{6.5}$$

where

- $G_0(\theta, q^{-1})$ is a nominal transfer function of known form and parameterized by the vector θ.
- $G_\Delta(q^{-1})$ is a representation of the unmodelled dynamics.
- $e(t)$ is a noise component accounting for the measurement error.

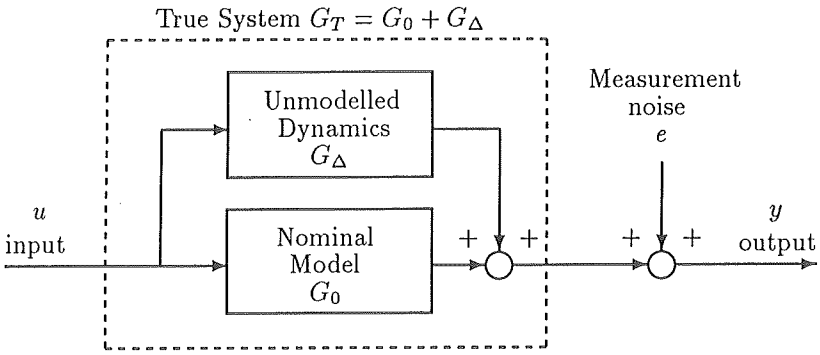


Figure 6.1: System description including additive unmodelled dynamics.

It is assumed that the nominal model and the unmodelled dynamics can be described by their impulse responses. Based on further assumptions of a-priori independent normal distribution of (1) the nominal parameters, (2) the impulse response representation of the unmodelled dynamics and (3) the noise component, standard calculations of conditional mean and variance lead to the a-posteriori distribution of θ as well as of the impulse response of the unmodelled dynamics. As it assumed that a complete description is obtained, the a-posteriori estimate can be used for evaluating the contributions of the unmodelled dynamics to the uncertainty of the nominal model estimate.

Subject to simplifying assumptions - justifiable when the unmodelled dynamics is of minor influence compared with the nominal model - the Bayesian estimation approach can be turned into a scheme, for which the nominal parameter estimates coincide with the LS estimates. However, the covariance for the parameter estimates still have contributions from the unmodelled dynamics. See Henningsen & Sejling (1990) for further details.

This estimation scheme is easily turned into a recursive algorithm, and based on assumptions of a first order model for the unmodelled dynamics the covariance for its impulse response can be evaluated. Hereby the need for prior covariance stipulation is removed. This is valuable since the

prior covariance is of significant influence for the resulting uncertainty assessment, which is quantified by the impulse response of the unmodelled dynamics.

Another point that can have a striking influence on LS estimates is the presence of outliers in data. In particular for outliers being additive effects LS estimates can be subject to severe bias. In Chapter 7 algorithms for recursive robust estimation of AR parameters for settings, where outliers are likely to occur, are derived.

In the following two sections RLS with exponential forgetting is applied for recursive and adaptive estimation of parameter-linear models. In Section 6.1.1 linear reformulations of the dynamic models of Chapter 5 is estimated. Subsequently in Section 6.1.2 a number of j-step predictors of different structure are applied and compared for $j = 1, 3, 6, 9, 12, 15, 18, 21, 24$. In all cases the estimation results are presented for the value of the forgetting factor λ giving the lowest prediction error variance.

Assessment of modelling and prediction performance is obtained by using the measures (5.25) and (5.26). For the modelling study these measures are calculated for the same intervals as used in Chapter 5, whereas the predictors are compared on two intervals, viz. 505-2172 and 2173-3843.

6.1.1 RLS ESTIMATION OF LINEAR MODELS WITH TRIGONOMETRIC PROFILES

An ardent question is how RLS estimation of a linear model using exponential forgetting compares with the modelling results in the Chapter 5. To apply the RLS algorithm on the model in (5.34) it has to be turned into a linear or pseudo-linear form. This can be obtained by multiplying with the nominator polynomiums in (5.34). The following linear model corresponds to such a reformulation of (5.34) in which, however, a number of the resulting parameters have been removed among these all the parameters in MA polynomium. Furthermore, the diurnal difference has been relaxed and

Period	SS	$SS^{\%}$
1009-3843	2.40	2.42
1009-1953	1.70	2.94
1954-2898	2.26	2.44
2899-3843	3.06	1.74

Table 6.1: Performance of model (6.6).

interchanged by diurnal parameters, and consequently two diurnal profiles is necessary. l is a level parameter introduced since some variables have mean values different from zero.

$$\begin{aligned} p(t) &= (\alpha_1 q^{-1} + \alpha_2 q^{-2} + \alpha_{24} q^{-24} + \alpha_{25} q^{-25} + \alpha_{26} q^{-26}) p(t) \\ &+ (\beta_{1,0} + \beta_{1,1} q^{-1} + \beta_{1,2} q^{-2} + \beta_{1,24} q^{-24} + \beta_{1,25} q^{-25} + \beta_{1,26} q^{-26}) s(t) \\ &+ (\beta_{2,0} + \beta_{2,1} q^{-1} + \beta_{2,24} q^{-24} + \beta_{2,25} q^{-25}) a(t) \\ &+ \mu_1 \sin\left(\frac{2\pi t}{24}\right) + \mu_2 \cos\left(\frac{2\pi t}{24}\right) + \mu_3 \sin\left(\frac{2\pi t}{12}\right) + \mu_4 \cos\left(\frac{2\pi t}{12}\right) \\ &+ \left(\mu_5 \sin\left(\frac{2\pi t}{24}\right) + \mu_6 \cos\left(\frac{2\pi t}{24}\right) + \mu_7 \sin\left(\frac{2\pi t}{12}\right) + \mu_8 \cos\left(\frac{2\pi t}{12}\right)\right) I_{\{\mathrm{A}\}}(t) \\ &+ l + e(t) \end{aligned} \tag{6.6}$$

This model is the result of an investigation, where different different model structures were estimated. For instance, it turned out that estimation of a moving average polynomial on the residuals did not improve the fit. The results of RLS estimation with the forgetting factor giving the best performance $\lambda = 0.997$ are listed in Table 6.1.

Comparing with the results in Table 5.2 for the linear transfer function model, it is seen that the adaptive estimation gives an improved fit for the first third of the criterion period, whereas for the later periods the fit of the off-line estimated linear transfer function model is better. The explanation to this result is probably that the transfer function model (5.34) is able

to give a better dynamic description, but the model estimate is adapted to the later part of data due to the excitation occurring in this period. The adaptive estimation of the linear model has implied a higher degree of adaption to the first period than the off-line estimate, for which reason a better fit is seen. Another explanation is the higher number of parameters in the linear model. Comparing with the results for the nonlinear models in Chapter 5 it is seen that all of these are superior to RLS estimation with exponential forgetting of the model (6.6).

Figure 6.2 shows the residuals from the recursive estimation, and in Figure 6.3 the autocorrelation of the residuals is shown. The residuals have a tendency of being smaller for the first part of data than for off-line estimation of the linear transfer function model (Figure 5.3), but the later part the two sets of residuals show a greater similarity. The autocorrelation in Figure 6.3 show that correlation remains in the residuals, especially at lags 23 and 48 is the correlation significant.

An interesting point is how the stationary gain of the transfer functions from supply temperature and ambient air temperature have evolved during the recursive estimation. Figure 6.4 shows the evolution of the stationary gain from supply temperature to heat load calculated from the varying parameter estimates as

$$H^s(1,t) = \frac{\hat{\beta}_{1,0}(t) + \hat{\beta}_{1,1}(t) + \hat{\beta}_{1,2}(t) + \hat{\beta}_{1,24}(t) + \hat{\beta}_{1,25}(t) + \hat{\beta}_{1,26}(t)}{1 - \hat{\alpha}_1(t) - \hat{\alpha}_2(t) - \hat{\alpha}_{24}(t) - \hat{\alpha}_{25}(t) - \hat{\alpha}_{26}(t)}. \quad (6.7)$$

The gain starts off at a level close to zero. This corresponds to the heat load being almost independent of supply temperature. However, the supply temperature is almost constant until observation no. 2016 meaning that there is almost no information in data about the stationary load dependence on supply temperature. Around observation no. 1510 the gain changes abruptly to a level around 6 MJ/°Cs. This point coincide with an observed drop in both supply temperature and heat load, and the abrupt level change occur here as this is the only point where there are any substantial excitation. This is in harmony with the discussion in Chapter 5 on the same property of the off-line estimates, and it is a situation where

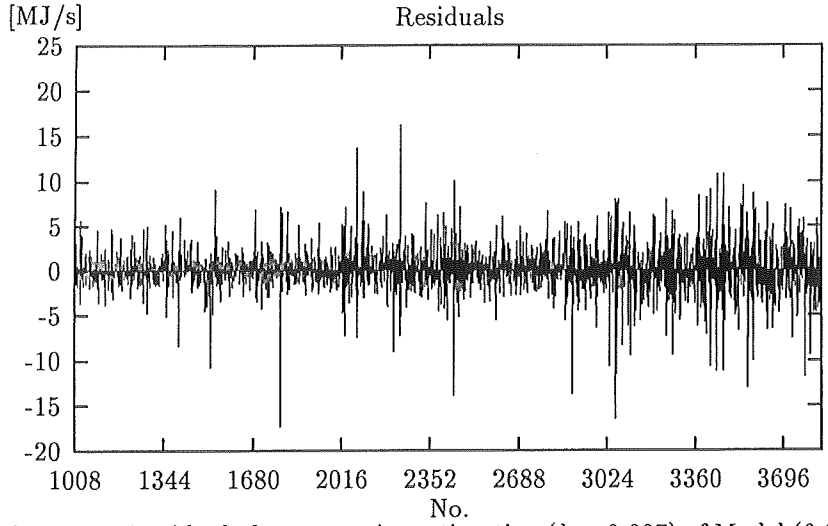


Figure 6.2: Residuals from recursive estimation ($\lambda = 0.997$) of Model (6.6).

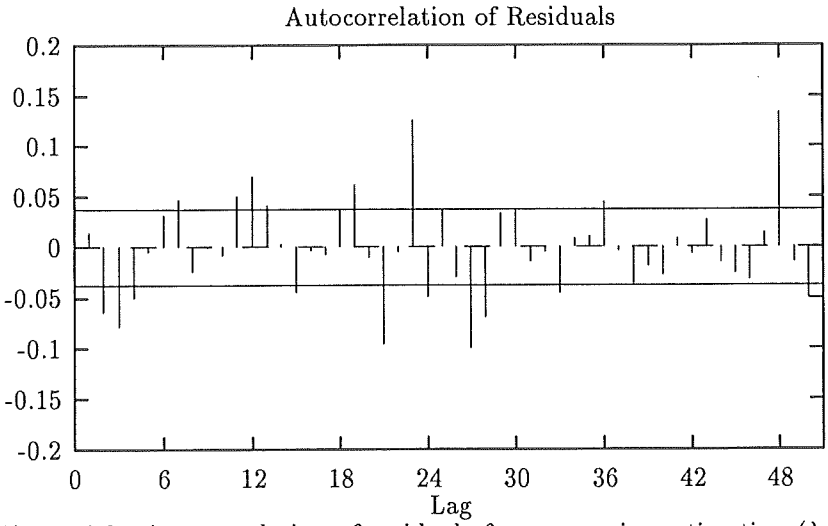


Figure 6.3: Autocorrelation of residuals from recursive estimation ($\lambda = 0.997$) of Model (6.6).

a selective forgetting factor algorithm could be advantageous. After observation no. 2016 the gain shows a behaviour that reflects the increased variation in supply temperature, and in the latest part of data the gain approaches a level about zero.

The stationary gain from ambient air temperature to heat load, calculated in the way as for the supply temperature, is close to -2 MJ/°Cs until observation no. 2016. Thereafter the gain shows a decrease ending at a level between -8 to -10 MJ/°Cs. The point, where the gain starts to change, probably coincides with the time when the ambient weather condition starts to influence room heating.

In order to introduce the same kind of nonlinear extensions as in Chapter 5, that is, modelling of the mass flow dependence of the transfer function from supply temperature as well as modelling of the dependence of the transfer function from ambient air temperature on the level of the ambient air temperature itself, the following model is put forward

$$\begin{aligned}
p(t) &= (\alpha_1 q^{-1} + \alpha_2 q^{-2} + \alpha_{24} q^{-24} + \alpha_{25} q^{-25} + \alpha_{26} q^{-26}) p(t) \\
&+ (\beta_{1,0} + \beta_{1,1} q^{-1} + \beta_{1,2} q^{-2} + \beta_{1,24} q^{-24} + \beta_{1,25} q^{-25} + \beta_{1,26} q^{-26}) s(t) \\
&+ (\beta_{2,0} + \beta_{2,1} q^{-1} + \beta_{2,24} q^{-24} + \beta_{2,25} q^{-25}) a(t) \\
&+ (\beta_{3,0} + \beta_{3,1} q^{-1} + \beta_{3,24} q^{-24} + \beta_{3,25} q^{-25}) \dot{m}_f^{-1}(t-1) s(t) \\
&+ (\beta_{4,0} + \beta_{4,1} q^{-1} + \beta_{4,24} q^{-24} + \beta_{4,25} q^{-25}) a_f(t) a(t) \\
&+ \mu_1 \sin\left(\frac{2\pi t}{24}\right) + \mu_2 \cos\left(\frac{2\pi t}{24}\right) + \mu_3 \sin\left(\frac{2\pi t}{12}\right) + \mu_4 \cos\left(\frac{2\pi t}{12}\right) \\
&+ \left(\mu_5 \sin\left(\frac{2\pi t}{24}\right) + \mu_6 \cos\left(\frac{2\pi t}{24}\right) + \mu_7 \sin\left(\frac{2\pi t}{12}\right) + \mu_8 \cos\left(\frac{2\pi t}{12}\right)\right) I_{\{A\}}(t) \\
&+ l + e(t),
\end{aligned} \tag{6.8}$$

where two additional explanatory variables are introduced being products of supply temperature and ambient air temperature, respectively, and two filtered variables calculated from the following fixed filters

$$\dot{m}_f^{-1}(t) = \frac{0.01}{1 - 0.99 q^{-1}} \frac{s(t) - r(t)}{p(t)} \tag{6.9}$$

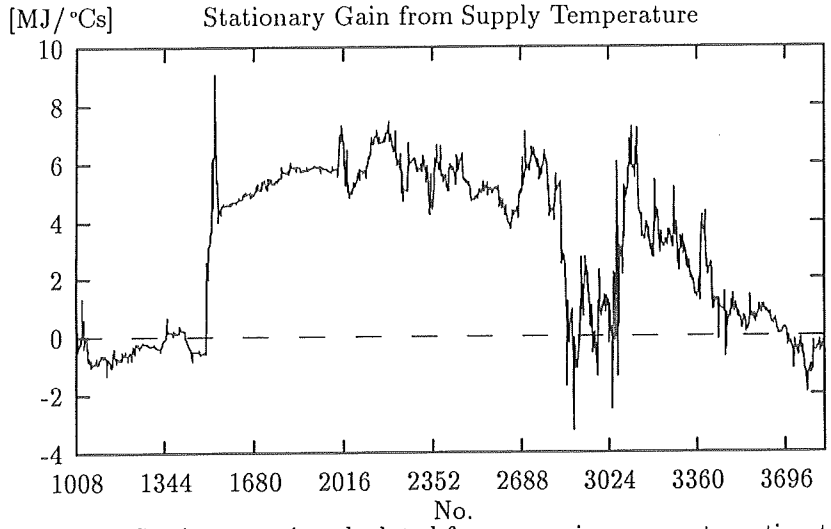


Figure 6.4: Stationary gain calculated from recursive parameter estimates.

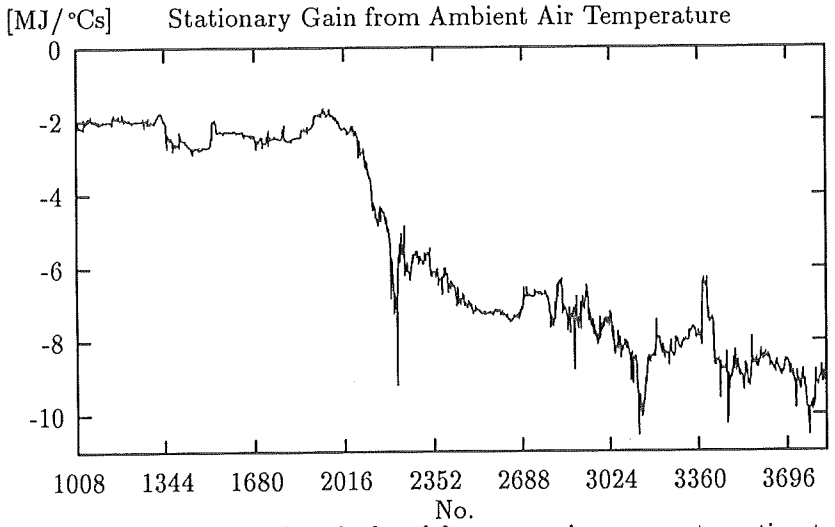


Figure 6.5: Stationary gain calculated from recursive parameter estimates.

Period	SS	$SS^{\%}$
1009-3843	2.38	2.44
1009-1953	1.70	2.98
1954-2898	2.24	2.43
2899-3843	3.02	1.73

Table 6.2: Performance of model (6.8).

and

$$a_f(t) = \frac{0.01}{1 - 0.99q^{-1}} a(t). \tag{6.10}$$

The parameters of these filters are chosen in accordance with the estimates of the models in Chapter 5.

The results of RLS estimation with $\lambda = 0.997$ of this model are shown in Table 6.2. The extension results in a slightly improved fit for the two last periods compared to (6.6). However, the off-line estimates of the nonlinear extensions to the transfer function model still results in better description of the data.

The residuals shown in Figure 6.6 are generally smaller than for the model (6.6) apart from a single observation close to no. 1510. This is the same point where a drop in supply temperature and heat load occurred indicating that this particular operating point deviates from normal operating conditions, for which reason it actually should be considered as containing erroneous information in relation to the desired model estimate. The recursive robust estimation algorithms, which are put forward in Chapter 7, are suitable for reducing the influence of such abnormal observations.

The autocorrelation of the residuals (Figure 6.7) shows almost the same pattern as for model (6.6), however, with even larger significant deviations from an assumption of white noise residuals. That is, even though the

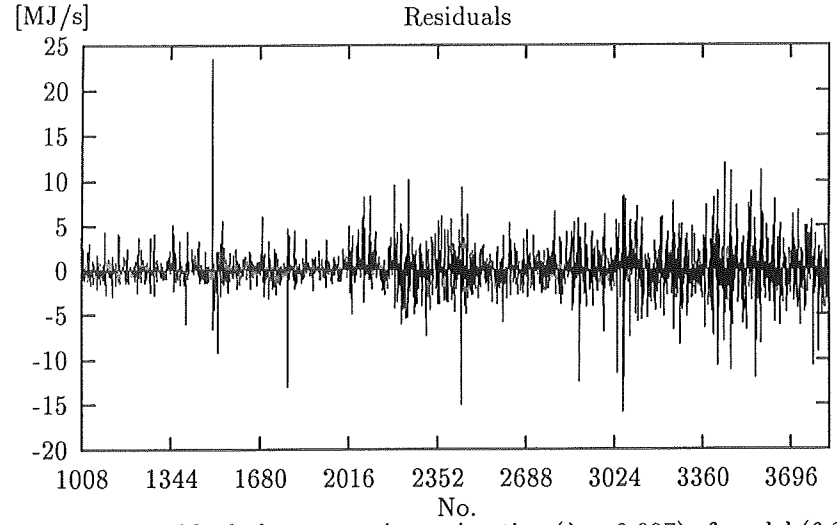


Figure 6.6: Residuals from recursive estimation ($\lambda = 0.997$) of model (6.8).

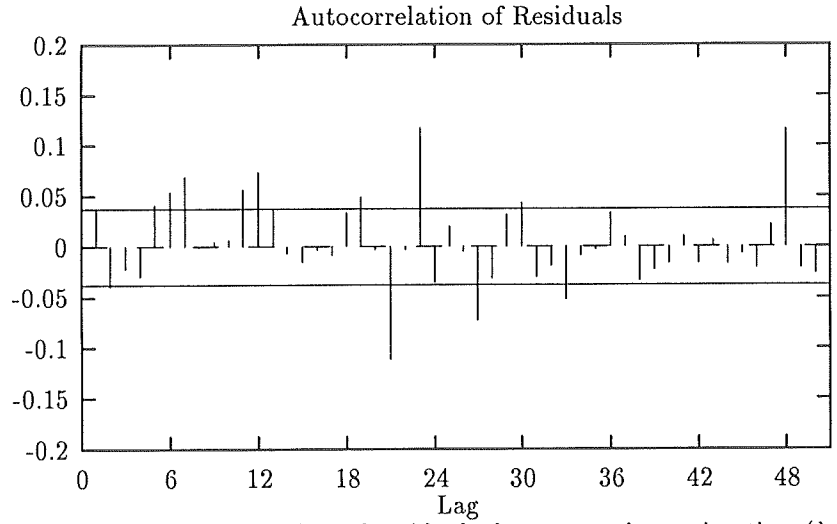


Figure 6.7: Autocorrelation of residuals from recursive estimation ($\lambda = 0.997$) of model (6.8).

model with second order explanatory variables (6.8) tends to imply a better description of the load dependence on supply temperature and ambient air temperature, a better modelling of the correlation of the load is not obtained.

6.1.2 RLS ESTIMATION OF j-STEP PREDICTORS

One of the results in Sejling (1987) was that the following model gave a good description of the supplied power for heating in the very same system as considered in this chapter

$$\begin{aligned}(1-\alpha_1 q^{-1}-\alpha_2 q^{-2})(1-\alpha_{24} q^{-24})p(t) = \\ (\beta_{1,0}+\beta_{1,1}q^{-1})\nabla s(t) + \\ (\beta_{2,0}+\beta_{2,1}q^{-1})a(t) + (\beta_{3,0}+\beta_{3,1}q^{-1})w(t) + \\ \mu_1(t) + \mu_2(t) I_{\{\mathrm{A}\}}(t) + l + e(t). \end{aligned} \tag{6.11}$$

$w(t)$ is the wind velocity, $\mu_1(t)$ is a diurnal profile common for all days, and $\mu_2(t)$ is an additional profile accounting for the difference between workdays and non-working days. All atypical days are classified as belonging to the group of non-working days.

Each of the profiles are modelled using trigonometric functions in the following way

$$\begin{aligned}\mu_*(t) = \\ m_{*1}\sin\left(\frac{2\pi t}{24}\right) + m_{*2}\cos\left(\frac{2\pi t}{24}\right) + m_{*3}\sin\left(\frac{2\pi t}{12}\right) + m_{*4}\cos\left(\frac{2\pi t}{12}\right) + \\ m_{*5}\sin\left(\frac{2\pi t}{8}\right) + m_{*6}\cos\left(\frac{2\pi t}{8}\right) + m_{*7}\sin\left(\frac{2\pi t}{6}\right) + m_{*8}\cos\left(\frac{2\pi t}{6}\right). \end{aligned} \tag{6.12}$$

In Chapter 4 it was found that none of the investigated methods were capable of giving a better description than the harmonic functions of the

diurnal variation of the load in Kulladal, Sweden. The trigonometric profile is indeed flexible, and a number of 8 parameters in the profile seems to be suitable for catching the significant characteristics of the systematic variation while at the same time smoothing the stochastic components of the load. The fact that also the diurnal autoregressive part in the model (6.11), which in fact is another way of describing the diurnal variation, was found to be significant in Sejling (1987), is difficult to explain. Perhaps the reason is that the profile adapts to slow changes while the autoregressive part takes care of faster changes, and when the model is estimated these two distinct parts are weighted to give the best prediction ability.

All the models presented until now in this chapter relate the heat load to the simultaneous weather input. Using these models for prediction requires prediction of the weather. In Sejling (1987) it was chosen not to make use of weather predictions explicitly, but only implicitly in the prediction models. For each prediction horizon a model is formulated, which only makes use of weather observations available at the instant of the prediction calculation. Hereby the prediction of weather is built into the model. Since the supply temperature is can be controlled this variable is not shifted in time in the estimation, and when using the model for predictions these can be computed as functions of the future supply temperature variation.

The model (6.11) in the linear form (6.1), where multiplicative terms is written

$$\begin{aligned} p(t) &= (\alpha_1 q^{-1} + \alpha_2 q^{-2} + \alpha_{24} q^{-24} + \alpha_{25} q^{-25} + \alpha_{26} q^{-26}) p(t) + \\ &\quad (\beta_{1,0} + \beta_{1,1} q^{-1}) \nabla s(t) + \\ &\quad (\beta_{2,1} q^{-1} + \beta_{2,2} q^{-2}) a(t) + (\beta_{3,1} q^{-1} + \beta_{3,2} q^{-2}) w(t) + \\ &\quad \mu_1(t) + \mu_2(t) I_{\{A\}}(t) + l + e(t), \end{aligned} \tag{6.13}$$

where the weather input now is shifted one step backwards in order to obtain one step predictions. This model for 1-hour predictions can be estimated adaptively by RLS with exponential forgetting.

The model for j-hour predictions can be obtained by using the weather inputs shifted j hours backwards instead of only one hour as in (6.13) and

then making iterative use of the model $j-1$ times. Hereby the explicit j-hour predictor, which in its linear form can be estimated by algorithm (6.3), can be found. However, it is easily realized that the number of parameters in this model will increase with the horizon of the prediction, j. In Sejling, Madsen, & Holst (1988) a proposed compromise for a reasonable number of parameters can be found. This linear j-hour predictor model is for $j < 23$

$$\begin{aligned} p(t) &= (\alpha_j q^{-j} + \alpha_{j+1} q^{-(j+1)} + \alpha_{24} q^{-24} + \alpha_{25} q^{-25} + \alpha_{26} q^{-26}) p(t) + \\ &\quad (\beta_{1,0} + \beta_{1,1} q^{-1} + \beta_{1,2} q^{-2} + \beta_{1,3} q^{-3}) \nabla s(t) + \\ &\quad (\beta_{2,j} q^{-j} + \beta_{2,j+1} q^{-j-1} + \beta_{2,j+2} q^{-j-2} + \beta_{2,j+3} q^{-j-3}) a(t) + \\ &\quad (\beta_{3,j} q^{-j} + \beta_{3,j+1} q^{-j-1} + \beta_{3,j+2} q^{-j-2} + \beta_{3,j+3} q^{-j-3}) w(t) + \\ &\quad \mu_1(t) + \mu_2(t) I_{\{A\}}(t) + l + e(t). \end{aligned} \tag{6.14}$$

For $j = 23, 24$ either one or two autoregressive parameters will in this linear model appear twice. Instead, it is chosen to add the autoregressive component $\alpha_{27} q^{-27}$ for $j = 23$ and additionally $\alpha_{28} q^{-28}$ for $j = 24$.

The results for application of these models on the data from Esbjerg (Data A) with j ranging from 1 to 24 and using the estimation algorithm (6.3) are listed in Table 6.3.

Not surprisingly does the standard deviation of the prediction errors increase with j. It is also obvious that the prediction errors are larger (numerically) in the period with the largest variation in load, but on the other hand, as seen from the values of $SS^{\%}$, the relative prediction ability is better in this period.

Comparing SS for $j = 1$ with the standard deviation of the first order difference of the load, it can be seen how the complicated predictor (6.13) compares to the simple random walk predictor (5.1). In the first low variance period it is remarkable how small the difference is, i.e. $SS = 1.59$ $(j = 1)$ compared to $\sqrt{V\{\nabla p(t)\}} = 2.77$. However, for the larger values of the load in the second period, the difference is more distinct, since $SS = 2.65$ and $\sqrt{V\{\nabla p(t)\}} = 7.24$.

*	SS		$SS^{\%}$	
j	505-2172	2173-3843	505-2172	2173-3843
1	1.59	2.65	2.69	1.88
3	3.00	5.47	5.00	3.95
6	4.36	8.22	7.64	5.97
9	4.82	10.83	8.68	7.37
12	5.08	12.74	9.14	8.27
15	5.43	14.15	9.63	8.94
18	5.83	15.34	10.25	9.58
21	6.08	16.07	10.59	9.99
24	6.28	16.53	10.69	10.41

Table 6.3: Prediction performance of the models (6.13) and (6.14).

The relatively good performance of the random walk predictor is of course due to the highly correlated load. A second order autoregressive model with the differenced supply temperature as the only external variable would probably do almost as good as the model (6.13) in predicting one hour ahead. However, when predicting further ahead the profile of the model becomes more valuable.

Concerning the 24-hour ahead prediction it is seen that SS for $j = 24$ equals 6.28 and 16.53, respectively, and $\sqrt{V\{\nabla_{24}p(t)\}}$ equals 6.75 and 18.81, respectively. This means that the simple predictor

$$\hat{p}(t \mid t-24) = p(t-24) \tag{6.15}$$

is close to the model (6.14) with $j = 24$ in performance.

IMPROVED TRANSFER FUNCTION MODELLING

In spite of the fact that the predictor models developed from model (6.11) have proved to work well in the implementation in Esbjerg, this model suffers from an objectionable quality in the transfer function from the independent variables. For instance, the transfer function from the ambient air temperature can be written

$$H_2(q^{-1}) = \frac{\beta_{2,0} + \beta_{2,1}q^{-1}}{(1-\alpha_1 q^{-1}-\alpha_2 q^{-2})(1-\alpha_{24}q^{-24})}. \tag{6.16}$$

This transfer function is indeed unreasonable, since there are no reason for the diurnal component $(1 - \alpha_{24}q^{-24})$ to occur in a physically realistic transfer function describing the influence of ambient air temperature on heat load, see Chapter 2 for a discussion of the dynamics of buildings. This goes for the supply temperature and wind velocity as well. Thus, a more reasonable model would be

$$\begin{aligned} p(t) \;=\; & \frac{\beta_{1,0} + \beta_{1,1}q^{-1}}{(1-\alpha_1 q^{-1}-\alpha_2 q^{-2})}\nabla s(t) + \\ & \frac{\beta_{2,0} + \beta_{2,1}q^{-1}}{(1-\alpha_1 q^{-1}-\alpha_2 q^{-2})}a(t) + \frac{\beta_{3,0} + \beta_{3,1}q^{-1}}{(1-\alpha_1 q^{-1}-\alpha_2 q^{-2})}w(t) + \\ & \mu_1(t) + \mu_2(t)I_{\{A\}}(t) + l + \\ & \frac{1}{(1-\alpha_1 q^{-1}-\alpha_2 q^{-2})(1-\alpha_{24}q^{-24})}e(t). \end{aligned} \tag{6.17}$$

The corresponding linear model for 1-hour predictions is

$$\begin{aligned} p(t) \;=\; & (\alpha_1 q^{-1} + \alpha_2 q^{-2} + \alpha_{24}q^{-24} + \alpha_{25}q^{-25} + \alpha_{26}q^{-26})p(t) + \\ & (\beta_{1,0} + \beta_{1,1}q^{-1} + \beta_{1,24}q^{-24} + \beta_{1,25}q^{-25})\nabla s(t) + \\ & (\beta_{2,1}q^{-1} + \beta_{2,2}q^{-2} + \beta_{2,24}q^{-24} + \beta_{2,25}q^{-25})a(t) + \\ & (\beta_{3,1}q^{-1} + \beta_{3,2}q^{-2} + \beta_{3,24}q^{-24} + \beta_{3,25}q^{-25})w(t) + \\ & \mu_1(t) + \mu_2(t)I_{\{A\}}(t) + l + e(t), \end{aligned} \tag{6.18}$$

and pursuing the principles for obtaining j-hour predictors as given in Sejling et al. (1988) the proposed general linear j-hour predictor model for

*	SS		$SS^{\%}$	
j	505-2172	2173-3843	505-2172	2173-3843
1	1.59	2.63	2.68	1.86
3	3.04	5.29	5.06	3.85
6	4.40	7.92	7.74	5.84
9	4.87	10.48	8.79	7.22
12	5.14	12.36	9.27	8.11
15	5.49	13.82	9.77	8.82
18	5.87	15.02	10.34	9.47
21	6.09	15.91	10.60	9.94
24	6.29	16.40	10.70	10.35

Table 6.4: Performance of the models (6.18) and (6.19).

$j < 21$ is

$$
\begin{aligned}
p(t) = \\
&(\alpha_j q^{-j} + \alpha_{j+1} q^{-(j+1)} + \alpha_{24} q^{-24} + \alpha_{25} q^{-25} + \alpha_{26} q^{-26}) p(t) + \\
&(\beta_{1,0} + \beta_{1,1} q^{-1} + \beta_{1,2} q^{-2} + \beta_{1,3} q^{-3} + \beta_{1,24} q^{-24} + \beta_{1,25} q^{-25}) \nabla s(t) + \\
&(\beta_{2,j} q^{-j} + \beta_{2,j+1} q^{-j-1} + \beta_{2,j+2} q^{-j-2} + \beta_{2,j+3} q^{-j-3} + \beta_{2,24} q^{-24} + \beta_{2,25} q^{-25}) a(t) + \\
&(\beta_{3,j} q^{-j} + \beta_{3,j+1} q^{-j-1} + \beta_{3,j+2} q^{-j-2} + \beta_{3,j+3} q^{-j-3} + \beta_{3,24} q^{-24} + \beta_{3,25} q^{-25}) w(t) + \\
&\mu_1(t) + \mu_2(t) I_{\{A\}}(t) + l + e(t). \qquad (6.19)
\end{aligned}
$$

For $j \in [21, 24]$ some of the regressors appear twice. In these cases, both for the autoregressive polynomial and for each of the transfer polynomials, a number of additional regressors, corresponding to the number of regressors appearing twice, is added at the end of each polynomial. These models are only a proposed compromise for the number of elements entering the linear predictor models. The results obtained using the algorithm (6.3) with $\lambda = 0.998$ and the models (6.18) and (6.19) are shown in Table 6.4.

For the first data period the improvement is dubious, whereas in the second period it seems to imply an improved prediction ability to use the more physically reasonable transfer function. The drawback is the increased number of parameters, the uncertainty of which contributes to the uncertainty of the predictions.

REMOVING THE DEPENDENCE ON WIND VELOCITY

*	SS		$SS^{\%}$	
j	505-2172	2173-3843	505-2172	2173-3843
1	1.59	2.63	2.68	1.86
3	3.02	5.27	5.03	3.83
6	4.37	7.88	7.64	5.79
9	4.82	10.43	8.65	7.17
12	5.07	12.33	9.10	8.06
15	5.41	13.79	9.56	8.76
18	5.77	15.02	10.08	9.40
21	5.97	15.91	10.32	9.87
24	6.15	16.40	10.38	10.29

Table 6.5: Performance of the models (6.18) and (6.19) without the transfer function from wind velocity.

One way of lowering the number of parameters is to remove the wind velocity from the model. The experience is that the influence of the wind is inferior to the influence of ambient air temperature. Moreover, though the correlation between wind and load exists, the wind velocity is difficult to predict, and therefore it is dubious, whether the use of wind velocity is of any good for long range predictions. The results obtained when omitting the wind transfer function in the models (6.18) and (6.19) are shown in Table 6.5 for $\lambda = 0.998$. For $j = 1, 3$ the results are comparable in Ta-

ble 6.4 and Table 6.5, but for higher values of j the prediction ability is better when the wind velocity is not used. The uncertainty arising as a consequence of more parameters to estimate is thus clearly exceeding the additional information obtained from using the wind velocity.

REMOVING THE RESTRICTION OF DIFFERENCING SUPPLY TEMPERATURE

In the modelling investigation in Chapter 5 it turned out that using the supply temperature as an explanatory variable instead of the differenced supply temperature was beneficial to the fit of the model. Consequently, it is appropriate to investigate whether the same applies for adaptive RLS estimation of a parameter linear model. By interchanging the forced differencing of the supply temperature by a free autoregressive parameter the following model for 1-hour predictions appears

$$\begin{aligned} p(t) &= (\alpha_1 q^{-1} + \alpha_2 q^{-2} + \alpha_{24} q^{-24} + \alpha_{25} q^{-25} + \alpha_{26} q^{-26}) p(t) \\ &+ (\beta_{1,0} + \beta_{1,1} q^{-1} + \beta_{1,2} q^{-2} + \beta_{1,24} q^{-24} + \beta_{1,25} q^{-25} + \beta_{1,26} q^{-26}) s(t) \\ &+ (\beta_{2,1} q^{-1} + \beta_{2,2} q^{-2} + \beta_{2,24} q^{-24} + \beta_{2,25} q^{-25}) a(t) \\ &+ \mu_1(t) + \mu_2(t) I_{\{A\}}(t) + l + e(t). \end{aligned} \tag{6.20}$$

The wind velocity does not enter the model, and the trigonometric profile with harmonics up to order three (6.12) is used. The corresponding model for j-hour predictions for $j < 21$ becomes

$$\begin{aligned} p(t) &= (\alpha_j q^{-j} + \alpha_{j+1} q^{-(j+1)} + \alpha_{24} q^{-24} + \alpha_{25} q^{-25} + \alpha_{26} q^{-26}) p(t) + \\ &+ (\beta_{1,0} + \beta_{1,1} q^{-1} + \beta_{1,2} q^{-2} + \beta_{1,3} q^{-3} + \beta_{1,4} q^{-4} + \\ &\quad \beta_{1,24} q^{-24} + \beta_{1,25} q^{-25} + \beta_{1,26} q^{-26}) s(t) \\ &+ (\beta_{2,j} q^{-j} + \beta_{2,j+1} q^{-j-1} + \beta_{2,j+2} q^{-j-2} + \beta_{2,j+3} q^{-j-3} + \\ &\quad \beta_{2,24} q^{-24} + \beta_{2,25} q^{-25}) a(t) \\ &+ \mu_1(t) + \mu_2(t) I_{\{A\}}(t) + l + e(t). \end{aligned} \tag{6.21}$$

For $j = 21, 24$ some parameters appear twice. In this case it is chosen, as previously, to introduce additional components following the diurnal com-

*	SS		$SS^{\%}$	
j	505-2172	2173-3843	505-2172	2173-3843
1	1.58	2.63	2.67	1.87
3	2.98	5.29	4.95	3.86
6	4.11	7.48	7.06	5.63
9	4.50	8.50	7.93	6.31
12	4.76	8.91	8.48	6.54
15	5.06	9.14	8.95	6.72
18	5.40	9.33	9.46	6.90
21	5.58	9.47	9.71	7.10
24	5.75	9.43	9.85	7.32

Table 6.6: Performance of the models (6.20) and (6.21).

ponents of the dependence on autoregressive load and ambient air temperature to end up with the same number of parameters.

Table 6.6 lists the results for RLS estimation with $\lambda = 0.998$ of these models. An evident improvement is seen compared to the model based on differenced supply temperature for the second criterion period, when j is increased. For the first period there is also seen an improvement, however, less pronounced.

This improvement is reasonable, considering the results of Chapter 5 and the estimated stationary gain from ambient air temperature shown in Figure 6.5. However, also in the control of supply temperature the actual ambient air temperature is used meaning that the supply temperature contains information about the ambient air temperature. Consequently, it is probable that the use of future supply temperature in the j-step predictors actually also utilizes the built-in information about future ambient air temperature.

INTRODUCING DEPENDENCE ON NONLINEAR EXPLANATORY VARIABLES

It is relevant to investigate whether introduction of the nonlinear explanatory variables, as in model (6.8), implies an improved prediction ability of the j-step predictors. The proposed model for predictions one hour ahead is

$$\begin{aligned} p(t) &= (\alpha_1 q^{-1} + \alpha_2 q^{-2} + \alpha_{24} q^{-24} + \alpha_{25} q^{-25} + \alpha_{26} q^{-26}) p(t) \\ &+ (\beta_{1,0} + \beta_{1,1} q^{-1} + \beta_{1,2} q^{-2} + \beta_{1,24} q^{-24} + \beta_{1,25} q^{-25} + \beta_{1,26} q^{-26}) s(t) \\ &+ (\beta_{2,1} q^{-1} + \beta_{2,2} q^{-2} + \beta_{2,24} q^{-24} + \beta_{2,25} q^{-25}) a(t) \\ &+ (\beta_{3,0} + \beta_{3,1}^{-1} + \beta_{3,24} q^{-24} + \beta_{3,25} q^{-25}) \dot{m}_f^{-1}(t-1) s(t) \\ &+ (\beta_{4,1} q^{-1} + \beta_{4,2} q^{-2} + \beta_{4,24} q^{-24} + \beta_{4,25} q^{-25}) a_f(t) a(t) \\ &+ \mu_1(t) + \mu_2(t) I_{\{A\}}(t) + l + e(t), \end{aligned} \tag{6.22}$$

where it is seen that the introduced observation of the mass flow is shifted one hour back as its value at time t is calculated from observations of load and return temperature up to time t. Similarly the proposed j-hour predictor becomes

$$\begin{aligned} p(t) &= (\alpha_j q^{-j} + \alpha_{j+1} q^{-(j+1)} + \alpha_{24} q^{-24} + \alpha_{25} q^{-25} + \alpha_{26} q^{-26}) p(t) + \\ &+ (\beta_{1,0} + \beta_{1,1} q^{-1} + \beta_{1,2} q^{-2} + \beta_{1,3} q^{-3} + \beta_{1,4} q^{-4} + \\ &\quad \beta_{1,24} q^{-24} + \beta_{1,25} q^{-25} + \beta_{1,26} q^{-26}) s(t) \\ &+ (\beta_{2,j} q^{-j} + \beta_{2,j+1} q^{-j-1} + \beta_{2,j+2} q^{-j-2} + \beta_{2,j+3} q^{-j-3} + \\ &\quad \beta_{2,24} q^{-24} + \beta_{2,25} q^{-25}) a(t) \\ &+ (\beta_{3,0} + \beta_{3,1} q^{-1} + \beta_{3,24} q^{-24} + \beta_{3,25} q^{-25}) \dot{m}_f^{-1}(t-j) s(t) \\ &+ (\beta_{4,j} q^{-j} + \beta_{4,j+1} q^{-j-1} + \beta_{4,24} q^{-24} + \beta_{4,25} q^{-25}) a_f(t) a(t) \\ &+ \mu_1(t) + \mu_2(t) I_{\{A\}}(t) + l + e(t), \end{aligned} \tag{6.23}$$

and for such values of j where components appear twice new components are added using the same principle as previously.

Table 6.7 contains the performance results when these models are estimated using RLS with $\lambda = 0.998$. For the period 505-2172 it is seen that it is better

*	SS		$SS^{\%}$	
j	505-2172	2173-3843	505-2172	2173-3843
1	1.59	2.59	2.70	1.84
3	3.03	5.20	5.09	3.81
6	4.24	7.25	7.42	5.45
9	4.64	8.05	8.29	6.05
12	4.89	8.41	8.78	6.23
15	5.20	8.66	9.25	6.42
18	5.55	8.92	9.76	6.67
21	5.71	9.17	10.01	6.94
24	5.78	9.50	10.03	7.38

Table 6.7: Performance of the models (6.22) and (6.23).

to use the models without the nonlinear variables. This is reasonable as in this period the supply temperature is almost constant as is the level of the ambient air temperature. On the contrary is seen an improved prediction ability, at least for $j = 24$, for the period 2173-3843 when the predictors with nonlinear explanatory variables are used.

6.2 STATE SPACE MODEL OF A DIURNAL AND WEEKLY PROFILE

A state space model allows for the description of states, which are subject to stochastic variations, and which are observed in noise. Hereby it is possible to describe the separation of time-variation of states (process noise) and observation noise. For instance, the parameters of a transfer function being linear in the parameters may be introduced into a state vector, where the dependent variable is the regressor determined linear combination of

the parameters. This implies that the parameters can be described by the covariance matrix of the process noise to have variation, which is of different character and according to a given correlation structure. In Sejling (1987) and Sejling et al. (1988) the heat load in a DH system is described by a state space model in which the covariance matrix for the parameter variations and the measurement noise are estimated, using the Kalman filter for parameter update, to match the underlying variation of the parameters in the load model.

Harvey (1989) gives an extensive account of the class of structural time series models introduced into state space formulations and updated by the Kalman filter. In this section the load is assumed to be separated into a nominal component corresponding to a systematic diurnal and weekly pattern and a residual component describing the correlation time and the dependence on explanatory variables. This approach has previously been applied for modelling the electricity power load, see e.g. Holst (1977), Holst & Jonsson (1984) and Holst & Ekelund (1987). In spite of the previous conclusion that weekly differencing is not useful in a heat load model, the systematic load variation is in this section described by a state for each hour in the week superposed a state for each hour in a day, where the last 24 states are introduced to obtain faster adaption of the model.

The above mentioned assumption that the load $p(t)$ can be separated into a nominal and a stochastic component gives

$$p(t) = p_n(t) + p_r(t). \tag{6.24}$$

The objective of the residual component $p_r(t)$ is to describe the faster changes in the load occurring as a consequence of the dynamical characteristics of the district heating system and the influence of weather condition. It is assumed that this component can be described by a transfer function model.

The objective of the nominal component $p_n(t)$ is to describe the slowly varying diurnal and weekly pattern of the load. The nominal component

is written as

$$p_n(t) = p_{n,1}(t) + p_{n,2}(t), \tag{6.25}$$

in which the two components accounting for the weekly and the diurnal pattern are obtained by the following exponential smoothing of the load

$$\begin{cases} p_{n,1}(t) &= (1-\alpha)p_{n,1}(t-168) + \alpha p(t) \\ p_{n,2}(t) &= \nu p_{n,2}(t-1) + \mu p_{n,2}(t-24) + (1-\nu-\mu)(p(t)-p_{n,1}(t)) \end{cases}, \tag{6.26}$$

with $\alpha, \nu, \mu \in [0,1]$. This is an extension of the well-known exponential smoothing algorithms, see e.g. Abraham & Ledolter (1983). Assuming that the diurnal and weekly patterns develop from white noise processes $\{e_1(t)\}$ and $\{e_2(t)\}$ with variances σ_1^2 and σ_2^2 gives by appropriate reformulation (Holst & Ekelund, 1987)

$$\begin{cases} p_{n,1}(t) &= p_{n,1}(t-168) + \frac{\alpha}{1-\alpha}\beta p_{n,2}(t-1) + \frac{\alpha}{1-\alpha}(1-\beta)p_{n,2}(t-24) + v_1(t) \\ p_{n,2}(t) &= \beta p_{n,2}(t-1) + (1-\beta)p_{n,2}(t-24) + v_2(t) \end{cases}, \tag{6.27}$$

where

$$E\begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix} = 0 \tag{6.28}$$

and

$$Cov\begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix} = \begin{bmatrix} \left(\frac{\alpha}{1-\alpha}\right)^2\sigma_2^2 + \sigma_1^2 & \frac{\alpha}{1-\alpha}\sigma_2^2 \\ \frac{\alpha}{1-\alpha}\sigma_2^2 & \left(\frac{\alpha}{1-\alpha}\right)^2\sigma_2^2 + \sigma_1^2 \end{bmatrix}. \tag{6.29}$$

To write the observational equation (6.24) and the processes (6.27) in a state-space formulation the following vectors are introduced

$$\begin{aligned} \boldsymbol{x}^{\mathrm{T}}(t) &= \begin{bmatrix} p_{n,1}(t) & p_{n,1}(t-1) & \ldots & p_{n,1}(t-167) & p_{n,2}(t) & \ldots & p_{n,2}(t-23) \end{bmatrix} \\ \boldsymbol{v}^{\mathrm{T}}(t) &= \begin{bmatrix} v_1(t) & v_2(t) \end{bmatrix} \end{aligned}. \tag{6.30}$$

Then the state-space formulation is

$$\boldsymbol{x}(t+1) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{D}\boldsymbol{v}(t) \tag{6.31}$$

$$p(t) = \boldsymbol{C}\boldsymbol{x}(t) + p_r(t) \tag{6.32}$$

$\boldsymbol{A}$ is a 192×192 matrix

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 & \beta\frac{\alpha}{1-\alpha} & 0 & \cdots & 0 & (1-\beta)\frac{\alpha}{1-\alpha} \\ 1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & \beta & 0 & \cdots & 0 & (1-\beta) \\ 0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \tag{6.33}$$

with the elements different different from zero given by

$$A_{1,168} = 1, \quad A_{1,169} = \beta\frac{\alpha}{1-\alpha}, \quad A_{1,192} = (1-\beta)\frac{\alpha}{1-\alpha},$$
$$A_{169,169} = \beta, \quad A_{169,192} = 1-\beta \quad \text{and}$$
$$A_{i,i-1} = 1 \quad \text{for} \quad i = 2, \ldots, 168, 170, \ldots, 192.$$

The $\boldsymbol{C}$-matrix picks out the 1st and 169th element of the state vector, i.e.

$$\boldsymbol{C} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \end{bmatrix}, \tag{6.34}$$

and the $\boldsymbol{D}$-matrix describes the correlation structure of the influence of the errors driving the changes in the profiles. The correlation between errors on current profile values and other profile values can, for instance, be assumed to decay exponentially with the distance in time. Thus, the proposal for parameterization of the $\boldsymbol{D}$-matrix is ($\rho, \gamma \in [0,1]$)

$$\boldsymbol{D}^{\mathrm{T}} = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^2 & \rho & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 & \gamma & \gamma^2 & \cdots & \gamma^2 & \gamma \end{bmatrix}. \tag{6.35}$$

The choice of D implies that the prediction error for a given hour in the week influences not only the corresponding state but also nearby the states.

Having specified the state space model of the load the Kalman filter can be applied, the recursions of which can be found in Appendix A.

To obtain the best prediction ability using the modelling scheme outlined above for predictions of heat load, the parameters entering the model have to be tuned to the actual system and data. Exactly as in the case of traditional transfer function modelling, optimal parameters with respect to a certain criterion can be found.

In this investigation the parameters $\alpha, \beta, \rho, \gamma, \sigma_1^2, \sigma_2^2$ and σ_y^2, where the last mentioned is the variance of the residual component in (6.32), are determined to give a minimum SS for $j = 1$. However, due to considerable computation is minimization not carried out by a minimization algorithm. It was carried out by applying the model repeatedly and changing the parameters to obtain an improved fit. Most effort was put down on finding the optimal values of α, β, ρ and γ as the fit was most sensitive to the values of these parameters.

To illustrate the consequences of having a correlation structure in the evolution of the states in the profile, optimal values are found both for $\gamma = \rho = 0.0$ and for γ and ρ free.

For $\gamma = \rho = 0.0$ the values that was found to give the lowest value of SS are

$$\alpha = 0.2, \quad \beta = 0.4$$
$$\sigma_1^2 = \sigma_2^2 = \sigma_y^2 = 1.0$$

The results with these parameter values using the state space model are listed in Table 6.8. Due to the correlation being zero in the state update, the measures are almost constant for $j \geq 6$. For small values of j the state space predictions are much worse than those obtained with the j-hour predictors with trigonometric profiles, but for the largest values of j the results are indeed comparable for the two different modelling approaches.

It is also noticeable that the one-hour ahead are worse than those obtained with the simple random walk model. Indeed the class of models within the state space profile model does not include the random walk model.

$\alpha = 0.2, \beta = 0.4, \gamma = \rho = 0.0$ $\sigma_1^2 = \sigma_2^2 = \sigma_y^2 = 1.0$				
*	SS		$SS^{\%}$	
j	505-2172	2173-3843	505-2172	2173-3843
1	5.35	13.60	9.22	8.49
3	6.77	17.18	11.63	10.68
6	7.01	17.72	12.02	11.00
9	7.05	17.73	12.05	10.99
12	7.07	17.72	12.07	10.97
15	7.09	17.72	12.08	10.96
18	7.09	17.72	12.09	10.95
21	7.10	17.71	12.09	10.94
24	7.16	17.71	12.22	10.95

Table 6.8: Prediction performance of the state space model with $\gamma = \rho = 0$, $\alpha = 0.2$, $\beta = 0.4$ and $\sigma_1^2 = \sigma_2^2 = \sigma_y^2 = 1$.

One important difference between the recursive transfer function modelling and the state space modelling is that the supply temperature is used in the estimation of the transfer function model and future supply temperatures are used in predictions with the transfer function model, whereas this information is neither used in the update mechanism of the state space description nor in the predictions calculated with state space model. This information is indeed valuable, since sudden changes in supply temperature implies instantaneous reactions in the load. When this information is not used in the state update, the states will adapt to the changes actually explained by the supply temperature, thus resulting in a worse prediction ability of the state space model. Moreover, at the time of the predicted value, information about the supply temperature until that moment, would also improve the predictions. Thus, the lacking use of the supply temperature deteriorate the prediction ability of the state space model in two

*	SS		$SS^{\%}$	
j	505-2172	2173-3843	505-2172	2173-3843
1	1.96	3.60	3.31	2.51
3	3.84	7.78	6.42	5.24
6	5.46	11.06	9.39	7.52
9	5.96	12.94	10.46	8.72
12	6.25	14.25	10.90	9.34
15	6.72	15.46	11.71	9.92
18	7.05	16.34	12.37	10.43
21	7.21	16.76	12.64	10.56
24	7.49	17.46	12.96	10.88

Table 6.9: Performance of the state space model as in Table 6.8 followed by residual modelling with respectively (6.18) and (6.19).

ways.

In Holst & Ekelund (1987) it is proposed to model the residual component $p_r(t)$, i.e. the prediction errors of the state space model, by using an transfer function model. In the present investigation the j-hour state space prediction errors are, for each value of j, modelled with the matching j-hour prediction model given either by (6.18) or (6.19) in which the trigonometric profile is left out. These models are estimated using the algorithm (6.3) with $\lambda = 0.998$. The results are shown in Table 6.9. It is seen that, when $\rho = \gamma = 0.0$ in the state space model, the transfer function model considerably improves the prediction ability for lower values of j. The improvement decreases with j, and for $j = 24$ the measures SS and $SS^{\%}$ are almost the same with and without the transfer function model. For all values of j the combined state space- and transfer function modelling procedure are clearly inferior to the recursively estimated j-hour predictor models with trigonometric profiles.

$\alpha = 0.55, \beta = 0.8, \gamma = \rho = 0.9$ $\sigma_1^2 = \sigma_2^2 = \sigma_y^2 = 1.0$				
*	SS		$SS^{\%}$	
j	505-2172	2173-3843	505-2172	2173-3843
1	2.28	5.44	3.99	3.58
3	4.06	11.03	7.11	7.03
6	5.48	14.99	9.43	9.38
9	6.20	16.80	10.61	10.37
12	6.71	17.85	11.48	10.90
15	7.22	18.84	12.32	11.47
18	7.60	19.50	12.93	11.87
21	7.90	19.75	13.41	12.03
24	8.20	20.08	13.94	12.20

Table 6.10: Prediction performance of the state space model with $\gamma = \rho = 0.9$, $\alpha = 0.55$, $\beta = 0.8$ and $\sigma_1^2 = \sigma_2^2 = \sigma_y^2 = 1$.

The parameters giving the lowest value of SS among those tried, when correlation in the state variations is allowed, are

$\alpha = 0.55, \quad \beta = 0.8$
$\rho = \gamma = 0.9$
$\sigma_1^2 = \sigma_2^2 = \sigma_y^2 = 1.0$

The prediction results obtained with these parameters in the state space model are listed in Table 6.10. For $j = 1$ the prediction ability is only a little inferior to the combined transfer function model with trigonometric profile. For increasing j the difference between these two approaches increases, and it end up to be worse than the state space model with $\rho = \gamma = 0.0$. Part of the explanation is of course that the pseudo optimization was carried out with respect to minimization of the 1-hour ahead prediction errors.

*	SS		$SS^{\%}$	
j	505-2172	2173-3843	505-2172	2173-3843
1	1.87	3.26	3.18	2.29
3	3.64	7.52	6.05	5.06
6	5.50	11.66	9.49	7.77
9	6.10	14.49	10.79	9.45
12	6.58	16.33	11.54	10.34
15	7.28	17.81	12.69	11.10
18	7.89	18.95	13.78	11.74
21	8.27	19.48	14.51	12.00
24	8.68	20.01	15.19	12.20

Table 6.11: Performance of the state space model as in Table 6.10 followed by residual modelling with respectively (6.18) and (6.19).

Probably the optimal parameter values are different for different prediction horizon. Therefore, a better performance than listed in Table 6.10 could be obtained for $j > 1$, but this would indeed be a time consuming process to find the optimal parameters for each value of j. It should not be forgotten that the optimal parameters also depend on the specific system. Moreover, the system might change implying the necessity of finding possible shifts of the optimal parameters.

As seen from Table 6.11 the improvement of the transfer function modelling of the residuals is not as clear as for $\rho = \gamma = 0.0$. This is due to the allowance of the correlation in the state variations to take care of the autocorrelation in the load. Hereby, less is left to the transfer function model.

6.3 CONCLUSION

For comparison with the modelling results in Chapter 5 two of the dynamic models are reformulated to parameter-linear forms. These models are estimated using RLS with simple exponential forgetting. The first model being linear in the explanatory variables supply temperature and ambient air temperature shows to adapt the variations of the load dependence on these variates. The fit, however, is not as good as for the linear transfer function model in Chapter 5. The second model is extended with dependence on two nonlinear transformations of explanatory variables, viz. supply temperature multiplied with low-pass filtered inverse mass flow and ambient air temperature multiplied with low-pass filtered ambient air temperature. This model slightly improves the fit, but still the nonlinear models of Chapter 5 are superior.

For prediction of heat load three distinct approaches are investigated, i.e.

1. Linear transfer function models with trigonometric diurnal profiles.
2. Linear transfer function models based on diurnal and weekly differences of load and independent variables.
3. State space models of a diurnal and weekly profile combined with an transfer function model of the difference between the actual load and the load prediction from the state space model.

These approaches are all tried with values of the prediction horizon, j, ranging from 1 to 24 hours. For all of the investigated values of j, it is found that a model of the first group is superior to the others. This particular model in its basic formulation is

$$\begin{aligned} p(t) \;=\; & \frac{\beta_{1,0}+\beta_{1,1}q^{-1}+\beta_{1,2}q^{-2}}{1-\alpha_1 q^{-1}-\alpha_2 q^{-2}}s(t) + \frac{\beta_{2,0}+\beta_{2,1}q^{-1}}{1-\alpha_1 q^{-1}-\alpha_2 q^{-2}}a(t) + \\ & \mu_1(t)+\mu_2(t)I_{\{\mathrm{A}\}}(t) + l + \\ & \frac{1}{(1-\alpha_1 q^{-1}-\alpha_2 q^{-2})(1-\alpha_{24}q^{-24})}e(t) \end{aligned}$$

with the trigonometric profile

$$\mu_*(t) = m_{*1}\sin\left(\frac{2\pi t}{24}\right) + m_{*2}\cos\left(\frac{2\pi t}{24}\right) + m_{*3}\sin\left(\frac{2\pi t}{12}\right) + m_{*4}\cos\left(\frac{2\pi t}{12}\right) + m_{*5}\sin\left(\frac{2\pi t}{8}\right) + m_{*6}\cos\left(\frac{2\pi t}{8}\right) + m_{*7}\sin\left(\frac{2\pi t}{6}\right) + m_{*8}\cos\left(\frac{2\pi t}{6}\right).$$

This model has shown to give a good description of the heat load, but it is not directly applicable for calculation of predictions, unless predictions of the ambient air temperature are available as input to the model.

Since predictions of ambient air temperature to some degree are uncertain, it is assessed that the prediction ability of the model is not deteriorated if the input of ambient air temperature is shifted backwards in time in such a way that only measurements of ambient air temperature available at the time of the prediction calculation is used in the model. The alternative would be to use the model as it is, hereby requiring predictions of ambient air temperature as input to the model in the calculation of the prediction.

One of the reasons for this assessment is that the ambient air temperature most of the time is slowly varying and cyclic with a diurnal period. When the ambient air temperature is behaving this way its influence is easily described implicitly in the prediction model.

On the other hand, predictions of ambient air temperature could indeed be valuable as input to load prediction calculations, if it is possible to predict abrupt changes in the ambient air temperature occurring as a consequence of passages of pressure fronts. This kind of ambient air temperature predictions may require the use of synoptic scale meteorological models, which on the other hand has to be rather exact at the location of the DH system when predicting accurately the hour of the front passage.

The application of different predictor models showed that the use of future supply temperature significantly improves the prediction ability as the prediction horizon increases. It is questionable whether this also will be the

case in a practical prediction situation, where the heat load is predicted as a function of future supply temperatures. The case is that in this investigation data has been used, for which momentary ambient air temperatures have been used in the control of supply temperature. This means that the future supply temperatures contain information about future ambient air temperatures, the use of which possesses the potential of improving the predictions. That is, in the data the supply temperature and the ambient air temperature is correlated due to the way the supply temperatures has been controlled, and this correlation implies that information about future ambient air temperatures are used when future supply temperatures are used as explanatory variables.

Concerning the wind velocity, it is found that this variable do not improve the load predictions for the considered system in Esbjerg for the analyzed data period. Even though the heat consumption evidently is correlated with the wind velocity, the influence is so small that the uncertainty arising as consequence of both additional parameters in the prediction model and the stochastic nature of the wind velocity itself, exceeds the improvement of using the wind velocity (non-future) in the predictions.

In Chapter 2 on load in DH systems it was stated that, because of the considerable transport delay, a description of the transmission and distribution net is necessary when modelling the load. In the model just above the net is taken into account by the transfer function from the supply temperature. Indeed, this is a simple way of modelling the network. Nevertheless, taking into consideration that no measurements from the network are used in the load model, the model seems to describe the effect of the distributed network very well.

The transfer function from the supply temperature simply describes the influence of a given course of variation of the supply temperature. That is, when the supply temperature is changed, it is not recognized at the consumers until the temperature front reaches the consumers with the delay of the transportation time. Thus, the change of supply temperature implies a momentary change of supplied heat, because a reaction in the flow in the

net is not occurring before the temperature front reaches the consumers. Since the consumers are geographically distributed with individual transport delays from the heat production unit, the effect of changing the supply temperature fades out as the temperature front reaches the consumers.

Another assessment in the search for the best model and method for prediction of heat load has been that the best prediction ability for a specific prediction horizon is obtained by using a prediction model dedicated for this particular prediction horizon. This assessment originates in the fact that the criterion to be minimized for estimation of the parameters in a certain model should match the purpose of the model. For instance, when a model is to be used for prediction one hour ahead, the criterion to be minimized ought to involve the one-hour ahead prediction errors, whereas for prediction 15 hours ahead the 15-hour prediction errors should enter the criterion. This approach clearly implies an increasing computational burden, when the number of prediction wanted horizons is increased.

The recommendation, based on the investigation in this chapter, of how to predict heat load is for $j \leq 24$ to apply a j-hour predictor based on an transfer function model with trigonometric profile.

.....

The implementation of the estimation schemes in this chapter were programmed Fortran, and the calculations were carried out on a HP-9000/750 using double precision arithmetic.

.....

7

Recursive Robust Estimation of AR Parameters

In the preceding chapters, the purpose of the investigation has been to develop models and methods for description and prediction of heat load in DH systems. One of the points, mentioned as an important basis for the practical use of heat load models, was the adaptivity of the models implying their ability to track time variations and system changes. This requires the application of on-line estimation using recursive methods. However, the occurrence of gross errors in data collected on-line is a matter that can have deteriorating influence on the resulting parameter estimates.

A desirable quality of the algorithm used to estimate the parameters is that it is able to protect the estimates from the destructive influence from any such gross errors while retaining the ability to track changes in the system. Indeed, this a demanding quality since it can obviously be difficult to separate abnormal prediction errors caused by gross errors in data from prediction errors arising as a consequence of abrupt changes in the system.

In this chapter the investigation is devoted to the problem of recursive estimation of models of AR structure in settings, where the possibility of erroneous observations exists, and where at the same time adaptivity in the estimation is required.

7.1 ERRONEOUS OBSERVATIONS

The classification of observations being erroneous is related to the application of a model. For the models discussed in the present thesis the purpose is to give a reliable description of the load variations, and furthermore, to be able to predict the load a number of hours in advance. However, these qualities of the models only apply to operating conditions classified as being normal. That is, it is not required that a model is capable of predicting any unusual load situation, occurring, for instance, as a consequence of a dropout of a turbine. Neither is it required that a model is in accordance with the actual load variations in such situations. The reason for this is that a DH system is so complex that a detailed physical model of the system would be necessary for a reliable description under non-normal operating conditions. In the preceding chapters, only models that approximate the system well under ordinary conditions are considered, and thus, whenever the system is in an unusual operating situation the observations are considered to be erroneous.

Another kind of erroneous observation arises when malfunction of the equipment, which takes care of recording or transmission of measurements, occurs. Typically this implies that measurements will be lacking or they will hit an upper or lower limit, specified in the software for recording and transmission. As such, these erroneous observations are easily classifiable.

Systematic deviations from actual values, due to bias in the recording equipment or lacking representativity (typically of weather observations), is also a frequent source of errors in the observations. However, the implication of such errors can be of minor significance due to the capability of the model estimate to adapt to the biased observations.

The discussion above considers the deviations of the observations from underlying true observations coursed by effects, which can be given another physical explanation than simply being process or measurement noise. They are to some degree easily identifiable as errors, at least in an off-line situation. In the on-line case it is also possible to set up rules for de-

tection of such errors in the observations. These may be simple rules in which knowledge about upper and lower limits or maximum gradients can be used to reject unlikely observations. Having detected erroneous observations, precautions in the recursive update can be taken, based on the knowledge that these observations should not influence the model estimate. The algorithms presented in this chapter include a detection of deviating observations as well as a robustified estimation.

Whenever observations are erroneous in such a way that consistence with model assumptions is lacking these observations are termed outliers. This means that the sources of errors mentioned above all lead to outlying observations.

Usually observations will be contaminated with noise coming from inaccuracy in the measuring phase as well as different sources of stochastic nature in the process. When these noise components are accounted for in the stochastic model, by assumption of a probability density function on the noise components, this is not a source of outliers. If, however, the assumed probability density function does not correspond to that of the actual noise components, the consistence with model assumptions is lost, and thus, outliers are present.

7.2 OUTLIER MODELS

In this section models of outliers, usually treated in the literature on analysis of estimator properties, are presented.

In the ARMA model

$$A(q^{-1})y(t) = C(q^{-1})e(t) \tag{7.1}$$

the innovation sequence $\{e(t)\}$ is traditionally assumed to be zero mean, independent and identically distributed normal random variables, $NIID(0;\sigma_e^2)$. As mentioned in Chapter 5 the ML estimator of the parameters in A and C in this case simplifies to a nonlinear LS estimator.

Traditionally two distinct kinds of outlier types are considered in the setting of ARMA models, viz. outliers in the innovations and outliers giving additive effects, see e.g., Denby & Martin (1979). By introducing the model

$$z(t) = y(t) + v(t), \tag{7.2}$$

where $z(t)$ is the observation and $y(t)$ is the process value, models for the situations without outliers, with innovation outliers and with additive outliers, respectively, can be described as in the following paragraphs.

NO OUTLIERS

There are no outliers when the actual innovation density function matches the assumed density, and the process is observed perfectly. For the traditional assumption of normal distributed innovations this means

$$\begin{aligned} e(t) &\in NIID(0;\sigma_e^2) \\ v(t) &\equiv 0. \end{aligned}$$

INNOVATIONS OUTLIERS

Innovation outliers occur when the density function of the innovations is different from the assumed density. Typically, the actual density function is one with heavier tails than of the assumed density. Thus, for the assumption of normal distributed innovations the following contaminated normal density gives rise to innovation outliers

$$\begin{aligned} e(t) &\in (1-\gamma)NIID(0;\sigma_{e_1}^2) + \gamma NIID(0;\sigma_{e_2}^2) \\ v(t) &\equiv 0, \end{aligned}$$

where γ typically is in the interval [0;0.25], and $\sigma_{e_2}^2$ is considerably bigger than $\sigma_{e_1}^2$. Actually the contaminated normal density is not heavy-tailed, but it is still an outlier generating density. This model for innovation outliers, which in the following will be referred to as $\mathrm{IO}(\gamma, \sigma_{e_1}, \sigma_{e_2})$, is used

in a succeeding simulation experiment. Examples of heavy-tailed densities, which also are used for modelling innovation outliers, are the Cauchy density

$$f(x) = \frac{\beta}{\pi} \frac{1}{\beta^2 + (x-\alpha)^2}, \qquad x \in \Re \tag{7.3}$$

or the Laplace density

$$f(x) = \frac{1}{2\beta} \exp\left[-\frac{1}{\beta}|x-\alpha|\right], \qquad x \in \Re. \tag{7.4}$$

ADDITIVE OUTLIERS

Additive outliers occur when there is a difference between the actual process values and the corresponding observations. The character of such additive effects depends on the source of the effects causing the deviations. They can, for instance, be modelled as

$$\begin{aligned} e(t) &\in NIID(0;\sigma_e^2) \\ v(t) &\in (1-\gamma)\delta(0) + \gamma NIID(0;\sigma_v^2), \end{aligned}$$

where $\delta(0)$ is a degenerate central component in zero. This means that the additive effects $\{v(t)\}$ occur with probability γ, and when appearing, they are independent and identically distributed normal random variables with mean zero and variance σ_v^2. This additive outlier model is referred to as $\mathrm{AO}(\gamma, \sigma_e, \sigma_v)$, and is used in the succeeding simulation experiments.

In accordance with the discussion above on erroneous observations, outliers might be correlated. This may be introduced in the outlier models by allowing a dependency structure in $\{e(t)\}$ for innovation outliers and in $\{v(t)\}$ for additive outliers.

The implication of the two outlier types on the set of data is clearly different. This can be illustrated, like in Martin (1981), by simulation of an

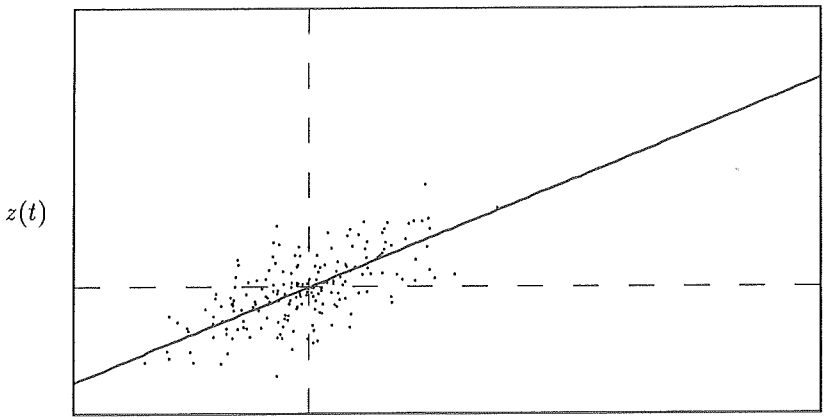


Figure 7.1: Observations simulated with an AR(1) model without outliers.

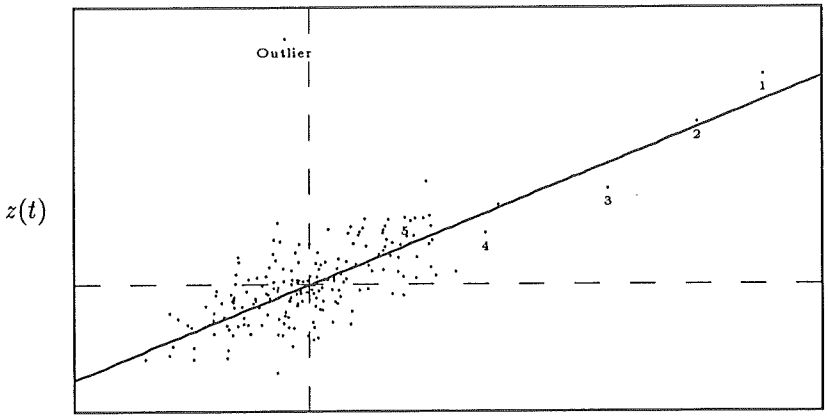


Figure 7.2: Observations simulated with an AR(1) model and a single innovation outlier. The observations succeeding the outlier are numbered.

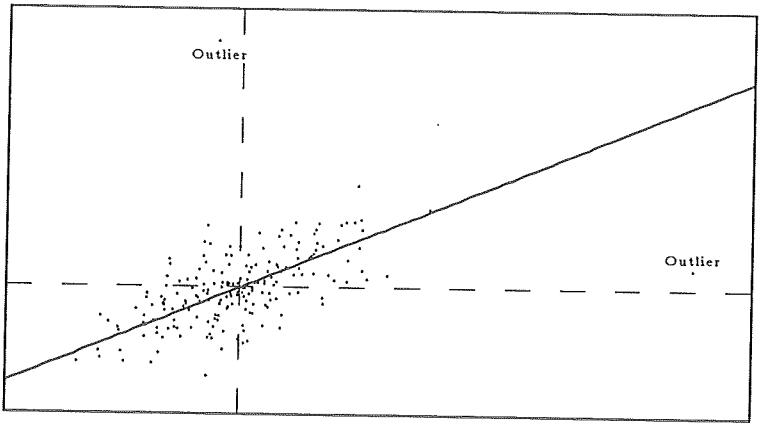


Figure 7.3: Observations simulated with an AR(1)-model and a single additive outlier.

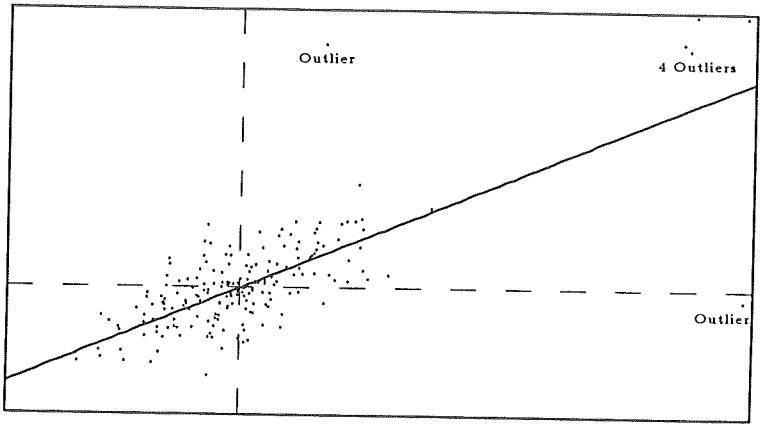


Figure 7.4: Observations simulated with an AR(1)-model and a group of additive outliers.

AR(1) model with different kinds of outliers present. Figure 7.1 shows a set of observations simulated with the model

$$y(t) = 0.75y(t-1) + e(t) \tag{7.5}$$

with $e(t) \in NIID(0;1)$. The slope of the line shown is equal to the AR parameter. In this figure and in the following three figures is the range of the abscissa and the ordinate axis the same (the scales are different).

In Figure 7.2 the result of a simulation where a single innovation has been interchanged with an outlier is shown (the innovation is added 10). The consequence of this outlier is clearly that the single outlying observation is far from the autoregression line. However, the subsequent observations (numbered) are consistent with the model, and due to the excitation of the large innovation they imply an increased precision in the estimation. This means that an innovation outlier influences the variance of the parameter estimate in both directions. In fact, Martin & Yohai (1985) point out, what they term as the distribution-free property of the LS estimator, i.e. the LS estimate of the parameters in A and C is central and has an asymptotic covariance matrix which is independent of the density of the innovations as long as the density is symmetric and the variance exists. On the other hand, Martin & Yohai (1985) show that the LS estimate of the mean of $y(t)$ (the average of the observations) is proportional to the variance of the innovations, thus lacking the distribution-free property.

Figure 7.3 and 7.4 show the same observations as in Figure 7.1 but with a single and a group of six successive observations, respectively, superposed an additive outlier equalling 10. It seems to be clear that such outliers do not imply an excitation of the process, and therefore such outliers do not imply an increase in precision. On the contrary the result is bias of the estimates. In the depicted single-outlier case the bias is obviously towards zero. Similarly, two of the outlying observations in the patchy outlier situation also give rise to bias towards zero, whereas the four outlying observations in between pull the estimate towards +1. In the general case the bias depends on the parameters of the underlying model as well as on the character of the additive effects causing the outliers.

If outliers coming from the sources of errors, discussed previously for load data in DH systems, should be classified to be either IO or AO, clearly malfunction of measurement equipment implies additive outliers to be present. Deviations from the assumed density of the noise components correspond to innovation outliers. Errors in the observations corresponding to situations where the process is not in an normal operation condition can be classified as being additive effects outliers. This is so for the part of the model inadequacy which can be described by additive model components. However, it is obvious that non-normal operation also implies an excitation of the system which evolves forward into the succeeding normal operation situation. This characteristic corresponds to that of innovation outliers.

It can be concluded that outliers of both types are expected to occur in the practical implementation of an on-line estimation of a load model. Therefore it is required that recursive algorithms, designed for handling the presence of both outlier types, are applied in a practical implementation.

7.3 ROBUST ESTIMATION

Solutions to the robustness issue for dynamic models have mainly appeared as off-line estimation methods. Due to the nonlinear nature of the robust estimation problem, these methods turn out to be implemented as iterative algorithms. Such off-line techniques are discussed in, e.g., Martin & Yohai (1985), Huber (1981) and Hampel, Ronchetti, Rousseuw, & Stahel (1986). Recently Allende & Heiler (1992) have proposed a multi-stage procedure for robust estimation of ARMA processes in the presence of additive outliers. In the first stage estimates of the innovations are found as the residuals of a robust autoregressive fit of high order. These residuals are then cleaned through a weight function, and subsequently an estimation, based on robust M-weighting of prediction errors supplied with bounds on the influence of regressors (generalized M-estimation), is carried out repeatedly until the parameter estimates have stabilized. For each step the prediction errors found in the preceding step are inserted as innovation estimates in the MA

part of the regressor vector.

The fact that such methods utilize the whole batch of data implies that all past data must be stored in order to obtain new robust estimates when also the effect of additional observations is to be included. Hence, apart from being a requisite for adaptive estimation, recursive algorithms also have considerable computational advantages compared to off-line techniques.

In the first place some robustness concepts, which are referred to later in this chapter, are outlined. Only a cursory account of the concepts are given. Refer to Martin & Yohai (1985) for a more stringent account.

7.3.1 ROBUSTNESS CONCEPTS

Let the estimate of the parameters in (7.1) be $\hat{\theta}_N = T_N(z(1), \ldots, z(N))$. This estimate is termed *efficiency robust* if it has high efficiency both at an assumed nominal distribution P_θ^n of the observations $z^n = (z(1), \ldots, z(n))$ and at deviating distributions which in some sense is close to the nominal distribution. The quintessence of efficiency robustness is that the estimate should not collapse, w.r.t. variance, when minor deviations from the distribution assumption are present. Having in mind that such deviations are always likely to be present in a practical setting, the efficiency robustness property is an important and necessary quality of an estimator.

In Hampel (1971) the concept of *qualitative robustness* for independent and identically distributed observations is introduced. Martin (1981) gives the following interpretation of this concept:

> Small changes in the distribution of data should produce only small changes in the distribution of the estimate, uniformly in sample size.

Martin & Yohai (1985) carries this over to the more general case of a proba-

bility model for the observations without special assumptions (observations may be correlated) by defining *qualitatively π-robustness*. They state the following description of this concept:

> The basic idea is to require that the estimate changes by only a small amount when the sample is changed by replacing a small fraction of observations by arbitrarily large outliers or by perturbing all the observations with small errors.

The *min-max robustness* concept is defined by the *min-max robust* estimate being the solution to the problem

$$\inf_{\theta \in \Theta} \sup_{P^\infty \in \boldsymbol{P}^\infty} V(\theta, P^\infty), \tag{7.6}$$

where $V(\theta, P^\infty)$ is the asymptotic variance of the estimate of θ at the distribution P^∞. Θ is a family of estimates, and $\boldsymbol{P}^\infty$ is a family of distributions for the process.

7.3.2 M-ESTIMATION

Huber (1964) introduced the M-estimator, given by

$$\hat{\mu} = \arg\min_{\mu} \sum_{i=1}^{N} \rho_c(z(i) - \mu), \tag{7.7}$$

for robust estimation of the location parameter μ, in the setting where $z(1), \ldots, z(N)$ are independent and identically distributed random variables. The application of this estimator is useful in situations where the distribution, $F(z)$, for $z(t)$ is only approximately known (Huber, 1964). For the min-max robustness concept, and assuming that F ranges over the set of all distributions $F = (1 - \epsilon)\Phi + \epsilon H$ for fixed ϵ and symmetric H,

Huber (1964) shows that the most robust of the M-estimators is uniquely determined and is obtained with

$$\rho_c(u) = \begin{cases} \frac{1}{2}u^2 & , \text{ for } |u| \leq c \\ c|u| - \frac{1}{2}c^2 & , \text{ for } |u| > c \end{cases} . \tag{7.8}$$

Φ is the standard normal cumulative and H is the unknown contaminating distribution. The M-estimator with this weight function is actually the ML estimator for a unique least favourable distribution F_0 with density

$$f_0(z) = (1-\epsilon)\frac{1}{\sqrt{2\pi}}\exp[-\rho_c(z)], \tag{7.9}$$

that is, a normal density for $z < c$ and an exponential density for $z \geq c$. When choosing c efficiency considerations, as described later in this chapter, can be used.

For the case where the scale, σ, of the normal component and the amount ϵ of contamination in the distribution F are unknown, Huber (1964) has three proposals for estimation of the location parameter. He recommends to use the second one, Proposal 2, in which the estimates of μ and σ are found as the solutions to the following equation pair

$$\frac{1}{N}\sum_{i=1}^{N}\psi_c\left(\frac{z(i)-\mu}{\sigma}\right) = 0 \tag{7.10}$$

$$\frac{1}{N}\sum_{i=1}^{N}\psi_c^2\left(\frac{z(i)-\mu}{\sigma}\right) = E_\Phi\psi_c^2(u). \tag{7.11}$$

$\psi_c(\cdot)$ is the derivative of $\rho_c(\cdot)$, i.e. $\psi_c(u) = \max(\min(c,u),-c)$. It is seen that (7.10) is the solution to (7.7) when σ is known. The term on the right-hand side of (7.11) implies that when innovations are normal distributed the scale estimate will be consistent.

In Huber (1973) it is demonstrated how the M-estimation is carried over to estimation in a linear model

$$y(t) = x^{\mathrm{T}}(t)\theta + e(t), \tag{7.12}$$

where the regressor, $x(t)$, consists of known explanatory variables and $e(t)$ are independent noise components. The M-estimate of θ is found by minimizing the weighted modelling errors, i.e.

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{N} \rho_c \left(\frac{z(i) - x^{\mathrm{T}}(i)\theta}{\sigma} \right) \sigma^2, \tag{7.13}$$

and for unknown variance of $e(t)$, the simultaneous estimate is found as the solution to

$$\frac{1}{N} \sum_{i} \psi_c \left(\frac{z(i) - x^{\mathrm{T}}(i)\theta}{\sigma} \right) x(i) = 0 \tag{7.14}$$

$$\frac{1}{N-(p+q)} \sum_{i} \psi_c^2 \left(\frac{z(i) - x^{\mathrm{T}}(i)\theta}{\sigma} \right) = E_{\Phi} \psi_c^2(u). \tag{7.15}$$

When the observations $y(1), \ldots, y(T)$ correspond to a stationary and invertible ARMA(p, q) with mean μ and zero mean i.i.d. innovations

$$A(q^{-1})(y(t) - \mu) = C(q^{-1})e(t), \tag{7.16}$$

the residuals are given by

$$\varepsilon_{\theta}(t) = C^{-1}(q^{-1})A(q^{-1})(y(t) - \mu). \tag{7.17}$$

The parameter θ contains the mean value μ and the parameters in $A(q^{-1})$ and $C(q^{-1})$. The residuals may be computed recursively for given initial values of $y(t)$ and $\varepsilon_{\theta}(t)$, and the M-estimate of θ may be found by taking the derivative of $\sum_{t=1}^{T} \rho_c(\varepsilon_{\theta}(t)/\sigma_e)$, where σ_e is the standard deviation of the innovations. Martin & Yohai (1985) give the following solution

$$0 = \sum_{t=p+1}^{T} \psi_c \left(\frac{\varepsilon_{\theta}(t)}{\sigma} \right) r_{\theta}(t-j), \quad j = 1, \ldots, p. \tag{7.18}$$

$$0 = \sum_{t=p+1}^{T} \psi_c \left(\frac{\varepsilon_{\theta}(t)}{\sigma} \right) s_{\theta}(t-j), \quad j = 1, \ldots, q. \tag{7.19}$$

$$0 = \sum_{t=p+1}^{T} \psi_c \left(\frac{\varepsilon_{\theta}(t)}{\sigma} \right), \tag{7.20}$$

where

$$r_\theta(t) = C^{-1}\left(q^{-1}\right)\left(z(t)-\mu\right) = A^{-1}\left(q^{-1}\right)\varepsilon_\theta(t) \quad (7.21)$$

$$s_\theta(t) = C^{-2}\left(q^{-1}\right)A\left(q^{-1}\right)\left(z(t)-\mu\right) = C^{-1}\left(q^{-1}\right)\varepsilon_\theta(t). \quad (7.22)$$

The simultaneous estimate of the scale can be found using (7.15).

In Section 7.2 the distribution-free property of the LS estimates implying that the asymptotic covariance matrix of the estimates of the parameters in $A(q^{-1})$ and $C(q^{-1})$ are not influenced by the presence of innovation outliers was mentioned. However, when innovation outliers, are present the LS estimates can be inefficient compared to the asymptotically efficient ML estimate (for known noise distribution). Martin & Yohai (1985) show, using the ratio of the trace of V_{ML} (the asymptotic covariance matrix) to the trace of V_{LS} as a measure of asymptotic efficiency of the LS estimate, that the asymptotic efficiency can be arbitrarily small for any ϵ-contaminated neighbourhood of the $N(0;\sigma^2)$ distribution. The lacking efficiency of the LS estimate can be explained to be the consequence of not fully utilizing the increased precision which can be obtained when the distribution of the innovations is heavy-tailed.

As the underlying distribution in general is unknown the application of an M-estimator, having large efficiency at heavy-tailed distributions, which in some sense is close to the normal distribution, is appropriate. The asymptotic efficiency for a given choice of weight function $\rho(u)$, $\frac{d}{du}\rho(u) = \psi(u)$, can be shown to be (Martin & Yohai, 1985)

$$EFF(\psi_c, F) = \frac{a(\Psi, F)}{a(\psi, F)} \quad (7.23)$$

where

$$a(\Psi, F) = \frac{E_F\Psi^2(u)}{E_F^2\Psi'(u)\cdot Var(F)}. \quad (7.24)$$

$F(u)$ is the distribution function for the innovations, $f(u)$ is the corresponding density and $\Psi(u) = -f'(u)/f(u)$. $a(\psi_c, F)$ is obtained by inserting $\psi_c(u) = \rho_c'(u)$ in (7.24).

7.3.3 GENERAL M-ESTIMATION

Despite the possibility of obtaining efficiency robustness with an M-estimator the estimates of the parameters in an ARMA-model will not be qualitatively robust. This can be seen from (7.18)-(7.22) by noting that even a small fraction of large outliers can have large effect on the estimates through r_θ and s_θ. To mend this shortcoming of the M-estimator the *general* M-estimators (GM) has emerged. The essential idea in these estimators is to introduce bounds on the influence of single observations. For this reason this type of estimator is also termed *bounded-influence* estimators.

For the linear regression model (7.12) Krasker & Welsch (1982) state that the M-estimation given by (7.14)-(7.15) corresponds to the following minimization problem

$$(\hat{\theta}, \hat{\sigma}) = \underset{\theta,\sigma}{\arg\min} \left(d\sigma + \sum_{i=1}^{N} \sigma \rho_c \left(\frac{z(i) - \boldsymbol{x}^{\mathrm{T}}(i)\theta}{\sigma} \right) \right), \tag{7.25}$$

where d is a necessary constant for the scale estimate to be consistent, and that the *influence function* for a single observation is

$$\Omega(z, \boldsymbol{x}) = \psi_c \left(\frac{z - \boldsymbol{x}^{\mathrm{T}}\theta}{\sigma} \right) \boldsymbol{B}^{-1}\boldsymbol{x}, \tag{7.26}$$

where $\boldsymbol{B}$ is a certain matrix specific for the batch of data. It is seen that the influence of the modelling error is bounded by the ψ_c-function. However, the regressor vector, $\boldsymbol{x}$, enters the influence function directly showing that the influence of errors in the regressors are not down-scaled in any way. This means that for pure AR models additive outliers enter the influence function directly as they will pass through the regressor vector, whereas the influence of innovation outliers will be bounded for the reason that they only enter Ω through ψ_c as modelling error. Of course, innovation outliers appear as large values in the $\boldsymbol{x}$-vector, but these large values are valuable since they are consistent with the system and offer high precision in the estimation.

For models with MA structure the prediction error also appears in the regressor vector meaning that the influence of innovation outliers, in the very sample where they occur, will be unbounded.

The idea of the bounded-influence estimation is to introduce bounds on the influence of the regressor vector in the criterion to be minimized. Krasker & Welsch (1982) consider the problem of estimating the parameters θ of the linear regression model with i.i.d. innovations as efficiently as possible subject to a bound on the sensitivity of the estimator. Their sensitivity measure of the estimator is

$$\gamma = \sup_{z,x} \sqrt{\Omega^{\mathrm{T}}(z,x) V^{-1} \Omega(z,x)} \tag{7.27}$$

with V being the asymptotic covariance matrix of the estimates, given by $V = E[\Omega(z,x)\Omega^{\mathrm{T}}(z,x)]$. They consider estimators of the form

$$0 = \sum_{i=1}^{N} w\left(z(i), x(i), \theta\right)\left(z(i) - x^{\mathrm{T}}(i)\theta\right) x(i), \tag{7.28}$$

where the weight function, w, is nonnegative, bounded and continuous.

They propose as estimator of $(\theta_N, \sigma_N, A_N)$ the solutions of the following equations, where A_N is the dispersion of the regressors

$$0 = \frac{1}{n}\sum_{i=1}^{N} \psi_c\left(\sqrt{x^{\mathrm{T}}(i) A_N^{-1} x(i)}\, \frac{\varepsilon_{\theta_N}(i)}{\sigma_N}\right) \frac{x(i)}{\sqrt{x^{\mathrm{T}}(i) A_N^{-1} x(i)}} \sigma_N, \tag{7.29}$$

$$0 = \frac{1}{n}\sum_{i=1}^{N} \chi\left(\frac{\varepsilon_{\theta_N}(i)}{\sigma_N}\right), \tag{7.30}$$

$$A_N = \frac{1}{n}\sum_{i=1}^{N} g_1\left(\frac{c}{\sqrt{x^{\mathrm{T}}(i) A_N^{-1} x(i)}}\right) x(i) x^{\mathrm{T}}(i). \tag{7.31}$$

$\chi(u)$ is a symmetric weight function, previously set to $\psi_c^2(u) - E_\Psi \phi_c^2(v)$, and the weight function in the regressor covariance estimation is given by

$$g_1(u) = E \min\left(z^2, u^2\right), \qquad z \sim N(0;1). \tag{7.32}$$

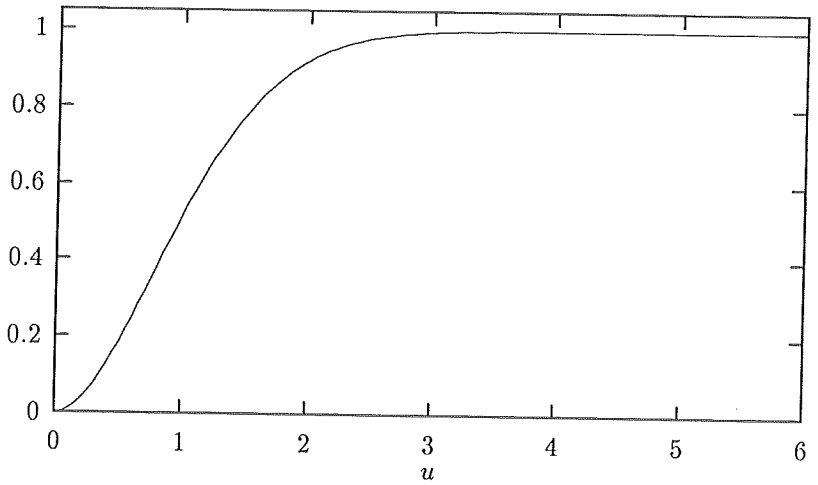


Figure 7.5: The function $g_1(u)$.

This function is depicted in Figure 7.5.

The idea of this proposal is to obtain a robust estimate $\boldsymbol{A}_N$ of the covariance of the regressors by the implicit expression (7.31). $\boldsymbol{A}_N$ is used in (7.29) to measure each of the regressors, and this measure is used to rescale the corresponding prediction error in relation to c. This actually means that when a regressor is deviating from the rest, the influence of the corresponding observation in the estimation of $\boldsymbol{\theta}$ is down-weighted. (7.30) corresponds to the previous scale estimators.

Krasker & Welsch (1982) analyze the appropriate choice of the boundary parameter c with respect to asymptotic efficiency compared to the LS estimator for normal distributed error structure (relative efficiency). To obtain a relative efficiency of, e.g., 0.95 or 0.99, they propose to choose c as 1.596 or 2.093 times $\sqrt{\dim \boldsymbol{x}(i)}$, respectively. This, however, cannot be carried over directly to the situation of dependent sequences and recursive estimation.

7.4 RECURSIVE ROBUST ESTIMATION

Robust estimation based on stochastic approximation is studied in Martin & Masreliez (1975), and used in Campbell (1982) for M-estimation of the parameters in an AR process. Recursive robust estimation of static systems is discussed by Poljak & Tsypkin (1980). In Kuh & Samarov (1986) recursive robust estimation is used in connection with detection of shifts in a regression. Masreliez (1975) makes a certain robustification of the Kalman filter, and his results are used by West (1981) in a study of sequential estimation of a location vector in linear regression.

Recursive M-estimators of location and scale are treated in the papers by Englund, Holst, & Ruppert (1988, 1989). The results are mostly theoretical concerning the asymptotic behaviour of the estimators. In Englund (1991) linear regression models are studied, and, by using stochastic approximation, recursive algorithms, based on the bounded-influence estimator in Krasker & Welsch (1982), are proposed. Poulsen & Holst (1982) study recursive robust estimation of parameters and scale in adaptive control systems. They use a recursive version of the Huber Criterion for the parameter estimation, and show that it is possible to separate the effects of outliers from systematic changes in the system description. McDougall (1994) investigate robustified version of the recursive maximum likelihood algorithm aimed at minimization of the effect of additive outliers.

In the following sections different approaches for recursive robust estimation are considered. First of all a general method for obtaining recursive robust parameter estimation algorithms is used to derive two different recursive robust estimation algorithms. The method originates from a formulation of the estimation problem as a recursive minimization of a criterion with respect to the parameters. It is a generalization of a method discussed in Sejling (1987), and it corresponds to one of the proposals for derivation of recursive algorithms in Ljung & Söderström (1983).

The first algorithm is based on the M-estimator proposed by Huber (1973). Cipra (1992) uses a different approach for the derivation of recursive M-

estimation, in which approximations in the off-line estimator is introduced allowing for the formulation of the algorithm as an approximative recursive solution to a set of normal equations. The third algorithm derived is based on the bounded-influence estimator by Krasker & Welsch (1982). These three algorithms are compared both to the RLS algorithm, see Section 6.1, and a modified RLS algorithm, in which observations are treated as missing if the corresponding prediction errors exceed a specified bound. This idea is applied for off-line estimation in Ljung (1993).

All the algorithms will be derived and presented with a forgetting factor to obtain adaptivity in the estimation. The derivation is for all algorithms based on the assumption that the model is linear in the parameters, which means that the algorithms are not derived for models with MA components. Hence the model can be written as

$$y(t) = x^{\mathrm{T}}(t)\theta + e(t). \tag{7.33}$$

The innovations are assumed to be zero mean independent random variables, and it is assumed that $x(t)$ and $e(t)$ are mutually independent.

7.4.1 RECURSIVE M-ESTIMATION

In this section two approaches for derivation of recursive M-estimation is applied. The objective is to determine the parameter estimate, $\hat{\theta}(t)$, that minimize the criterion of Huber (1964, 1973)

$$V_{\mathrm{Hu}}^{\lambda}(t,\theta) = \sum_{i=1}^{t} \lambda^{t-i} \rho_c\left(\frac{\varepsilon_\theta(i)}{\sigma}\right)\sigma^2 \tag{7.34}$$

with $\rho_c(u)$ defined in (7.8).

Both of the two approaches can be considered as counterparts to methods which are used for derivation of the RLS algorithm.

1. RLS can be derived as the exact recursive minimization of a quadratic criterion function. In Section 7.4.1.1 it is shown how this is carried

over to the recursive minimization of the non-quadratic M-criterion function (7.34). This requires introduction of some approximations, which indeed are justifiable when the parameter estimates are close their optimum. These approximations and the evolution of the criterion between consecutive steps are used to obtain update expressions for the parameter estimate and the covariance estimate.

2. Another way of obtaining RLS is to apply a recursive formulation of the explicit LS estimator (the Normal Equations). Cipra (1992) shows how approximations can be introduced in the off-line M-estimator allowing for a formulation of the algorithm as a recursive solution to a set of normal equations. The derivation of Cipra (1992) is outlined in Section 7.4.1.2.

Naturally these approaches can also be applied for the derivation of recursive M-estimation with other weight functions than $\rho_c(u)$ in (7.8).

7.4.1.1 RECURSIVE M-ESTIMATION VIA RECURSIVE MINIMIZATION OF THE M-CRITERION

In this section the algorithm, implementing a recursive approximative minimization of the M-criterion of Huber (7.34), is derived. It is assumed that the scale parameter, σ, is known. In Section 7.4.3 it is shown how a recursive estimation of the scale parameter is obtained from Huber's Proposal 2 (7.11).

Let $\hat{\theta}(t-1)$ be the estimate at time $t-1$. Then the result of applying the algorithm on an observation at time t is supposed to be the estimate $\hat{\theta}(t)$ implementing an approximative minimization of the criterion $V_{\text{Hu}}^{\lambda}(t,\theta)$ (7.34).

A Taylor expansion of the criterion around $\hat{\theta}(t-1)$ gives

$$\begin{aligned} V_{\text{Hu}}^{\lambda}(t,\theta) &= V_{\text{Hu}}^{\lambda}\Big(t,\hat{\theta}(t-1)\Big) + \Big[\theta-\hat{\theta}(t-1)\Big]^{\text{T}} \nabla_{\theta} V_{\text{Hu}}^{\lambda}(t,\theta)_{|\hat{\theta}(t-1)} + \\ &\quad \tfrac{1}{2}\Big[\theta-\hat{\theta}(t-1)\Big]^{\text{T}} \nabla_{\theta}\nabla_{\theta} V_{\text{Hu}}^{\lambda}(t,\theta)_{|\hat{\theta}(t-1)} \Big[\theta-\hat{\theta}(t-1)\Big] + \\ &\quad \text{o}\Big(|\theta-\hat{\theta}(t-1)|^2\Big), \end{aligned} \tag{7.35}$$

where ∇_{θ} denotes differentiation with respect to θ, and $\text{o}(x)$ is a function for which $\text{o}(x)/|x| \to 0$ as $|x| \to 0$. Taking the derivative of this expression with respect to θ, and applying the assumption of the gradient of the criterion being zero in every time step when evaluated in the recursively obtained parameter estimates, i.e,

$$\forall t : \nabla_{\theta} V_{\text{Hu}}^{\lambda}(t,\theta)_{|\hat{\theta}(t)} = 0, \tag{7.36}$$

lead to the following parameter update

$$\begin{aligned} \hat{\theta}(t) &= \hat{\theta}(t-1) - \Big[\nabla_{\theta}\nabla_{\theta} V_{\text{Hu}}^{\lambda}(t,\theta)_{|\hat{\theta}(t-1)}\Big]^{-1} \nabla_{\theta} V_{\text{Hu}}^{\lambda}(t,\theta)_{|\hat{\theta}(t-1)} + \\ &\quad \text{o}\Big(|\hat{\theta}(t)-\hat{\theta}(t-1)|\Big). \end{aligned} \tag{7.37}$$

Taking the derivative of the criterion (7.34) written as a recursion,

$$V_{\text{Hu}}^{\lambda}(t,\theta) = \lambda V_{\text{Hu}}^{\lambda}(t-1,\theta) + \rho_c\left(\frac{\varepsilon_{\theta}(t)}{\sigma}\right)\sigma^2, \tag{7.38}$$

and using the assumption (7.36) at $t-1$ give

$$\nabla_{\theta} V_{\text{Hu}}^{\lambda}(t,\theta)_{|\hat{\theta}(t-1)} = -x(t)\psi_c\left(\frac{\varepsilon_{\hat{\theta}(t-1)}(t)}{\sigma}\right)\sigma. \tag{7.39}$$

Evaluation of the two times derivative of (7.38) in $\hat{\theta}(t-1)$ gives

$$\begin{aligned} \nabla_{\theta}\nabla_{\theta} V_{\text{Hu}}^{\lambda}(t,\theta)_{|\hat{\theta}(t-1)} &= \lambda \nabla_{\theta}\nabla_{\theta} V_{\text{Hu}}^{\lambda}(t-1,\theta)_{|\hat{\theta}(t-1)} + \\ &\quad x(t)x^{\text{T}}(t)\psi_c'\left(\frac{\varepsilon_{\hat{\theta}(t-1)}(t)}{\sigma}\right) \end{aligned} \tag{7.40}$$

with $\psi_c'(u) = \frac{d}{du}\psi_c(u)$.

If it is assumed that $\hat{\theta}(t)$ is close to $\hat{\theta}(t-1)$ the following two approximations can be justified

- ▷ The term $o\Big(|\hat{\theta}(t)-\hat{\theta}(t-1)|\Big)$ in (7.37) can be neglected.

- ▷ $\nabla_\theta \nabla_\theta V_{\text{Hu}}^\lambda(t-1,\theta)_{|\hat{\theta}(t-1)}$ can be interchanged by $\nabla_\theta \nabla_\theta V_{\text{Hu}}^\lambda(t-1,\theta)_{|\hat{\theta}(t-2)}$ in (7.40).

For a quadratic criterion the difference in the first approximation vanishes, and the second order derivative is independent of θ. This means that if the criterion is approximately quadratic at the location of the parameter estimate the term in the first approximation can reasonably be neglected, and the second order derivative of the criterion is close to being independent of the estimate. Furthermore, when the algorithm has been applied for a reasonable number of recursions the change of the estimate in consecutive steps is likely to be small, if the forgetting factor λ is not chosen too small. Hence, this is also a reason for the justification of the approximations. The questionable thing about the validity of the approximations is clearly in the initial phase of the algorithm application, where steps of substantial size of the parameter estimate occur, and where the criterion can be far from being quadratic.

Finally, by denoting the inverse of the second order derivative of the criterion as

$$\boldsymbol{P}_\lambda(t) \equiv \Big[\nabla_\theta \nabla_\theta V_{\text{Hu}}^\lambda(t,\theta)_{|\hat{\theta}(t-1)}\Big]^{-1}, \tag{7.41}$$

and using the Matrix Inversion Lemma on (7.40), the algorithm is

RH1(c).....

$$\varepsilon_{\hat{\theta}(t-1)}(t) = z(t) - \boldsymbol{x}^{\mathrm{T}}(t)\hat{\theta}(t-1) \tag{7.42}$$

$$\boldsymbol{P}_\lambda(t) = \frac{\boldsymbol{P}_\lambda(t-1)}{\lambda} - \frac{1}{\lambda}\psi_c'\left(\frac{\varepsilon_{\hat{\theta}(t-1)}(t)}{\sigma}\right)\frac{\boldsymbol{P}_\lambda(t-1)\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{P}_\lambda(t-1)}{\lambda + \boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{P}_\lambda(t-1)\boldsymbol{x}(t)} \tag{7.43}$$

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \boldsymbol{P}_\lambda(t)\boldsymbol{x}(t)\psi_c\left(\frac{\varepsilon_{\hat{\theta}(t-1)}(t)}{\sigma}\right)\sigma. \tag{7.44}$$

.....

Note that the expression in (7.43) has been simplified. Actually $\psi_c'(\cdot)$ appears in the denominator of the fraction on the right-hand side multiplied on $\boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{P}_\lambda(t-1)\boldsymbol{x}(t)$, but since $\psi_c'(u)$ is either zero or one and it appears in front of the fraction, it is redundant in the denominator.

7.4.1.2 RECURSIVE M-ESTIMATION VIA RECURSIVE FORMULATION OF NORMAL EQUATIONS

Recently another approach for derivation of the recursive M-estimation of a parameter-linear model has been proposed by Cipra (1992). His approach is to rewrite Equation (7.14), which solves the M-estimation problem, as

$$\sum_{i=1}^{t} \lambda^{t-i} \sigma \frac{\psi_c\left(\frac{\varepsilon_{\hat{\theta}(t)}(i)}{\sigma}\right)}{\varepsilon_{\hat{\theta}(t)}(i)} x(i)\varepsilon_{\hat{\theta}(t)}(i) = 0, \tag{7.45}$$

where also a forgetting factor has been introduced. This is approximated by the expression

$$\sum_{i=1}^{t} \lambda^{t-i} w_i(t-1) x(i)\varepsilon_{\hat{\theta}(t)}(i) = 0, \tag{7.46}$$

where the weight factor is given by

$$w_i(t-1) = \psi_c\left(\frac{\varepsilon_{\hat{\theta}(t-1)}(i)}{\hat{\sigma}(t-1)}\right) \frac{\hat{\sigma}(t-1)}{\varepsilon_{\hat{\theta}(t-1)}(i)}. \tag{7.47}$$

The approximations consist of using $\hat{\theta}(t-1)$ instead of $\hat{\theta}(t)$ for calculation of the prediction error in the weight factor, and of using a scale parameter estimate $\hat{\sigma}(t-1)$ estimated by a separate method.

Equation (7.46) has the form of a normal equation for the weighted LS method which, due to the nonlinearity, should be solved iteratively, but Cipra (1992) proposes that these are replaced by recursions leading to the algorithm below.

RH2(c) .....

$$\varepsilon_{\hat{\theta}(t-1)}(t) = z(t) - \boldsymbol{x}^{\mathrm{T}}(t)\hat{\theta}(t-1) \tag{7.48}$$

$$w(t-1) = \psi_c\left(\frac{\varepsilon_{\hat{\theta}(t-1)}(t)}{\hat{\sigma}(t-1)}\right)\frac{\hat{\sigma}(t-1)}{\varepsilon_{\hat{\theta}(t-1)}(t)} \tag{7.49}$$

$$\boldsymbol{P}_\lambda(t) = \frac{1}{\lambda}\left[\boldsymbol{P}_\lambda(t-1) - \frac{\boldsymbol{P}_\lambda(t-1)\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{P}_\lambda(t-1)}{\boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{P}_\lambda(t-1)\boldsymbol{x}(t) + \lambda/w(t-1)}\right] \tag{7.50}$$

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{\boldsymbol{P}_\lambda(t-1)\boldsymbol{x}(t)}{\boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{P}_\lambda(t-1)\boldsymbol{x}(t) + \lambda/w(t-1)}\varepsilon_{\hat{\theta}(t-1)}(t) \tag{7.51}$$

.....

7.4.2 RECURSIVE GENERAL M-ESTIMATION

In this section a recursive algorithm based on the off-line estimator by Krasker & Welsch (1982) is derived. In the first place a number of modifications of the original estimator is introduced. Instead of having only one tuning parameter one parameter c, is used in the part of the estimator which measures the prediction error in Equation (7.52), and another tuning parameter, a, is used in the covariance estimation in Equation (7.54). Finally, a forgetting factor has been introduced in the expressions for model and scale parameter estimation.

$$0 = \frac{1}{t}\sum_{i=1}^{t}\lambda^{t-i}\psi_c\left(\sqrt{\boldsymbol{x}^{\mathrm{T}}(i)\hat{\boldsymbol{A}}^{-1}(t)\boldsymbol{x}(i)}\,\frac{\varepsilon_{\hat{\theta}(t)}(i)}{\hat{\sigma}(t)}\right)\frac{\boldsymbol{x}(i)\hat{\sigma}(t)}{\sqrt{\boldsymbol{x}^{\mathrm{T}}(i)\hat{\boldsymbol{A}}^{-1}(t)\boldsymbol{x}(i)}} \tag{7.52}$$

$$0 = \frac{1}{t}\sum_{i=1}^{t}\lambda^{t-i}\chi\left(\frac{\varepsilon_{\hat{\theta}(t)}(i)}{\hat{\sigma}(t)}\right), \tag{7.53}$$

$$\hat{\boldsymbol{A}}(t) = \frac{1}{t}\sum_{i=1}^{t}g_1\left(\frac{a}{\sqrt{\boldsymbol{x}^{\mathrm{T}}(i)\hat{\boldsymbol{A}}^{-1}(t)\boldsymbol{x}(i)}}\right)\boldsymbol{x}(i)\boldsymbol{x}^{\mathrm{T}}(i), \tag{7.54}$$

In the derivation of the recursive algorithm, corresponding to Equa-

tions (7.52)-(7.53), the estimation of the model parameters and the scale parameter are considered separately. That is, when deriving the recursive counterpart of Equation (7.52) σ is assumed to be known, and vice versa for Equation (7.53). The robust estimation of the covariance matrix is as it stands independent of the other parts of the estimator.

The algorithm for estimation of θ is obtained as the recursive and approximative minimization of the criterion

$$V_{\mathrm{KW}}^{\lambda}(t,\theta)=\sum_{i=1}^{t}\lambda^{t-i}\frac{\sigma^2}{x^{\mathrm{T}}(i)\hat{A}^{-1}(t)x(i)}\rho_c\left(\sqrt{x^{\mathrm{T}}(i)\hat{A}^{-1}(t)x(i)}\,\frac{\varepsilon_\theta(i)}{\sigma}\right). \quad (7.55)$$

The recursive minimization of (7.55) is derived using the same procedure as in the derivation of RH1. The first step is to write down a second order Taylor expansion of the criterion and use the assumption that the derivative of the criterion at time t is zero when evaluated in $\hat{\theta}(t)$. As outlined below, however, it will be necessary to introduce a modified criterion, $V_{\mathrm{KW}}^{\lambda,r}(t,\theta)$, to be able to obtain a recursive expression for the criterion. For this reason the second order expansion is applied to the modified criterion leading to the following expression for the parameter update

$$\begin{aligned}\hat{\theta}(t) &= \hat{\theta}(t-1)-\left[\nabla_\theta\nabla_\theta V_{\mathrm{KW}}^{\lambda,r}(t,\theta)_{|\hat{\theta}(t-1)}\right]^{-1}\nabla_\theta V_{\mathrm{KW}}^{\lambda,r}(t,\theta)_{|\hat{\theta}(t-1)}+\\ &\quad o\Big(|\hat{\theta}(t)-\hat{\theta}(t-1)|\Big). \end{aligned} \quad (7.56)$$

To obtain expressions for the first and second order derivative of the criterion in (7.56), the criterion (7.55) must be written as a recursive expression. This requires that the batch covariance estimate $\hat{A}^{-1}(t)$ in the criterion (7.55) is interchanged by the, at time i, available covariance estimate , i.e.

$$V_{\mathrm{KW}}^{\lambda,r}(t,\theta)=\sum_{i=1}^{t}\lambda^{t-i}\frac{\sigma^2}{x^{\mathrm{T}}(i)\hat{A}^{-1}(i)x(i)}\rho_c\left(\sqrt{x^{\mathrm{T}}(i)\hat{A}^{-1}(i)x(i)}\,\frac{\varepsilon_\theta(i)}{\sigma}\right). \quad (7.57)$$

This criterion can be written recursively as

$$V_{\text{KW}}^{\lambda,r}(t,\theta) = \lambda V_{\text{KW}}^{\lambda,r}(t-1,\theta) + \frac{\sigma^2}{x^{\text{T}}(t)\hat{A}^{-1}(t)x(t)}\rho_c\left(\sqrt{x^{\text{T}}(t)\hat{A}^{-1}(t)x(t)}\,\frac{\varepsilon_\theta(t)}{\sigma}\right) \quad (7.58)$$

Computing the gradients in (7.58), and using the assumption of the criterion gradient at time t being zero when evaluated in the parameter estimate at the same time gives

$$\nabla_\theta V_{\text{KW}}^{\lambda,r}(t,\theta)_{|\hat{\theta}(t-1)} = \frac{-x(t)\sigma}{\sqrt{x^{\text{T}}(t)\hat{A}^{-1}(t)x(t)}}\psi_c\left(\sqrt{x^{\text{T}}(t)\hat{A}^{-1}(t)x(t)}\,\frac{\varepsilon_{\hat{\theta}(t-1)}(t)}{\sigma}\right). \quad (7.59)$$

The second order derivative of the criterion expression, evaluated in $\hat{\theta}(t-1)$, becomes

$$\nabla_\theta\nabla_\theta V_{\text{KW}}^{\lambda,r}(t,\theta)_{|\hat{\theta}(t-1)} = \lambda\nabla_\theta\nabla_\theta V_{\text{KW}}^{\lambda,r}(t-1,\theta)_{|\hat{\theta}(t-1)} + x(t)x^{\text{T}}(t)\psi_c'\left(\sqrt{x^{\text{T}}(t)\hat{A}^{-1}(t)x(t)}\,\frac{\varepsilon_{\hat{\theta}(t-1)}(t)}{\sigma}\right) \quad (7.60)$$

As in the derivation of RH1 the following approximations are made

- ▷ The term $o\left(|\hat{\theta}(t)-\hat{\theta}(t-1)|\right)$ in (7.56) can be neglected.
- ▷ $\nabla_\theta\nabla_\theta V_{\text{KW}}^{\lambda,r}(t-1,\theta)_{|\hat{\theta}(t-1)}$ can be interchanged by $\nabla_\theta\nabla_\theta V_{\text{KW}}^{\lambda,r}(t-1,\theta)_{|\hat{\theta}(t-2)}$ in (7.60).

Applying the Matrix Inversion Formula on (7.60) and introducing the notation

$$P_\lambda(t) \equiv \left[\nabla_\theta\nabla_\theta V_{\text{KW}}^{\lambda,r}(t,\theta)_{|\hat{\theta}(t-1)}\right]^{-1} \quad (7.61)$$

give (7.67). Inserting the expression (7.59) in (7.56) gives the final expression of the parameter update (7.68).

To obtain the recursive formula for the covariance estimate, the batch estimate, $\hat{\boldsymbol{A}}_n$, entering g_1 in summand i in (7.31) is interchanged by the estimate which is available at time $i-1$. This gives

$$\hat{\boldsymbol{A}}(t) = \frac{1}{t}\sum_{i=1}^{t} g_1\left(\frac{a}{\sqrt{\boldsymbol{x}^{\mathrm{T}}(i)\hat{\boldsymbol{A}}^{-1}(i-1)\boldsymbol{x}(i)}}\right)\boldsymbol{x}(i)\boldsymbol{x}^{\mathrm{T}}(i), \tag{7.62}$$

which allows the estimator to be written recursively as

$$\hat{\boldsymbol{A}}(t) = \hat{\boldsymbol{A}}(t-1) + \frac{1}{t}\left(g_1\left(\frac{a}{\sqrt{\boldsymbol{x}^{\mathrm{T}}(t)\hat{\boldsymbol{A}}^{-1}(t-1)\boldsymbol{x}(t)}}\right)\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{T}}(t) - \hat{\boldsymbol{A}}(t-1)\right). \tag{7.63}$$

Comparing Equation (7.62) to the batch version Equation (7.54) shows that the difference between the recursive estimator and the batch estimator can be written as

$$\Delta_{\hat{\boldsymbol{A}}(t)} = \frac{1}{t}\sum_{i=1}^{t}\left(g_1^a\left(\hat{\boldsymbol{A}}(i-1), \boldsymbol{x}(i)\right) - g_1^a\left(\hat{\boldsymbol{A}}(t), \boldsymbol{x}(i)\right)\right)\boldsymbol{x}(i)\boldsymbol{x}^{\mathrm{T}}(i), \tag{7.64}$$

where $g_1^a(\boldsymbol{A}, \boldsymbol{x})$ is short for the dependency of g_1 on a, $\boldsymbol{A}$ and $\boldsymbol{x}$.

By use of the Matrix Inversion Formula (7.63) is transformed into (7.65). This completes the derivation of the RKW algorithm.

RKW(c,a).....

$$\hat{\boldsymbol{A}}^{-1}(t) = \frac{t}{t-1}\left(\hat{\boldsymbol{A}}^{-1}(t-1) - g_1\left(\frac{a}{\sqrt{\boldsymbol{x}^{\mathrm{T}}(t)\hat{\boldsymbol{A}}^{-1}(t-1)\boldsymbol{x}(t)}}\right)\times \right.$$

$$\left.\frac{\hat{\boldsymbol{A}}^{-1}(t-1)\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{T}}(t)\hat{\boldsymbol{A}}^{-1}(t-1)}{t+\boldsymbol{x}^{\mathrm{T}}(t)\hat{\boldsymbol{A}}^{-1}(t-1)\boldsymbol{x}(t)g_1\left(\frac{a}{\sqrt{\boldsymbol{x}^{\mathrm{T}}(t)\hat{\boldsymbol{A}}^{-1}(t-1)\boldsymbol{x}(t)}}\right)}\right) \tag{7.65}$$

$$\varepsilon_{\hat{\theta}(t-1)}(t) = z(t) - \boldsymbol{x}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1) \tag{7.66}$$

$$\boldsymbol{P}_\lambda(t) = \frac{\boldsymbol{P}_\lambda(t-1)}{\lambda} -$$

$$\frac{1}{\lambda}\psi_c'\left(\sqrt{\boldsymbol{x}^{\mathrm{T}}(t)\hat{\boldsymbol{A}}^{-1}(t)\boldsymbol{x}(t)}\,\frac{\varepsilon_{\hat{\theta}(t-1)}(t)}{\sigma}\right)\frac{\boldsymbol{P}_\lambda(t-1)\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{P}_\lambda(t-1)}{\lambda+\boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{P}_\lambda(t-1)\boldsymbol{x}(t)} \tag{7.67}$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) +$$

$$\boldsymbol{P}_\lambda(t)\boldsymbol{x}(t)\frac{\sigma}{\sqrt{\boldsymbol{x}^{\mathrm{T}}(t)\hat{\boldsymbol{A}}^{-1}(t)\boldsymbol{x}(t)}}\psi_c\left(\sqrt{\boldsymbol{x}^{\mathrm{T}}(t)\hat{\boldsymbol{A}}^{-1}(t)\boldsymbol{x}(t)}\,\frac{\varepsilon_{\hat{\theta}(t-1)}(t)}{\sigma}\right) \tag{7.68}$$

.....

Note that (7.67) has been simplified by using that $\psi_c'(u)$ is either zero or one.

7.4.3 RECURSIVE ESTIMATION OF SCALE

To develop the recursive scale estimation using a χ-function as proposed by Krasker & Welsch (1982), cf. Equation (7.54), the function

$$X_\lambda(t,\sigma) = \sum_{i=1}^{t} \lambda^{t-i} \chi\left(\frac{\varepsilon_\theta(i)}{\sigma}\right) \tag{7.69}$$

is introduced. In Huber's Proposal 2 the χ-function given by, cf. (7.11),

$$\chi(u) = \psi_c^2(u) - E_\Phi \psi_c^2(u) \tag{7.70}$$

is used. In the derivation of a recursive solution to $\chi_\lambda(t,\sigma) = 0$, the set of model parameters, θ, is assumed to be known.

The function $X_\lambda(t,\sigma)$ can be written recursively as

$$X_\lambda(t,\sigma) = \lambda X_\lambda(t-1,\sigma) + \chi\left(\frac{\varepsilon_\theta(t)}{\sigma}\right). \tag{7.71}$$

By applying a first order Taylor expansion of the X_λ-function around $\hat{\sigma}(t-1)$, and assuming that the scale estimate at time t satisfies the estimator definition, i.e. $X_\lambda(t,\hat{\sigma}(t)) = 0$, give

$$\begin{aligned} 0 \;=\; & X_\lambda\left(t,\hat{\sigma}(t-1)\right) + \left[\hat{\sigma}(t) - \hat{\sigma}(t-1)\right] \nabla_\sigma X_\lambda(t,\sigma)_{|\hat{\sigma}(t-1)} + \\ & o(|\hat{\sigma}(t) - \hat{\sigma}(t-1)|). \end{aligned} \tag{7.72}$$

An expression for the first term on the right-hand side of (7.72) is found by evaluating (7.71) in $\hat{\sigma}(t-1)$, using the assumption $X_\lambda(t-1,\hat{\sigma}(t-1)) = 0$. This gives

$$X_\lambda(t,\hat{\sigma}(t-1)) = \chi\left(\frac{\varepsilon_\theta(t)}{\hat{\sigma}(t-1)}\right). \tag{7.73}$$

Evaluating the derivative of (7.71) in $\hat{\sigma}(t-1)$ gives

$$\nabla_\sigma X_\lambda\left(t,\sigma\right)_{|\hat{\sigma}(t-1)} = \lambda \nabla_\sigma X_\lambda\left(t-1,\sigma\right)_{|\hat{\sigma}(t-1)} + \nabla_\sigma \chi\left(\frac{\varepsilon_\theta(t)}{\sigma}\right)_{|\hat{\sigma}(t-1)}. \tag{7.74}$$

For Huber's Proposal 2 this becomes

$$\nabla_\sigma X_\lambda(t,\sigma)_{|\hat{\sigma}(t-1)} = \lambda \nabla_\sigma X_\lambda(t-1,\sigma)_{|\hat{\sigma}(t-1)} - 2\frac{\varepsilon_\theta^2(t)}{\hat{\sigma}^3(t-1)} I_{\{|\varepsilon_\theta(t)| \le c\hat{\sigma}(t-1)\}} \quad (7.75)$$

Now the following approximations are made

- ▷ The term $o(|\hat{\sigma}(t)-\hat{\sigma}(t-1)|)$ in (7.72) can be neglected.
- ▷ $\nabla_\sigma X_\lambda(t-1,\sigma)_{|\hat{\sigma}(t-1)}$ can be interchanged by $\nabla_\sigma X_\lambda(t-1,\sigma)_{|\hat{\sigma}(t-2)}$ in (7.75).

By introducing $h_\lambda(t) \equiv -\nabla_\sigma X_\lambda(t,\sigma)_{|\hat{\sigma}(t-1)}$ this completes the derivation of the recursive scale estimator using Huber's Proposal 2

.....

$$\varepsilon_\theta(t) = z(t) - x^{\mathrm{T}}(t)\theta \quad (7.76)$$

$$h_\lambda(t) = \lambda h_\lambda(t-1) + 2\frac{\varepsilon_\theta^2(t)}{\hat{\sigma}^3(t-1)} I_{\{|\varepsilon_\theta(t)| \le c\hat{\sigma}(t-1)\}} \quad (7.77)$$

$$\hat{\sigma}(t) = \hat{\sigma}(t-1) + \frac{\chi_c\left(\frac{\varepsilon_\theta(t)}{\hat{\sigma}(t-1)}\right)}{h_\lambda(t)}. \quad (7.78)$$

.....

Due to the symmetry of χ_c the scale estimate can converge to both a positive value and its negative counterpart. To avoid a negative scale estimate the set of permissible values is restricted to the positive real line. In practice this is of significance only in the initial phase when the algorithm is applied.

7.4.4 THE RLS ALGORITHM TREATING CLASSIFIED OUTLIERS AS MISSING OBSERVATIONS

This algorithm is an extension of RLS. Each residual is compared to a recursive estimate of the residual variance. If the numerical value of the residual exceeds c times the variance estimate, neither the estimate of the models parameters, the estimate of the P-matrix nor the residual variance estimate are updated. Hence, the algorithm is

RMO(c) .....

$$k_\sigma(t) = \max\left(\frac{1}{t}, 1-\lambda\right) \tag{7.79}$$

$$\varepsilon_{\hat{\theta}(t-1)}(t) = z(t) - \boldsymbol{x}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1) \tag{7.80}$$

$$\hat{\sigma}^2(t) = \hat{\sigma}^2(t-1) + k_\sigma(t)\left(d_c\varepsilon^2_{\hat{\theta}(t-1)}(t) - \hat{\sigma}^2(t-1)\right) I_{\{|\varepsilon_{\hat{\theta}(t-1)}(t)|<c\hat{\sigma}(t-1)\}} \tag{7.81}$$

$$\boldsymbol{P}_\lambda(t) = \frac{\boldsymbol{P}_\lambda(t-1)}{\lambda} - \frac{1}{\lambda}\frac{\boldsymbol{P}_\lambda(t-1)\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{P}_\lambda(t-1)}{\lambda + \boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{P}_\lambda(t-1)\boldsymbol{x}(t)} I_{\{|\varepsilon_{\hat{\theta}(t-1)}(t)|<c\hat{\sigma}(t-1)\}} \tag{7.82}$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \boldsymbol{P}_\lambda(t)\boldsymbol{x}(t)\varepsilon_{\hat{\theta}(t-1)}(t)\, I_{\{|\varepsilon_{\hat{\theta}(t-1)}(t)|<c\hat{\sigma}(t-1)\}} \tag{7.83}$$

.....

The gain $k_\sigma(t)$ is introduced to enable tracking of a slowly varying scale. Likewise, the scale factor d_c is used to obtain a consistent scale estimate, when the residuals are zero mean normal distributed random variables. This is obtained with

$$d_c^{-1} = E\left\{\phi_c^2(z)\right\}, \quad z \sim \mathrm{N}(0,1) \quad \text{and} \quad \phi_c(u) = \begin{cases} u & , \ |u| \le c \\ 0 & , \ |u| > c \end{cases}. \tag{7.84}$$

This method has the obvious disadvantage that if the estimates are too far from the true parameter values, then the prediction errors will always be too big, and neither the parameters nor the P-matrix will be updated. Furthermore, there is a large risk of accepting outlying observations. A non-recursive treatment of outliers as missing observations after detection is discussed by Ljung (1993).

7.5 SIMULATION AND ESTIMATION RESULTS

In this section the performance of the proposed algorithms is tested on simulated AR processes with different outlier schemes and contamination. In the first part this is done for a first order AR model with constant parameter. Thereafter, the simulation is based on a first order model with a smooth deterministic variation of the AR parameter, and the algorithms are applied with a forgetting factor below one. Finally, the simulation is carried out using a second order AR model with constant parameters.

In all simulations it applies that each data sequence has a length of 3005 observations. When a sequence is simulated with outliers, these will not occur during the first 5 observations. Due to this and the way the estimation algorithms are applied, the observations are numbered from -4 to 3000.

For each model and each outlier scheme 1000 replicates of the data simulation with application of the different algorithms are carried out.

The recursive estimation in all of the algorithms is started by using the RLS algorithm on the first 5 observations. At $t=1$ the algorithm to be studied is initialized with the parameters, $\hat{\theta}(0)_{RLS}$, and covariance matrix, $\boldsymbol{P}(0)_{RLS}$, obtained from RLS. RLS is initialized with the covariance matrix equal to 100 times the unit matrix, and so is the robust covariance estimate of RKW, i.e. $\hat{\boldsymbol{A}}^{-1}(0) = 100\boldsymbol{I}$. The starting value of $\hat{\sigma}^2(0)$ is the known simulation variance of the outlier free innovations, and the starting value of $h(0)$ in the recursive version of Huber's Proposal 2, Equation (7.77), is set to 1.

The three parts of the simulation study are described in the paragraphs below.

I: AR(1) MODEL WITH CONSTANT PARAMETERS

In this part of the study the nominal model is

$$\left(1 - 0.8q^{-1}\right) y(t) = e(t), \tag{7.85}$$

where $\{e(t)\}$ is $NIID(0;1)$ apart from when innovation outliers occur. This model is used for simulation with five different outlier schemes, viz.

1. NO: Without outliers.
2. IO(1,.05,2.5): Innovation outliers with 5% contamination with variance 2.5^2.
3. IO(1,.05,7.5): Innovation outliers with 5% contamination with variance 7.5^2.
4. AO(1,.05,2.5): Additive outliers with 5% contamination with variance 2.5^2.
5. AO(1,.05,7.5): Additive outliers with 5% contamination with variance 7.5^2.

The different characteristics of these noise schemes is illustrated in Figure 7.6. The outlier contamination has variance 7.5^2. With both kinds of outliers (IO or AO), it is possible to observe the existence of outlying observations. Also the different characteristics of the influence on the observations resulting from the different types of outliers can be observed. The innovation outliers affect the process itself and hereby imply a lasting excitation, whereas the additive outliers are additive spikes in the very moment they occur.

For each data sequence RH1(c), RH2(c) and RMO(c) are applied with $c = 1.5, 2, 3, 4$. Likewise RKW(c, a) is applied using the same c-values, and

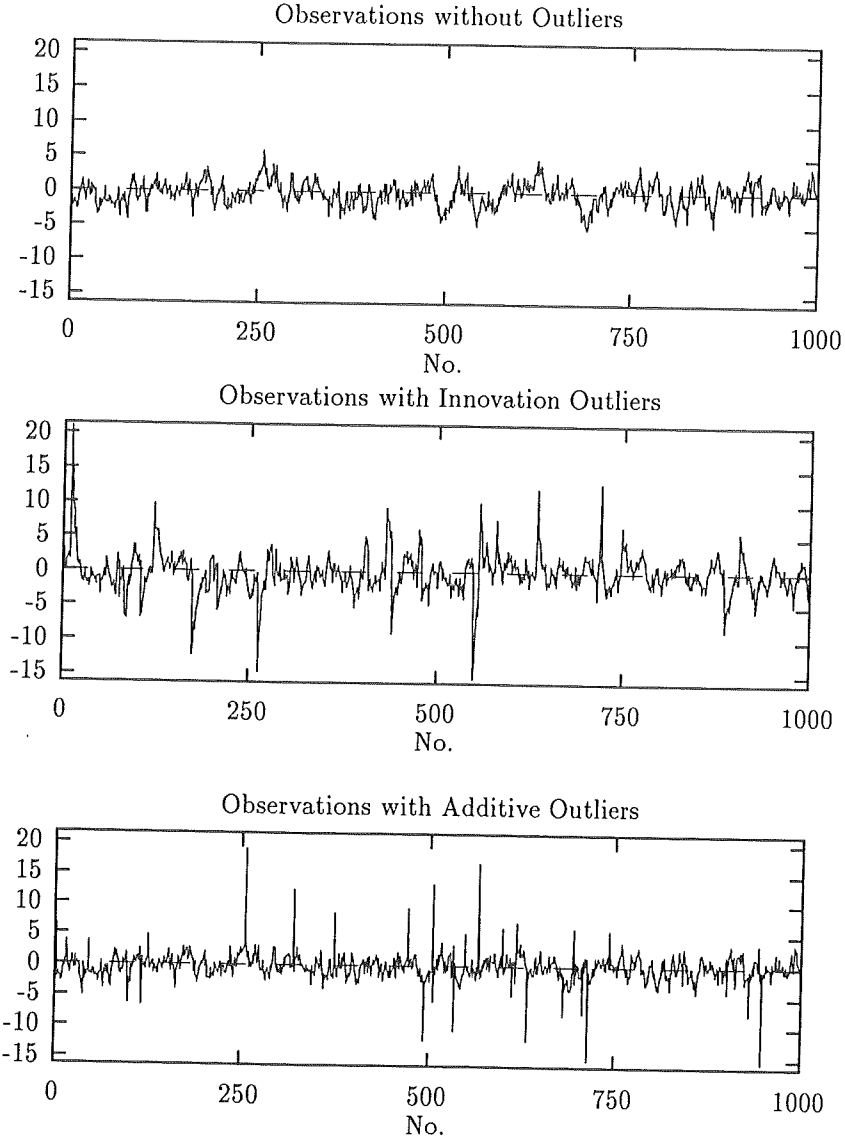


Figure 7.6: Observations simulated with the first order model and using three different outlier types with the same basic noise realization.

for each c with $a = 1.5, 2, 3, 4$. Finally, also RLS is applied. Since the data in this of part is simulated with constant parameter, all algorithms are applied with the forgetting factor $\lambda = 1$.

For comparison the off-line Krasker & Welsch estimator (7.29)-(7.32), denoted KW(c, a), is applied with the combinations (c, a) =(2,3),(3,3) and (3,4). The off-line estimation is based on observations for $t \in (1, 3000)$, and is applied to all three types of data sequences.

The performance of the algorithms is illustrated in trajectory plots in Figures 7.7, 7.8 and 7.9 for observations without outliers, with innovation outliers and with additive outliers, respectively. In the simulations with outlier presence, the contamination is 5 % with variance 7.5^2. Since the trajectories of RH1 and RH2 are almost identical only RH1 is shown. For the trajectories shown the algorithms has been applied with $c = 2$ and $a = 3$. Note that the scales on the parameter axis are different for the three plots. The basic sequence of innovations was the same in all of the estimations shown.

When observations are simulated without outliers RLS, RH1 and RKW are close in performance. With this particular noise realization RKW is closer to the actual parameter than RH1 which in turn is closer than RLS. For another noise realization the order might be different. RMO, however, seems to show a basically different trajectory, the explanation of which seems to lie in the initial phase. Actually this is not surprising as the prediction errors in the beginning possibly can be larger than twice the estimated standard deviation of the innovations in which case the update is omitted. In all cases the estimate seems to converge to the underlying simulation parameter.

Also in the presence of innovation outliers convergence to 0.8 takes place. Recognizing the reduced range on the parameter axis it is clear that the estimates are closer to 0.8 than without outliers. As the basic noise realization is the same the explanation is reasonably ascribed to an increased excitation coming from the outliers.

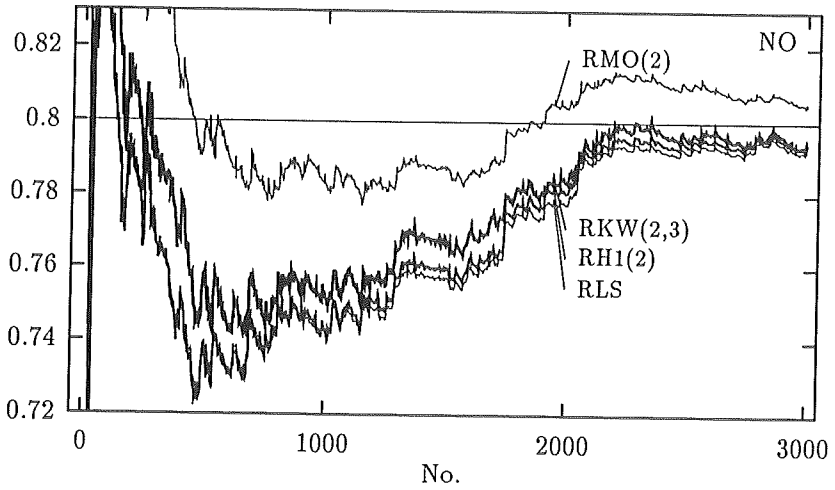


Figure 7.7: Trajectories of the parameter estimates in the first order system without outliers. The basic innovation sequence was the same as in Figure 7.6.

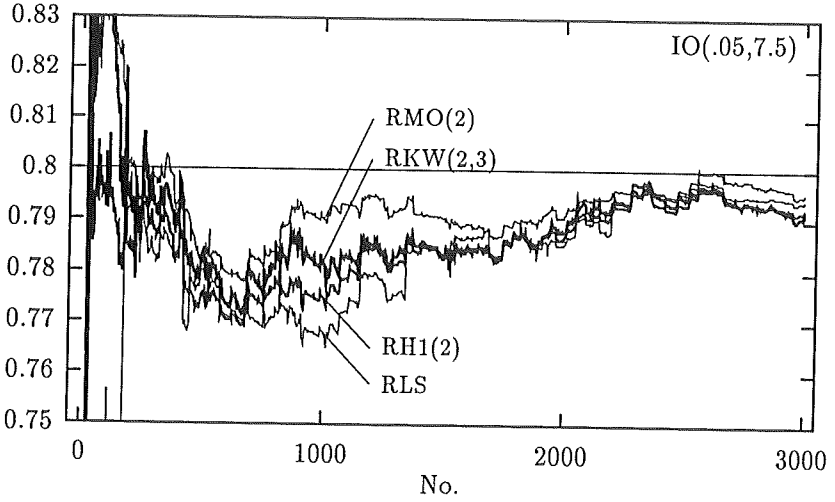


Figure 7.8: Trajectories of the parameter estimates in the first order system with innovation outliers. The basic innovation sequence was the same as in Figure 7.6.

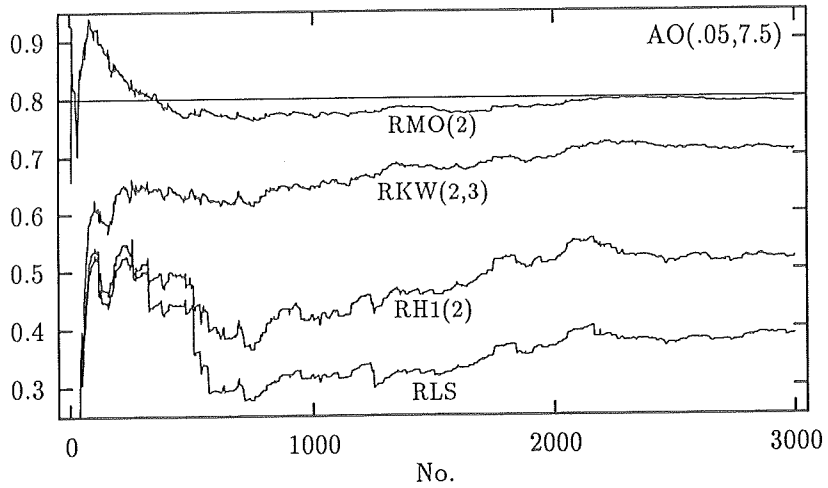


Figure 7.9: Trajectories of the parameter estimates in the first order system for additive outliers. The basic innovation sequence was the same as in Figure 7.6.

Clearly the additive outliers have a drastic effect on the estimates produced by all algorithms. The most clear effect is observed on the RLS estimate, but also the RH1 algorithm seems to allow considerable steps in the estimate; yet the bias is reduced by RH1. On the contrary the RKW(2,3) algorithm shows a more smooth course of the estimate, and it gives a considerable reduction of bias compared to RH1. The RMO algorithm is doing even better than RKW although the trajectory is likely to be more jagged. RMO is obviously doing an easy job classifying some of the additive outliers.

II: AR(1) MODEL WITH SMOOTH VARIATION OF THE AR PARAMETER

To investigate the performance of the algorithms in a situation where the AR parameter shows a smooth deterministic time variation the simulation is in this part based on the nominal model

$$\left(1 - a(t)q^{-1}\right) y(t) = e(t), \tag{7.86}$$

where $\{e(t)\}$ is $NIID(0;1)$ when there is no innovation outliers, and the time varying parameter is given by

$$a(t) = 0.8 + 0.1 \sin\left(\frac{2\pi t}{2000}\right). \tag{7.87}$$

This model is used for simulation with three different outlier schemes

1. NO: Without outliers.
2. IO(1,.05,7.5): Innovation outliers with 5% contamination with variance 7.5^2.
3. AO(1,.05,7.5): Additive outliers with 5% contamination with variance 7.5^2.

To allow the algorithms follow the time varying parameter they are all applied with $\lambda = 0.995$. This means that the estimates are based on effectively 200 observations.

For each outlier scheme RLS is applied. RH1, RH2 and RMO are applied with $c = 2$, and RKW is applied with (c, a)=(2,2) and (2,3).

In Figures 7.10-7.11 the trajectories of RLS, RH1, RKW and RMO when applied with $c = 2$, $a = 3$ and $\lambda = 0.995$ in the presence of innovation and additive outliers, respectively,are found. Also the underlying time varying simulation parameter is shown.

Comparing with the previous situation with constant parameter and $\lambda = 1$ it is evident that the trajectories are more noisy. Part of the explanation is

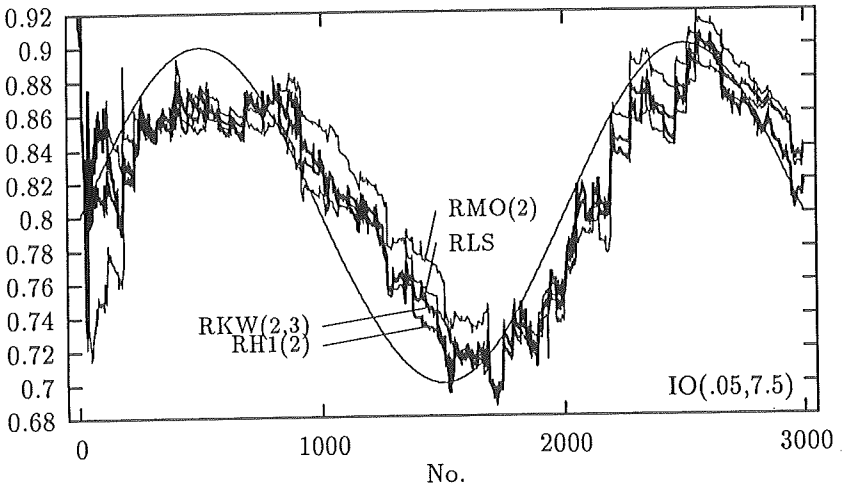


Figure 7.10: Trajectories of the parameter estimate in the time varying first order system with innovation outliers with variance 7.5^2.

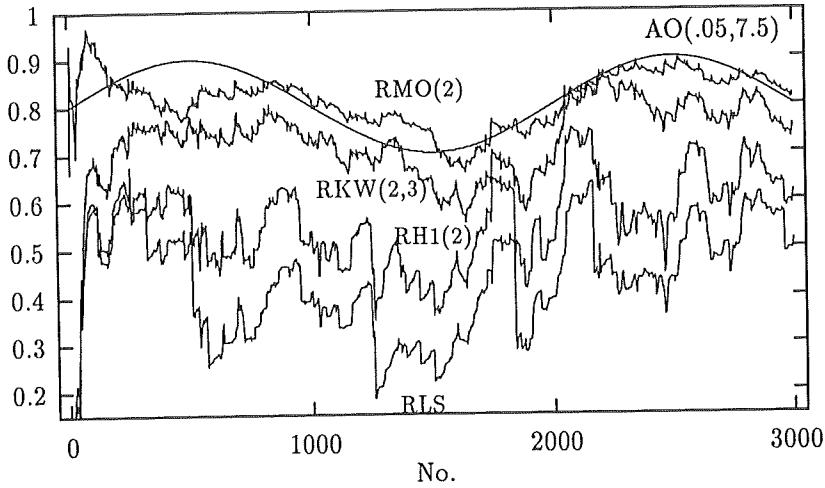


Figure 7.11: Trajectories of the parameter estimate in the time varying first order system with additive outliers with variance 7.5^2.

the continuous discount of information implying that the effective number of observations used in the estimation converge to and never exceeds a limit being $(1-\lambda)^{-1}$. The delayed tracking of exponential forgetting is rendered visible by the trajectories of all the algorithms. This deficiency of exponential forgetting as a remedy of obtaining a locally valid model estimate also contributes to the noisy trajectories due the general increase of the size of the prediction errors.

The actual comparison of the algorithms from Figure 7.10 is blurred by the non-perfect tracking capability of exponential forgetting. This matter of comparison is postponed until the quantified performance evaluation. However, it should be expected that the down-weighting or exclusion of information which takes place in the robust algorithms has a negative influence on the tracking capability.

In the additive outlier setting the picture is the same as earlier, i.e. RMO is closest to the underlying parameter variation though it is seems to be delayed. RLS shows a considerable bias as does RH1 and the trajectories are more ragged than for RKW and RMO. RKW is significantly closer to the true parameter variation than RH1.

III: AR(2) MODEL WITH CONSTANT PARAMETERS

In the third part of the study the simulation is carried out with the following second order model with constant parameters

$$\left(1 - 1.20q^{-1} + 0.52q^{-2}\right) y(t) = e(t), \tag{7.88}$$

where $\{e(t)\}$ is $NIID(0;1)$ for innovation outlier-free observations. and with the poles

$$z = 0.6 \pm 0.4i, \qquad |z| = 0.72. \tag{7.89}$$

RLS, RH1, RKW and RMO are applied using $c = 2$ and $a = 5$, and the following three noise schemes are used

1. NO: Without outliers.

2. IO(1,.05,2.5): Innovation outliers with 5% contamination with variance 2.5^2.

3. AO(1,.05,2.5): Additive outliers with 5% contamination with variance 2.5^2.

Since the data are simulated with constant parameters, all algorithms are applied with the forgetting factor $\lambda = 1$.

7.5.1 PERFORMANCE EVALUATION

For each sequence of parameter estimates the sum of squared deviations from the true parameter from observation no. 2001 till no. 3000 is computed

$$SSD = \sum_{i=2001}^{3000} \left(\hat{\theta}(i) - \theta\right)^2, \tag{7.90}$$

as is the average $\overline{SSD}$ of the 1000 values of SSD. Likewise, the average $\overline{\hat{\theta}(t)}$, the standard deviation $s_{\hat{\theta}(t)}$ and the 5, 50 and 95 % fractiles $\hat{\theta}(t)_{.05}$, $\hat{\theta}(t)_{.50}$ and $\hat{\theta}(t)_{.95}$ of the recursively obtained estimates at $t = 3000$ are computed. In addition $\overline{\hat{\theta}(t)}$ and $s_{\hat{\theta}(t)}$ are computed for $t = 2000$. The average, the standard deviation and the 50 % fractile are also computed for the estimates obtained from the off-line Krasker & Welsch algorithm.

To give a visual interpretation of the distribution of parameter estimates a nonparametric kernel density estimator (Silverman, 1986)

$$\hat{f}(u) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{u - \theta_i}{h}\right) \tag{7.91}$$

has been applied, where $\{\theta_i\}$ are the estimates entering the density estimation. h is the bandwidth (smoothing parameter), n is the number of estimates used in calculation and K is the kernel (weight function). In this

study the Epanechnikov Kernel below is used

$$K(u) = \begin{cases} \frac{3}{4}(1-u^2) & , \quad |u| < 1 \\ \\ 0 & , \quad |u| \geq 1 \end{cases} . \tag{7.92}$$

7.5.2 DISCUSSION OF THE RESULTS

I: AR(1) MODEL WITH CONSTANT PARAMETERS

In Tables 7.1-7.5 the performance statistics of the simulation and algorithm application for the first order model with constant parameters are listed.

For data simulated without outliers it appears that none of the algorithms give a better result than RLS. This is, of course, as expected since LS for normal distributed innovations corresponds to ML estimation. However, the other algorithms seems to give an unbiased estimate of the AR parameter as well, and for the larger values of c RH1, RH2 and RMO seem to be quite as efficient as RLS. For decreasing values of c the standard deviation of the estimates at $t = 2000$ and 3000 increases. This is reasonable since no erroneous information is present, and lowering of c implies down-weighting or removal of influence when prediction errors exceed the bound specified by c. For RKW the performance approaches that of RLS for increasing values of c and a, though it does not reach the same performance for the values tried in the tests. This is, of course, only a question of a sufficiently increasing the values of c and a.

Figure 7.12 shows the kernel estimate of the density of the estimates at $t = 3000$ obtained with $c = 2$ and $a = 3$ on data without outliers. RLS obviously has the most narrow density, but RH1(2) is only slightly inferior. RKW(2,3), however, shows a density with a little wider tails and a lower top level, and this is even more pronounced for RMO(2). This can be ascribed to the down-weighting or exclusion of information that actually is in accordance with the model. In Fig. 7.13 the density of the RKW(2,3) esti-

NO	$\theta = 0.8$	$t = 2000$		$t = 3000$				
Algorithm	$\overline{SSD}$	$\overline{\hat{\theta}(t)}$	$s_{\hat{\theta}(t)}$	$\overline{\hat{\theta}(t)}$	$s_{\hat{\theta}(t)}$	$\hat{\theta}(t)_{.05}$	$\hat{\theta}(t)_{.50}$	$\hat{\theta}(t)_{.95}$
RLS	.143	.799	.0133	.799	.0108	.780	.800	.817
RH1(1.5)	.149	.799	.0136	.799	.0110	.780	.799	.817
RH1(2)	.145	.799	.0134	.799	.0108	.780	.800	.816
RH1(3)	.143	.799	.0133	.799	.0108	.780	.800	.817
RH1(4)	.143	.799	.0133	.799	.0108	.780	.800	.817
RH2(1.5)	.150	.799	.0136	.799	.0110	.780	.799	.817
RH2(2)	.145	.799	.0134	.799	.0108	.780	.800	.816
RH2(3)	.143	.799	.0133	.799	.0108	.780	.800	.817
RH2(4)	.143	.799	.0133	.799	.0108	.780	.800	.817
RKW(1.5,1.5)	.225	.799	.0167	.799	.0134	.777	.800	.822
RKW(1.5,2)	.191	.799	.0153	.799	.0124	.778	.800	.819
RKW(1.5,3)	.178	.799	.0147	.799	.0120	.778	.800	.818
RKW(1.5,4)	.174	.799	.0146	.799	.0119	.778	.799	.818
RKW(2,1.5)	.210	.799	.0162	.799	.0130	.777	.800	.821
RKW(2,2)	.180	.799	.0149	.799	.0121	.778	.800	.819
RKW(2,3)	.166	.799	.0143	.799	.0116	.779	.799	.818
RKW(2,4)	.163	.799	.0141	.799	.0115	.780	.799	.818
RKW(3,1.5)	.192	.799	.0154	.799	.0124	.778	.800	.819
RKW(3,2)	.165	.799	.0143	.799	.0115	.779	.799	.818
RKW(3,3)	.154	.799	.0138	.799	.0112	.780	.799	.817
RKW(3,4)	.152	.799	.0137	.799	.0111	.780	.799	.817
RKW(4,1.5)	.179	.799	.0149	.799	.0120	.778	.800	.818
RKW(4,2)	.156	.799	.0139	.799	.0112	.780	.799	.817
RKW(4,3)	.149	.799	.0136	.799	.0109	.780	.799	.817
RKW(4,4)	.149	.799	.0136	.799	.0109	.780	.799	.817
RMO(1.5)	.434	.796	.0230	.797	.0184	.767	.797	.826
RMO(2)	.215	.798	.0163	.799	.0131	.776	.799	.820
RMO(3)	.147	.799	.0134	.799	.0109	.780	.800	.817
RMO(4)	.144	.799	.0133	.799	.0108	.780	.800	.817

Table 7.1: Results with no outliers and first order model.

mates is compared to the density of its off-line counterpart. The similarity between the densities is evident indicating that the effect of the approximations in the recursive version of the estimator by Krasker & Welsch is vanishing with an increasing number of observations.

In Table 7.2 is seen that with innovation outliers with variance 2.5^2 in data RH1, RH2, RKW and RMO improve compared to the situation with no outliers, whereas for RLS the result is very much the same without outliers and with innovation outliers. This is in accordance with the result by Martin & Yohai (1985), who show that the asymptotic covariance of the LS estimates of AR and MA parameters are independent of the innovation distribution. However, for a heavy-tailed innovation distribution LS estimates can be inefficient compared to ML estimates for the reason why LS does not make the most out of the increased precision attainable with heavy-tailed innovation distributions. When the distribution is not known, M-estimation can be applied for better utilization of the increased excitation for generally not known heavy-tailed innovation distributions Martin & Yohai (1985). Certainly, this is the reason why RH1, RH2, RKW and RMO, for some values of c and a, do better than RLS, although the differences are rather small. Also RKW is for a few combinations of (c, a) performing better than RLS, but the main conclusion is that RKW is not superior to RLS, and indeed, it seems to be preferable to use one of the Huber algorithms or RMO. The reason is probably that the bounded-influence quality of RKW is needless for innovation outliers in pure AR processes as all the regressors are coherent with the model assumptions.

Kernel estimates of the recursive parameter estimates at $t = 3000$ in the first order system with innovation outliers present is given in Fig. 7.14. The picture is almost the same as without outliers, except that RH1(2) is slightly superior to RLS. Fig. 7.15 shows that the agreement between RKW(2,3) and KW(2,3) also is found with innovation outliers in data.

When the variance of the innovation outliers is increased from 2.5^2 to 7.5^2 the differences among the algorithms become clearer. That is, it is evident that both of the algorithms based on the minimization of Huber's criterion

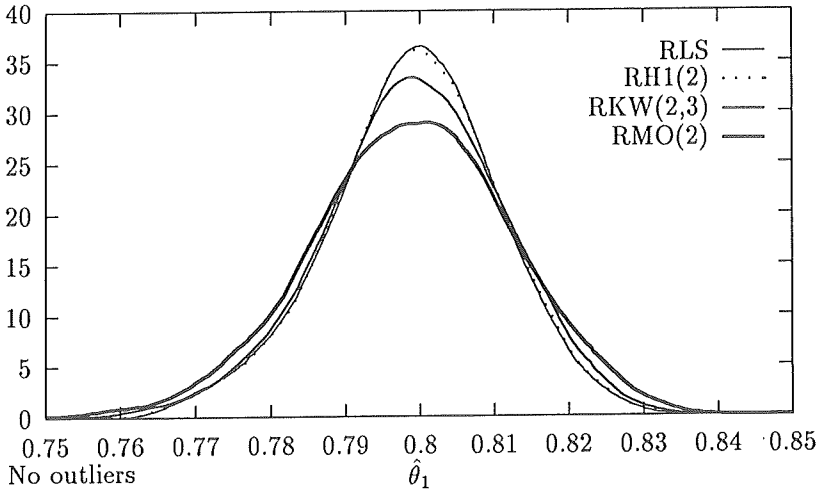


Figure 7.12: Kernel estimate of the density of the parameter estimate at $t = 3000$ for data without outliers. The kernel is Epanechnikov with bandwidth $h = 0.009$.

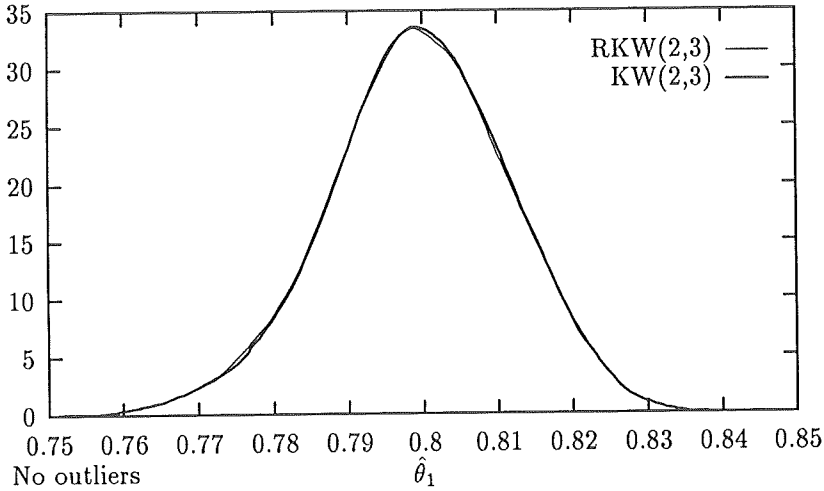


Figure 7.13: Kernel estimate of the density of the recursive and off-line Krasker & Welsch parameter estimate for data without outliers. The kernel is Epanechnikov with bandwidth $h = 0.009$.

IO(.05,1,2.5)	$\theta = 0.8$	$t = 2000$		$t = 3000$				
Algorithm	$\overline{SSD}$	$\overline{\hat{\theta}(t)}$	$s_{\hat{\theta}(t)}$	$\overline{\hat{\theta}(t)}$	$s_{\hat{\theta}(t)}$	$\hat{\theta}(t)_{.05}$	$\hat{\theta}(t)_{.50}$	$\hat{\theta}(t)_{.95}$
RLS	.144	.799	.0134	.799	.0109	.780	.800	.817
RH1(1.5)	.130	.799	.0127	.799	.0102	.780	.800	.817
RH1(2)	.130	.799	.0127	.799	.0103	.781	.800	.816
RH1(3)	.136	.799	.0130	.799	.0105	.780	.800	.816
RH1(4)	.141	.799	.0132	.799	.0107	.780	.800	.816
RH2(1.5)	.131	.799	.0127	.799	.0103	.781	.799	.815
RH2(2)	.130	.799	.0127	.799	.0103	.781	.800	.815
RH2(3)	.136	.799	.0130	.799	.0105	.781	.800	.816
RH2(4)	.141	.799	.0132	.799	.0107	.780	.800	.816
RKW(1.5,1.5)	.196	.799	.0156	.800	.0127	.778	.800	.821
RKW(1.5,2)	.165	.799	.0143	.799	.0116	.780	.800	.818
RKW(1.5,3)	.153	.799	.0138	.799	.0112	.781	.800	.817
RKW(1.5,4)	.151	.799	.0137	.799	.0111	.781	.800	.817
RKW(2,1.5)	.182	.799	.0151	.799	.0122	.779	.800	.819
RKW(2,2)	.156	.799	.0140	.799	.0113	.781	.800	.818
RKW(2,3)	.146	.799	.0135	.799	.0109	.781	.800	.816
RKW(2,4)	.145	.799	.0134	.799	.0108	.781	.800	.816
RKW(3,1.5)	.165	.799	.0144	.799	.0115	.781	.800	.818
RKW(3,2)	.146	.799	.0135	.799	.0109	.781	.800	.817
RKW(3,3)	.145	.799	.0135	.799	.0108	.781	.799	.816
RKW(3,4)	.143	.799	.0134	.799	.0107	.781	.799	.817
RKW(4,1.5)	.155	.799	.0140	.799	.0112	.781	.800	.817
RKW(4,2)	.142	.799	.0133	.799	.0107	.781	.799	.816
RKW(4,3)	.151	.799	.0138	.799	.0110	.780	.799	.816
RKW(4,4)	.164	.799	.0145	.799	.0115	.780	.800	.816
RMO(1.5)	.436	.796	.0230	.797	.0185	.769	.797	.826
RMO(2)	.187	.798	.0152	.799	.0122	.779	.798	.819
RMO(3)	.132	.799	.0127	.799	.0104	.782	.799	.816
RMO(4)	.136	.799	.0130	.799	.0105	.781	.800	.816

Table 7.2: Results with innovation outliers for first order model ($\gamma = 0.05$, $\sigma_{e_2} = 2.5$).

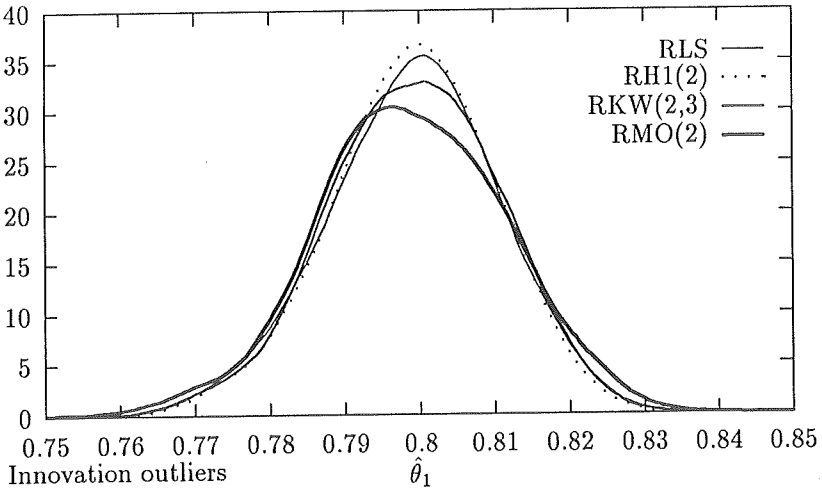


Figure 7.14: Kernel estimate of the density of the parameter estimate at $t = 3000$ for data with innovation outliers with variance 2.5^2. The kernel is Epanechnikov with bandwidth $h = 0.009$.

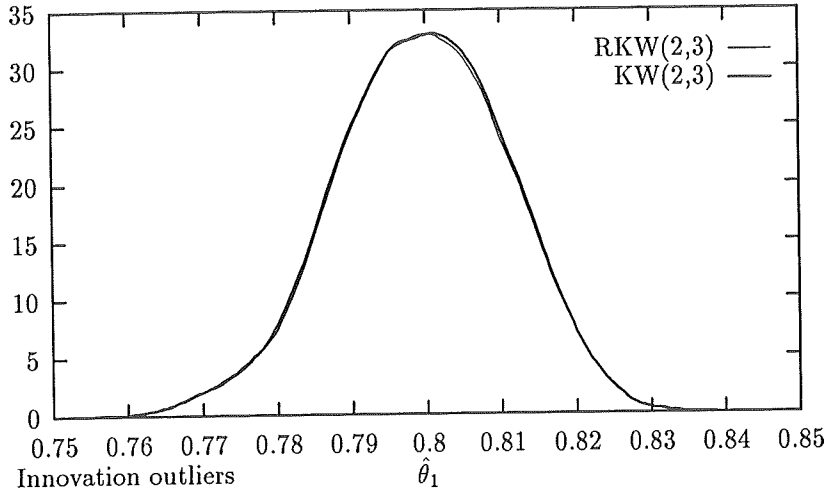


Figure 7.15: Kernel estimate of the density of the recursive and off-line Krasker & Welsch parameter estimate for data with innovation outliers with variance 2.5^2. The kernel is Epanechnikov with bandwidth $h = 0.009$.

IO(.05,1,7.5)	$\theta = 0.8$	$t = 2000$		$t = 3000$				
Algorithm	$\overline{SSD}$	$\overline{\hat{\theta}(t)}$	$s_{\hat{\theta}(t)}$	$\overline{\hat{\theta}(t)}$	$s_{\hat{\theta}(t)}$	$\hat{\theta}(t)_{.05}$	$\hat{\theta}(t)_{.50}$	$\hat{\theta}(t)_{.95}$
RLS	.150	.799	.0136	.799	.0110	.780	.799	.817
RH1(1.5)	.046	.799	.0076	.800	.0062	.789	.800	.809
RH1(2)	.048	.799	.0077	.800	.0063	.789	.800	.809
RH1(3)	.058	.799	.0086	.799	.0069	.788	.799	.810
RH1(4)	.078	.799	.0099	.799	.0079	.786	.799	.811
RH2(1.5)	.047	.799	.0076	.800	.0062	.789	.800	.810
RH2(2)	.048	.799	.0078	.800	.0063	.789	.800	.810
RH2(3)	.059	.799	.0086	.799	.0069	.788	.799	.810
RH2(4)	.078	.799	.0099	.799	.0080	.786	.799	.811
RKW(1.5,1.5)	.098	.800	.0112	.800	.0089	.786	.800	.815
RKW(1.5,2)	.076	.800	.0098	.800	.0079	.787	.800	.812
RKW(1.5,3)	.067	.799	.0092	.800	.0074	.787	.799	.812
RKW(1.5,4)	.064	.799	.0090	.800	.0073	.788	.799	.811
RKW(2,1.5)	.089	.800	.0106	.800	.0085	.786	.800	.814
RKW(2,2)	.070	.799	.0094	.800	.0076	.787	.800	.812
RKW(2,3)	.062	.799	.0088	.800	.0072	.787	.799	.811
RKW(2,4)	.063	.799	.0088	.800	.0072	.787	.800	.811
RKW(3,1.5)	.076	.799	.0099	.800	.0079	.787	.800	.812
RKW(3,2)	.063	.799	.0089	.800	.0072	.788	.800	.811
RKW(3,3)	.064	.799	.0089	.800	.0073	.787	.800	.811
RKW(3,4)	.068	.799	.0093	.799	.0075	.787	.800	.811
RKW(4,1.5)	.068	.799	.0093	.800	.0075	.787	.799	.812
RKW(4,2)	.062	.799	.0088	.800	.0072	.788	.799	.811
RKW(4,3)	.076	.799	.0098	.800	.0078	.787	.800	.811
RKW(4,4)	.113	.800	.0123	.800	.0092	.787	.800	.812
RMO(1.5)	106.9	.812	.3602	.809	.2907	.779	.799	.818
RMO(2)	5.886	.803	.1290	.800	.0294	.787	.799	.811
RMO(3)	.042	.800	.0072	.800	.0059	.789	.800	.810
RMO(4)	.042	.800	.0073	.800	.0059	.790	.800	.809

Table 7.3: Results with innovation outliers for first order model ($\gamma = 0.05$, $\sigma_{e_2} = 7.5$).

give a more efficient estimate of the AR parameter than RLS. The optimal c-value in RH1 and RH2 is 1.5. However, an even lower standard deviation of the estimates are obtained with RMO for c=3,4. The reason might be that it is better to totally discard the influence at large prediction errors than just down-weight the influence. On the contrary the result for RMO with c=1.5,2 is extremely poor. This result covers the incidence that the algorithm has lost track of the parameter estimates for some of the simulated data sequences. That is, the parameter estimate has become so wrong that the size of the prediction error, at least for a period, exceeds c times the estimated standard deviation. Indeed, this is an extremely undesirable property of the RMO algorithm which also is relevant for higher values of c. This makes the application of RMO in practical settings questionable. RKW does for the innovation outlier contamination with large variance perform better than RLS for all the tested values of c and a. The worst result is obtained with (c, a)=(1.5,1.5) or (4,4) corresponding to either too much down-weighting of good information and too little down-weighting of erroneous information. The interaction between c and a on the performance of the algorithm is seen from the fact that the best result is obtained with either (c, a)=(2,3-4) or (4,2). Finally, note that the result for RLS is almost the same as the previous results without outliers and with innovation outliers with variance 2.5^2.

Turning to the results for the estimation on data simulated with additive outliers with variance 2.5^2 in Table 7.4, a large bias is seen with all algorithms. The smallest bias is obtained with RMO(2). A plausible reason for this is that it is easy for the algorithm to classify the outliers as such. RKW(2,3) also shows a good performance evidently because it takes care of the outliers both by comparing the residuals to estimated variance and by calculating the length of the regressor, $\boldsymbol{x}(t)$, measured in the Euclidean norm in the space expanded by the estimated dispersion, $\hat{\boldsymbol{A}}^{-1}(t)$. RH1(1.5) and RH2(1.5) also reduce the effect of additive outliers but not quite enough, and as expected the bias is largest for RLS. It is evident that smaller values of c and a now give the best results. Note that a large amount of $\overline{SSD}$ stems from the bias in the estimates.

AO(.05,1,2.5)	$\theta = 0.8$	$t = 2000$		$t = 3000$				
Algorithm	$\overline{SSD}$	$\overline{\hat{\theta}(t)}$	$s_{\hat{\theta}(t)}$	$\overline{\hat{\theta}(t)}$	$s_{\hat{\theta}(t)}$	$\hat{\theta}(t)_{.05}$	$\hat{\theta}(t)_{.50}$	$\hat{\theta}(t)_{.95}$
RLS	7.132	.718	.0227	.718	.0182	.687	.719	.747
RH1(1.5)	3.399	.744	.0196	.744	.0158	.717	.745	.770
RH1(2)	4.359	.736	.0205	.737	.0166	.708	.738	.763
RH1(3)	6.129	.724	.0221	.724	.0178	.693	.725	.753
RH1(4)	6.880	.719	.0226	.720	.0182	.688	.721	.748
RH2(1.5)	3.410	.744	.0200	.744	.0161	.716	.746	.770
RH2(2)	4.338	.737	.0206	.737	.0167	.707	.738	.763
RH2(3)	6.084	.724	.0220	.725	.0178	.694	.726	.753
RH2(4)	6.852	.720	.0225	.720	.0182	.688	.721	.748
RKW(1.5,1.5)	1.132	.771	.0193	.771	.0156	.745	.772	.796
RKW(1.5,2)	1.275	.768	.0182	.768	.0147	.743	.769	.790
RKW(1.5,3)	1.431	.766	.0178	.766	.0144	.742	.766	.789
RKW(1.5,4)	1.485	.765	.0177	.765	.0143	.741	.765	.788
RKW(2,1.5)	1.174	.770	.0188	.770	.0152	.745	.771	.793
RKW(2,2)	1.414	.766	.0179	.766	.0144	.741	.766	.789
RKW(2,3)	1.671	.762	.0177	.762	.0143	.738	.763	.785
RKW(2,4)	1.758	.761	.0178	.761	.0144	.737	.762	.784
RKW(3,1.5)	1.302	.768	.0182	.768	.0146	.742	.768	.790
RKW(3,2)	1.801	.761	.0178	.761	.0145	.736	.761	.783
RKW(3,3)	2.301	.755	.0182	.755	.0148	.729	.756	.778
RKW(3,4)	2.465	.753	.0184	.753	.0149	.727	.754	.777
RKW(4,1.5)	1.467	.765	.0180	.765	.0145	.740	.766	.788
RKW(4,2)	2.270	.755	.0182	.755	.0148	.729	.756	.779
RKW(4,3)	3.008	.748	.0190	.748	.0155	.721	.749	.772
RKW(4,4)	3.278	.746	.0203	.746	.0163	.718	.747	.771
RMO(1.5)	1.308	.776	.0311	.777	.0253	.737	.779	.816
RMO(2)	.968	.775	.0215	.776	.0174	.747	.777	.803
RMO(3)	2.377	.755	.0207	.755	.0171	.724	.757	.782
RMO(4)	5.090	.732	.0227	.732	.0185	.699	.733	.762

Table 7.4: Results with additive outliers for first order model ($\gamma = 0.05$, $\sigma_v = 2.5$).

Obviously, it seems to be impossible to completely remove the bias of the parameter estimates with any of these recursive methods when additive outliers are present. In the off-line setting multi-stage methods, e.g. Allende & Heiler (1992) , involving outlier detection and filtering, can be applied. However, it is problematic to carry this over to a recursive estimation algorithm.

It is seen that the estimated parameter corresponds to a faster system than what is simulated. This seems reasonable, because the additive outliers occur without affecting the dynamics of the system, and therefore the contaminated observations seem to belong to a system with a smaller time constant.

The nonparametric density estimates of the parameter estimate given in Fig. 7.16 for the case with additive outliers clearly show the improvement obtained by using RKW and RMO, and Fig. 7.17 once again show the agreement between RKW and KW.

For additive outliers with the large variance (7.5^2) of the contamination RLS gives disastrous result both with respect to bias and variance of the estimates. The improvement of using RH1 and RH2 is only minor lying in a decrease of the bias. RKW is clearly better with the best result for $(c, a) = (1.5, 1.5)$. However, RMO(2) is superior and again clearly better than RKW(1.5,1.5).

As seen from the kernel estimates and also in Table 7.6 the resulting estimates are very much the same whether the off-line estimator proposed by Krasker & Welsch (1982) or the recursive variant is used.

To illustrate the performance of the recursive scale estimator proposed in Section 7.4.3, kernel estimates of the density of both the recursive ($t = 3000$) and the off-line scale estimates obtained with the Krasker & Welsch estimator for (c, a)=(2,3) are shown in Figures 7.18-7.20. Furthermore, Table 7.7 lists the average and standard deviation of the estimates at $t = 3000$ when outliers are present.

AO(.05,1,7.5)	$\theta = 0.8$	$t = 2000$		$t = 3000$				
Algorithm	$\overline{SSD}$	$\overline{\hat{\theta}(t)}$	$s_{\hat{\theta}(t)}$	$\overline{\hat{\theta}(t)}$	$s_{\hat{\theta}(t)}$	$\hat{\theta}(t)_{.05}$	$\hat{\theta}(t)_{.50}$	$\hat{\theta}(t)_{.95}$
RLS	162.5	.400	.0441	.398	.0355	.342	.398	.457
RH1(1.5)	49.08	.582	.0448	.583	.0362	.520	.584	.638
RH1(2)	76.71	.527	.0523	.527	.0424	.454	.529	.595
RH1(3)	134.9	.436	.0525	.435	.0427	.364	.435	.508
RH1(4)	157.5	.406	.0465	.405	.0375	.346	.405	.470
RH2(1.5)	49.30	.582	.0482	.583	.0401	.510	.586	.642
RH2(2)	73.97	.533	.0527	.532	.0442	.457	.535	.599
RH2(3)	128.2	.447	.0527	.444	.0436	.371	.443	.517
RH2(4)	153.1	.413	.0470	.410	.0380	.349	.410	.476
RKW(1.5,1.5)	2.474	.754	.0206	.754	.0166	.726	.755	.779
RKW(1.5,2)	3.156	.746	.0200	.747	.0161	.719	.748	.772
RKW(1.5,3)	4.024	.739	.0205	.739	.0164	.711	.741	.764
RKW(1.5,4)	4.475	.736	.0210	.736	.0168	.707	.737	.762
RKW(2,1.5)	2.710	.751	.0203	.751	.0163	.723	.752	.777
RKW(2,2)	3.927	.740	.0202	.740	.0163	.712	.741	.765
RKW(2,3)	5.527	.728	.0216	.728	.0174	.698	.729	.755
RKW(2,4)	6.354	.723	.0225	.723	.0182	.692	.724	.751
RKW(3,1.5)	3.656	.743	.0208	.742	.0167	.714	.743	.768
RKW(3,2)	7.183	.719	.0235	.717	.0192	.686	.718	.747
RKW(3,3)	11.91	.694	.0283	.694	.0231	.653	.695	.730
RKW(3,4)	14.26	.684	.0303	.684	.0247	.641	.686	.723
RKW(4,1.5)	5.532	.729	.0227	.728	.0186	.695	.729	.756
RKW(4,2)	13.72	.687	.0297	.685	.0249	.641	.687	.724
RKW(4,3)	23.50	.651	.0366	.650	.0304	.596	.652	.698
RKW(4,4)	27.88	.637	.0396	.637	.0328	.579	.639	.688
RMO(1.5)	2.356	.780	.0505	.782	.0400	.738	.786	.824
RMO(2)	.838	.782	.0264	.784	.0209	.753	.786	.812
RMO(3)	2.469	.760	.0336	.762	.0280	.726	.766	.793
RMO(4)	20.29	.677	.0762	.678	.0689	.526	.696	.750

Table 7.5: Results with additive outliers for first order model ($\gamma = 0.05$, $\sigma_v = 7.5$).

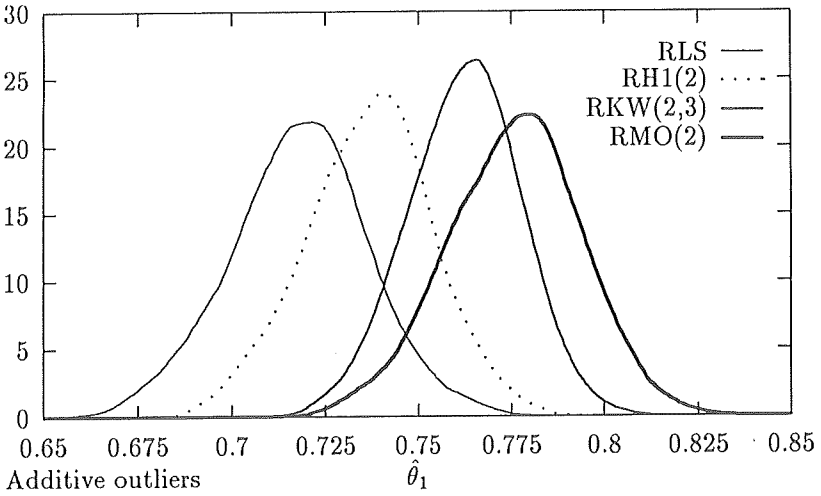


Figure 7.16: Kernel estimate of the density of the parameter estimate at $t = 3000$ for data with additive outliers with variance 2.5^2. The kernel is Epanechnikov with bandwidth $h = 0.011$.

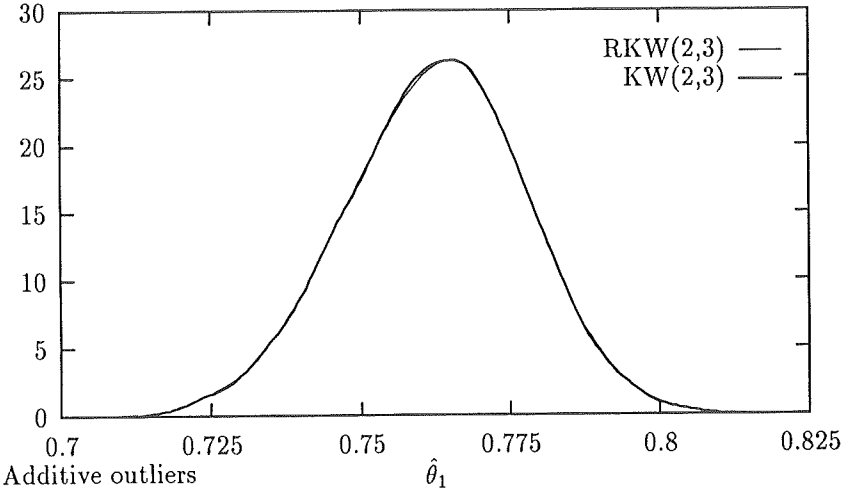


Figure 7.17: Kernel estimate of the density of the recursive and off-line Krasker & Welsch parameter estimate for data with additive outliers with variance 2.5^2. The kernel is Epanechnikov with bandwidth $h = 0.011$.

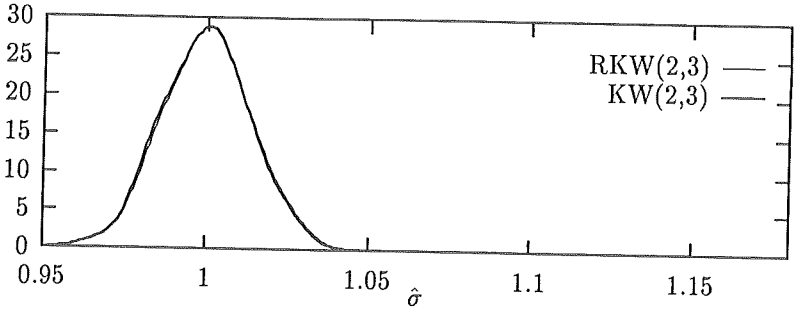


Figure 7.18: Kernel estimate of the density of the recursive and off-line Krasker & Welsch scale estimate for data without outliers. The kernel is Epanechnikov with bandwidth $h = 0.009$.

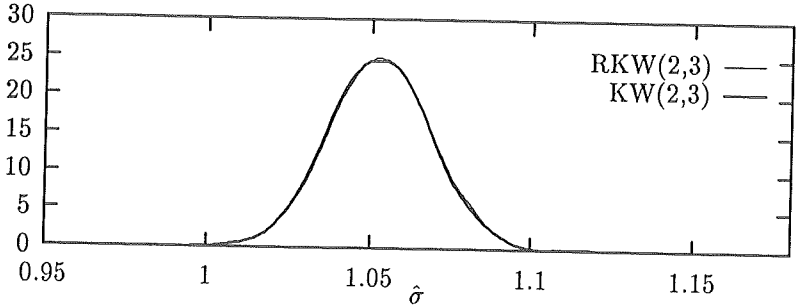


Figure 7.19: Kernel estimate of the density of the recursive and off-line Krasker & Welsch scale estimate for data with innovation outliers with variance 2.5^2. The kernel is Epanechnikov with bandwidth $h = 0.009$.

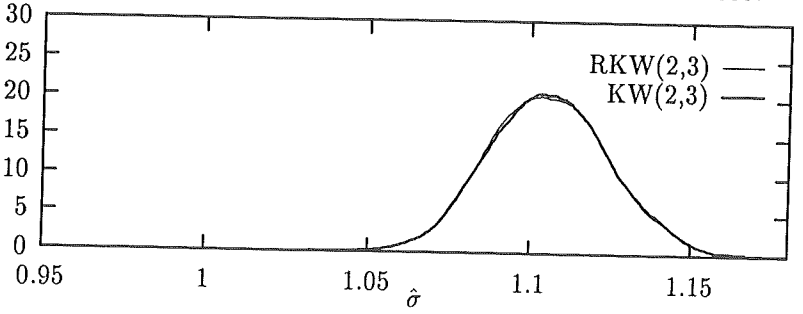


Figure 7.20: Kernel estimate of the density of the recursive and off-line Krasker & Welsch scale estimate for data with additive outliers with variance 2.5^2. The kernel is Epanechnikov with bandwidth $h = 0.011$.

$\theta = 0.8$			Off-line			Recursive at $t = 3000$		
Outliers	c	a	$\overline{\hat{\theta}}$	$s_{\hat{\theta}}$	$\hat{\theta}_{.50}$	$\overline{\hat{\theta}(t)}$	$s_{\hat{\theta}(t)}$	$\hat{\theta}(t)_{.50}$
None	2	3	0.799	0.012	0.799	0.799	0.012	0.799
None	3	3	0.799	0.011	0.800	0.799	0.011	0.800
None	3	4	0.799	0.011	0.800	0.799	0.011	0.799
Innovation	2	3	0.799	0.011	0.800	0.799	0.011	0.800
Innovation	3	3	0.799	0.011	0.800	0.799	0.011	0.799
Innovation	3	4	0.799	0.011	0.800	0.799	0.011	0.799
Additive	2	3	0.762	0.014	0.763	0.762	0.014	0.763
Additive	3	3	0.755	0.015	0.756	0.755	0.015	0.756
Additive	3	4	0.753	0.015	0.754	0.753	0.015	0.754

Table 7.6: Results of Krasker-Welsch estimation of first order AR-model.

From the nonparametric estimates it is seen that the recursive estimates have densities being similar to the densities of the off-line estimates for the recursive version of Huber's Proposal 2 combined with the method by Krasker & Welsch.

Figure 7.18 indicates that for data without outliers the recursive version of Huber's Proposal 2 results in a consistent scale estimate.

With innovation outliers in data the scale estimate is generally larger than for the outlier free observations. However, Table 7.7 shows that the bias is significantly reduced by all of the robust proposals when compared to the RLS scale estimate. The empirical standard deviation of the scale estimate is also clearly reduced compared to RLS. For RH1, RH2 and RKW the bias increases with the choice of c, and it generally applies that the bias increases with the degree of contamination. RMO is different as the smallest bias and standard deviation is obtained for $c = 3$. This result should be considered in connection with the result for the estimation of the AR parameter which shows a bad result for RMO with $c = 1.5, 2$. For these values of c it turned out that RMO during some simulations periodically has lost track of the

Algorithm	IO(.05,1,2.5)		IO(.05,1,7.5)		AO(.05,1,2.5)		AO(.05,1,7.5)	
	$\overline{\hat{\sigma}(t)}$	$s_{\hat{\sigma}(t)}$	$\overline{\hat{\sigma}(t)}$	$s_{\hat{\sigma}(t)}$	$\overline{\hat{\sigma}(t)}$	$s_{\hat{\sigma}(t)}$	$\overline{\hat{\sigma}(t)}$	$s_{\hat{\sigma}(t)}$
RLS	1.126	.0217	1.956	.1437	1.224	.0312	2.182	.1880
RH1(1.5)	1.043	.0162	1.071	.0173	1.088	.0182	1.211	.0401
RH1(2)	1.053	.0157	1.096	.0179	1.108	.0190	1.297	.0517
RH1(3)	1.079	.0192	1.198	.0317	1.156	.0281	1.527	.0744
RH1(4)	1.100	.0271	1.367	.0652	1.189	.0335	1.734	.0991
RH2(1.5)	1.043	.0162	1.071	.0172	1.088	.0183	1.209	.0434
RH2(2)	1.052	.0157	1.096	.0179	1.107	.0191	1.290	.0525
RH2(3)	1.079	.0194	1.198	.0317	1.156	.0280	1.520	.0749
RH2(4)	1.100	.0267	1.366	.0652	1.189	.0337	1.731	.0991
RKW(1.5,1.5)	1.044	.0162	1.072	.0173	1.086	.0240	1.147	.0285
RKW(1.5,2)	1.044	.0163	1.072	.0173	1.086	.0181	1.148	.0226
RKW(1.5,3)	1.043	.0163	1.071	.0173	1.086	.0180	1.149	.0229
RKW(1.5,4)	1.043	.0163	1.071	.0173	1.086	.0180	1.150	.0230
RKW(2,1.5)	1.053	.0156	1.097	.0177	1.105	.0246	1.205	.0336
RKW(2,2)	1.053	.0156	1.097	.0177	1.105	.0259	1.206	.0348
RKW(2,3)	1.053	.0156	1.096	.0177	1.105	.0186	1.209	.0287
RKW(2,4)	1.053	.0157	1.096	.0178	1.105	.0186	1.210	.0289
RKW(3,1.5)	1.080	.0190	1.199	.0311	1.154	.0327	1.454	.0767
RKW(3,2)	1.079	.0190	1.199	.0312	1.155	.0279	1.454	.0706
RKW(3,3)	1.079	.0188	1.199	.0314	1.154	.0281	1.455	.0713
RKW(3,4)	1.079	.0189	1.199	.0319	1.154	.0279	1.456	.0707
RKW(4,1.5)	1.100	.0267	1.367	.0657	1.190	.0348	1.781	.1280
RKW(4,2)	1.100	.0271	1.366	.0667	1.189	.0351	1.765	.1200
RKW(4,3)	1.100	.0279	1.367	.0680	1.189	.0349	1.752	.1170
RKW(4,4)	1.100	.0277	1.369	.0687	1.189	.0337	1.749	.1142
RMO(1.5)	1.185	.0813	1.180	.0795	1.201	.0884	1.186	.0908
RMO(2)	1.053	.0340	1.045	.0357	1.076	.0386	1.058	.0430
RMO(3)	1.039	.0165	1.023	.0161	1.081	.0208	1.057	.0315
RMO(4)	1.071	.0179	1.054	.0201	1.144	.0255	1.171	.0774

Table 7.7: Scale estimation performance in the presence of outliers at $t = 3000$, and variance of outlier-free innovation equal to one.

AR parameter for which reason the scale estimate has increased.

In the additive outlier setting it is also seen that the robust algorithms imply a clear reduction in bias, and that for RH1, RH2 and RKW the best result is obtained for the lowest value of c. For RMO the best scale estimation performance is for $c = 2$ coinciding with the result for the estimation of the AR parameter. Generally the bias as well as the empirical standard deviation of the scale estimate exceed the results in the innovation outlier settings. Partially this is due to a larger variance of the contaminating densities, but also the biased AR estimates give rise to an increase in the scale parameter estimate.

By determining σ in $E[\chi(e/\sigma)] = 0$ using Huber's Proposal 2 with $c = 2$ for e given by the three noise models corresponding to Figures 7.18-7.20, respectively (assuming that the model parameter is known) the expected scale parameter estimates, $\tilde{\sigma}_{NO} = 1.0$, $\tilde{\sigma}_{IO} = 1.054$ and $\tilde{\sigma}_{AO} = 1.088$, are obtained. For no outliers and for innovation outliers the center of the densities of the scale parameter estimates are close to the corresponding expected values. However, for additive outliers the center lies above the expected value. The explanation is most likely that for additive outliers the model parameter estimate is biased for which reason the prediction errors differ from the simulated noise components and therefore give rise an increased scale parameter estimate.

In the RH and RKW the standard deviation of the scale is used to measure the size of the prediction errors with the purpose of detecting outliers. Consequently, the standard deviation used for this purpose should correspond to the outlier-free innovations, and therefore it is not desirable that the estimate of the standard deviation exceeds this value. Nor is it desirable that the scale estimate increases with increasing outlier contamination since the optimal value of c thus depends on the contamination degree. However, for little contamination the bias will be small and the optimal c-value will not change too much. Thus it is appropriate to search for methods giving smaller bias. For instance, a $\psi(u)$-function, used in Huber's Proposal 2, which is zero for large values of u could be used, but it is clear that when

the model parameters are biased this will also be the case for the scale parameter estimate.

II: AR(1) MODEL WITH SMOOTH VARIATION OF THE AR PARAMETER

NO	$\theta = \theta(t)$	$t = 2000$		$t = 3000$				
Algorithm	$\overline{SSD}$	$\overline{\hat{\theta}(t)}$	$s_{\hat{\theta}(t)}$	$\overline{\hat{\theta}(t)}$	$s_{\hat{\theta}(t)}$	$\hat{\theta}(t)_{.05}$	$\hat{\theta}(t)_{.50}$	$\hat{\theta}(t)_{.95}$
RLS	2.158	.755	.0362	.849	.0241	.808	.851	.886
RH1(2)	2.159	.755	.0365	.849	.0242	.807	.851	.886
RH2(2)	2.193	.754	.0357	.850	.0238	.808	.852	.886
RKW(2,2)	2.254	.755	.0381	.847	.0283	.796	.850	.890
RKW(2,3)	2.206	.755	.0374	.847	.0270	.799	.850	.889
RMO(2)	2.193	.753	.0370	.854	.0250	.808	.856	.892

Table 7.8: Results with no outliers and first order model with time varying parameter.

Tables 7.8-7.10 lists the results from the application of the algorithms on observations simulated with a first order model with time varying AR parameter. The robust algorithms are applied with $c = 2$ and RKW with $a = 2, 3$.

For observations without outliers the algorithms give comparable results. It is evident that all algorithms have troubles tracking the time varying parameter. This is seen from the result that the estimates at $t = 2000$ are clearly below 0.8 and at $t = 3000$ clearly above 0.8, where 0.8 is the actual value of the time varying parameter at these time instants.

In the presence of innovation outliers with variance 7.5^2, RLS has the worst performance. The lowest value of $\overline{SSD}$ is obtained with RH and RKW. The reduction in $\overline{SSD}$ seems to be obtained by a reduction of the empirical standard deviation of the estimates, as seen from the values of $s_{\hat{\theta}}(t)$, more than a reduction in bias.

IO(.05,1,7.5)	$\theta = \theta(t)$	$t = 2000$		$t = 3000$				
Algorithm	$\overline{SSD}$	$\overline{\hat{\theta}(t)}$	$s_{\hat{\theta}(t)}$	$\overline{\hat{\theta}(t)}$	$s_{\hat{\theta}(t)}$	$\hat{\theta}(t)_{.05}$	$\hat{\theta}(t)_{.50}$	$\hat{\theta}(t)_{.95}$
RLS	2.141	.755	.0343	.850	.0239	.808	.853	.886
RH1(2)	1.607	.757	.0230	.851	.0159	.823	.851	.875
RH2(2)	1.677	.756	.0222	.852	.0155	.825	.852	.876
RKW(2,2)	1.593	.757	.0256	.849	.0180	.818	.849	.877
RKW(2,3)	1.577	.757	.0251	.849	.0172	.820	.850	.876
RMO(2)	1.971	.755	.0207	.855	.0152	.829	.856	.880

Table 7.9: Results with innovation outliers for first order model with time varying parameter ($\gamma = 0.05$, $\sigma_{e'} = 7.5$).

AO(.05,1,7.5)	$\theta = \theta(t)$	$t = 2000$		$t = 3000$				
Algorithm	$\overline{SSD}$	$\overline{\hat{\theta}(t)}$	$s_{\hat{\theta}(t)}$	$\overline{\hat{\theta}(t)}$	$s_{\hat{\theta}(t)}$	$\hat{\theta}(t)_{.05}$	$\hat{\theta}(t)_{.50}$	$\hat{\theta}(t)_{.95}$
RLS	149.1	.358	.0919	.490	.0966	.335	.492	.646
RH1(2)	70.22	.449	.1070	.627	.0880	.472	.635	.751
RH2(2)	71.77	.457	.0871	.640	.0729	.513	.644	.749
RKW(2,2)	8.009	.682	.0529	.792	.0387	.726	.795	.849
RKW(2,3)	9.808	.667	.0568	.783	.0409	.715	.786	.842
RMO(2)	3.611	.734	.0413	.843	.0272	.796	.844	.888

Table 7.10: Results with additive outliers for first order model with time varying parameter ($\gamma = 0.05$, $\sigma_v = 7.5$).

For the situation with additive outliers the same pattern applies as previously with constant parameter. That is, RMO is doing surprisingly good (almost as good as for innovation outliers) and RKW also results in a considerable reduction of the bias compared to RLS and the RH algorithms. The bias for the two last-mentioned algorithms is, however, clear, and they both seem to be inapplicable for data with additive outlier effects of some kind.

III: AR(2) MODEL WITH CONSTANT PARAMETERS

Table 7.11 shows the results obtained, when estimating the parameters of the second order model. When innovation outliers are exciting the system, the RMO(2) algorithm gives a more flat distribution of the parameter estimates and a more disturbed parameter estimate trajectory than the rest of the algorithms. The differences are, however, small, and all of the algorithms are obviously working well. In the additive outlier case, however, the low value of the cut-off limit in RMO(2) forces the algorithm to expel of that large a part of the data that what remains is enough to make a good performance. The RLS estimates are brought far away from the true values of the parameters and hence, as is well known, RLS is a dangerous algorithm when applied to data with additive outliers. Irrespective of the nature of the outliers, the RKW algorithm performs satisfactory.

7.6 CONCLUSION

The successful application of parameter estimation algorithms to real data demands that the algorithms are robust in order to reduce errors in the estimates originating from outliers of different kinds. The traditionally formulated methods for robust estimation only consider the off-line situation. However, for several purposes, e.g. adaptive forecasting and control or efficient treatment of large amounts of data, on-line estimation is desirable.

Outliers/	$\theta = 1.20$				$\theta = -0.52$			
Algorithm	$\overline{SSD}$	$\overline{\hat{\theta}(t)}$	$s_{\hat{\theta}(t)}$	$\hat{\theta}(t)_{.50}$	$\overline{SSD}$	$\overline{\hat{\theta}(t)}$	$s_{\hat{\theta}(t)}$	$\hat{\theta}(t)_{.50}$
NO								
RLS	0.292	1.200	0.016	1.200	0.293	-0.520	0.016	-0.521
RH1(2)	0.298	1.200	0.016	1.200	0.298	-0.520	0.016	-0.521
RKW(2,5)	0.330	1.199	0.017	1.200	0.330	-0.520	0.016	-0.521
RMO(2)	0.452	1.199	0.019	1.199	0.468	-0.520	0.020	-0.521
IO(.05,1,7.5)								
RLS	0.289	1.200	0.016	1.200	0.295	-0.520	0.016	-0.521
RH1(2)	0.263	1.200	0.015	1.200	0.270	-0.520	0.015	-0.521
RKW(2,5)	0.295	1.200	0.016	1.200	0.296	-0.520	0.016	-0.521
RMO(2)	0.391	1.199	0.018	1.200	0.405	-0.520	0.018	-0.520
AO(.05,1,7.5)								
RLS	71.276	0.935	0.033	0.934	55.084	-0.287	0.029	-0.287
RH1(2)	43.156	0.995	0.029	0.996	36.346	-0.332	0.027	-0.331
RKW(2,5)	13.554	1.086	0.022	1.086	12.145	-0.412	0.022	-0.412
RMO(2)	7.529	1.120	0.030	1.121	8.245	-0.436	0.031	-0.437

Table 7.11: Summary of the recursively obtained estimates at $t = 3000$ with second order model.

In this chapter two recursive algorithms for robust estimation of AR parameters are derived. One is based on M-estimation for linear regression using Huber's weight function (Huber, 1973) and the other is derived from a bounded-influence estimator (Krasker & Welsch, 1982).

These derivations demonstrate a general approach for deriving recursive robust estimation algorithms, whenever the estimates can be found as the minimizing arguments of a criterion function. Depending on the actual criterion, the derivation involves the introduction of approximations, which influences the performance of the algorithm.

A simulation study is carried out to investigate the performance of the

estimators in settings, where innovation outliers and additive outliers are present. A comparison with another algorithm for recursive M-estimation (Cipra, 1992) as well as with RLS and a modification of RLS with rejection of the observation's influence for large prediction errors is made.

An important conclusion from the simulation study is that the influence of the approximations, introduced in the algorithm derivations, vanishes with increasing number of observations, and thus the approximations only influence the initial phase of an application.

Furthermore, algorithms based on M-estimation are appropriate choices for estimation in AR models in settings, where only innovation outliers might occur. The reason is that M-estimation is capable of utilizing the increased precision attainable from heavy-tailed innovation densities, and the bounded-influence quality is not necessary.

However, when additive outliers are present, RLS and also the M-estimation algorithms are not recommendable for estimation as considerable bias occur. The bounded-influence algorithm implies a significant reduction of the bias, but it is not capable of removing it totally.

The RLS algorithm extended with a facility for treating detected outliers as missing observations shows the best performance. The likely explanation is that whenever an outlier is present it is rewarding to remove its influence. The difficulty lies in the detection of outliers. Since the outlier detection in the current algorithm use the parameter estimates there is a risk of ending in a situation, at least periodically, where biased estimates implies all observations to be classified as outliers. Indeed, this counts in disfavour of this algorithm.

Whenever parameters are time-varying the important thing to mention is that the exponential forgetting does not extend the outlined algorithms in a way that makes them capable of tracking the time varying parameter without a delay. This is a deficiency that always exists whenever the parameter variation is not predictable. Apart for this result the same conclusions can

be drawn as from the constant parameter setting.

Based on these conclusions it is recommended to use either the proposed algorithm for recursive bounded-influence estimation or RLS extended with detection and rejection of outliers, whenever gross errors being additive effects are likely to occur. When using the last-mentioned algorithm it seems to be necessary to improve the detection of outliers to avoid the drawback of loosing track of the parameter.

.....

All programming in the simulation and estimation study was performed in Pascal and Fortran on a HP-9000/750 using double precision arithmetic. For the generation of Gaussian and uniform random variables the pseudo random number generators of the IMSL Stat/Library (Inc., 1987a) DRNNOF and DRNUNF were used. For calculation of the solution to the off-line estimation using Equations (7.29)-(7.32) the nonlinear-equation solver DNEQNF of the IMSL Math/Library (Inc., 1987b) was used.

.....

8

CONCLUSIONS

Predictions of heat load can in DH systems be used as input to control and optimization of production and distribution of consumer-required heat supply. Even a minimum increase in forecast precision can lead to large savings in energy with resulting economical savings and environmental benefits. The present thesis should be considered as a contribution to the research in gaining increased forecast precision of load in DH systems as well as in energy systems in general and other kinds of dynamic system. Moreover, the contents in each of the chapters represent different aspects of model building and estimation.

The conclusions of the thesis can be summarized as follows

Chapter 2. A DH system is a distributed parameter system. The dominating time constants of that part of the system, which are found in the consumers' buildings, can lie in the range from about 24 hours to one week. The actual value depends on the heat capacity of each building. This is determining the dynamics of the influence of ambient weather condition on heat load. Moreover, the distribution of heat takes place with considerable transport delays in the transmission and distribution net. However, due to the propagation of pressure fronts with a velocity close to that of sound, changes of flow demand from the consumers imply almost immediate change of flow at the location of the heat supply. The huge heat capacity of the DH water means that change of supply temperature implies an additional heat production

used for increase of accumulated energy in the water.

Chapter 3. It is demonstrated how physical considerations concerning flow rate and heat transfer in a heat exchanger can be used to obtain a simplified lumped parameter description in continuous time of the outlet temperatures as functions of inlet temperatures and mass flows.

Chapter 4. Nonparametric regression is applied for estimation of the dependence of heat load on time of day and night as well as on ambient air temperature and supply temperature. In doing this it is uncovered how the heat load depends in a nonlinear way on the ambient air temperature.

Chapter 5. A model belonging to the class of linear transfer function models using time and type of day, ambient air temperature and supply temperature is estimated. By estimating the same model on three disjoint data intervals the insufficiency of the linear model class is indicated. It is shown how nonlinear extensions, suggested by the nonparametric identification study along with physical considerations, gives an improved fit to data. These models consist of a smooth transfer function model and a model where explanatory variables enters linearly and squared. Furthermore, two-layered feed-forward neural network models are estimated to give an even better fit to data. However, when testing the estimated models on a validation set, the nonlinear extensions to the linear transfer function model outperform the neural network model. This point out the weak spot of the neural network approach being the difficulty or impossibility of controlling the flexibility the model.

Chapter 6. Models that are linear in the parameters are estimated using RLS with simple exponential forgetting. The variations of the parameter estimates disclose the time-varying dependence on supply and ambient air temperature. The introduction of suitable transformations of the observations as explanatory variables, as in Chapter 5, slightly improves the fit of the adaptively estimated model. j-step predictors for $j \in [1;24]$ which are linear in the parameters and are obtained by reformulations of transfer function models, where the use

of ambient air temperature is shifted j hours backwards, are estimated in the same way. Also a state space formulation of a diurnal and a weekly profile, updated by Kalman filter recursions, are investigated. For increasing prediction horizon is seen an increased dependence on a profile and a decrease of dependence on the dynamic modelling in the prediction meaning that, as both methods contain a diurnal profile, the performance of the two methods approaches each other.

Chapter 7. The likely occurrence of outliers in real data motivates the use of recursive estimation algorithms being robust against the erroneous influence of such outliers. Two robust algorithms are derived using a generally applicable approach for derivation of an algorithm as a recursive minimization of a criterion. In the first algorithm large residuals are weighted linearly instead of quadratically (M-estimation), and the second algorithm is an extension where the influence of regressors are bounded (bounded-influence-/general M-estimation). In a simulation study it is demonstrated that the influence of approximations, required in the derivation, vanishes with increasing number of observations. The algorithms are compared to a modification of RLS where large prediction errors implies the algorithm to exclude the influence of the corresponding observation. The last algorithm turns out to perform well, at least in the presence of outliers giving additive effects. However, the performance has a clear dependence on the tuning parameter, and it possesses the undesirable property that it periodically can loose track of the parameters. The bounded-influence implies a significant reduction in bias for additive outliers, and an increased precision in the presence of innovation outliers, whereas the M-estimation algorithm does not perform well when additive outliers are present. The investigation in this chapter has demonstrated that whenever outliers being additive effects may occur RLS estimates will be subject to considerable bias. Consequently, in this situation it is recommendable to apply an estimation algorithm that minimize the influence of such outliers.

A

ML Estimation of Continuous Time Models

In this appendix is outlined how the parameters of a model, which is formulated in continuous time by a set of linear differential equations, can be estimated by using a maximum likelihood method (ML).

Let the system under consideration be described by the system of deterministic, linear differential equations

$$\frac{d\boldsymbol{T}(t)}{dt} = \boldsymbol{A}\boldsymbol{T}(t) + \boldsymbol{B}\boldsymbol{U}(t). \tag{A.1}$$

$\boldsymbol{T}(t)$ is the state vector of the model, and $\boldsymbol{U}(t)$ is a vector of input to the system. The dynamics of the system is accounted for by the system matrix $\boldsymbol{A}$, and the influence of the external variable is described by the matrix $\boldsymbol{B}$.

Since the system of differential equations in Equation (A.1) is a model of a system and, as such, only give an approximative description of the system, it is required to introduce a stochastic component to account for model inadequacies. To arrive at a model in discrete time, where the model error is a sequence of normal distributed random variables that are independent in time, the stochastic process in the continuous time formulation is assumed to a be Wiener process. This means that the stochastic counterpart of Equation (A.1) is

$$d\boldsymbol{T}(t) = \boldsymbol{A}\boldsymbol{T}(t)dt + \boldsymbol{B}\boldsymbol{U}(t) + d\boldsymbol{w}(t), \tag{A.2}$$

where $\boldsymbol{w}(t)$ is the Wiener process with incremental covariance matrix $\boldsymbol{R}_w dt$.

Assume that a linear combination of the states of $\boldsymbol{T}(t)$ are measured at discrete instants $t = t_1, t_2, \ldots, t_n$ described by the measurement equation

$$\boldsymbol{T}_r(t_i) = \boldsymbol{C}\boldsymbol{T}(t_i) + \boldsymbol{e}(t_i), \qquad t_i \in [t_1, \ldots, t_n]. \tag{A.3}$$

The matrix $\boldsymbol{C}$ describes what is measured in the state vector. $\{\boldsymbol{e}(t_i)\}$ is a sequence of zero mean, independent and indentically, normal distributed variables with covariance matrix $\boldsymbol{R}_e$. In addition, it is assumed that $\boldsymbol{w}(t)$ and $\boldsymbol{e}(t_i)$ are mutually independent.

Having obtained measurements from the system given by Equation (A.3) the parameter estimation is carried out by maximizing the likelihood function according to the probabilistic part of the model, viz. the assumed probability density function for $\boldsymbol{w}(t)$ and $\boldsymbol{e}(t_i)$. To do this a discrete time formulation of the process in Equation (A.2) is necessary. This is obtained by integrating the continuous time model through the sampling intervals $t_2 - t_1, t_3 - t_2, \ldots, t_n - t_{n-1}$, respectively.

A.1 OBTAINING THE MODEL IN DISCRETE TIME

The discrete time model is obtained by integrating the continuous time model (A.2) through a given sampling interval Δt. Analytically this gives

$$\begin{aligned} \boldsymbol{T}(t+\Delta t) &= \exp\left[\boldsymbol{A}\Delta t\right]\boldsymbol{T}(t) + \int_t^{t+\Delta t} \exp\left[\boldsymbol{A}(t+\Delta t-s)\right]\boldsymbol{B}\boldsymbol{U}(s)ds \\ &\quad + \int_t^{t+\Delta t} \exp\left[\boldsymbol{A}(t+\Delta t-s)\right]d\boldsymbol{w}(s). \end{aligned} \tag{A.4}$$

With $\boldsymbol{U}(t)$ constant during the sampling interval the discrete time model takes the simplified expression

$$\boldsymbol{T}(t+\Delta t) = \boldsymbol{\Phi}(\Delta t)\boldsymbol{T}(t) + \boldsymbol{\Gamma}(\Delta t)\boldsymbol{U}(t) + \boldsymbol{v}(t;\Delta t), \tag{A.5}$$

where

$$\Phi(\Delta t) = \exp[\boldsymbol{A}\Delta t] \tag{A.6}$$

$$\Gamma(\Delta t) = \int_0^{\Delta t} \exp[\boldsymbol{A}s]\,\boldsymbol{B}ds \tag{A.7}$$

$$v(t;\Delta t) = \int_t^{t+\Delta t} \exp[\boldsymbol{A}(t+\Delta t-s)]\,d\boldsymbol{w}(s). \tag{A.8}$$

When $\{\boldsymbol{w}(t)\}$ is a Wiener process $\boldsymbol{v}(t;\Delta t)$ will consist of zero mean, independent normal distributed random variables with covariance (Åström, 1970; Gard, 1988)

$$\boldsymbol{R}_v(\Delta t) = E[\boldsymbol{v}(t;\Delta t)\boldsymbol{v}(t;\Delta t)^{\mathrm{T}}] = \int_0^{\Delta t} \boldsymbol{\Phi}(s)\boldsymbol{R}_w\boldsymbol{\Phi}^{\mathrm{T}}(s)ds. \tag{A.9}$$

When the sampling interval is constant the discrete time model is notationally simplified to the following difference equation

$$\boldsymbol{T}(t+1) = \boldsymbol{\Phi}\boldsymbol{T}(t) + \boldsymbol{\Gamma}\boldsymbol{U}(t) + \boldsymbol{v}(t). \tag{A.10}$$

$\boldsymbol{\Phi}$ and $\boldsymbol{\Gamma}$ are in this situation constant for all samples, and the length of the sampling interval is normalized to one time unit.

A.2 MAXIMUM LIKELIHOOD ESTIMATION

The likelihood function is given as the joint probability density function for all the observations, for a given set of parameter values, i.e.,

$$L'(\theta, \mathbf{T}_r(t_n)) = p(\mathbf{T}_r(t_n)|\theta). \tag{A.11}$$

By repeated application of the definition of conditional probability (Grimmett & Stirzaker, 1982) the likelihood function can be expressed as

$$\begin{aligned} L'(\theta, \mathbf{T}_r(t_n)) &= p(\mathbf{T}_r(t_n)|\mathbf{T}_r(t_{n-1}), \theta)p(\mathbf{T}_r(t_{n-1})|\theta) \\ &= \left(\prod_{i=1}^{n} p(\mathbf{T}_r(t_i)|\mathbf{T}_r(t_{i-1}), \theta)\right) p(\boldsymbol{T}_r(t_0)|\theta) \end{aligned} \tag{A.12}$$

where $\mathbf{T}_r(t_i)$ is a data-structure containing all observations up to and including time t, i.e.,

$$\mathbf{T}_r(t_i) = [\boldsymbol{T}_r(t_i), \boldsymbol{T}_r(t_{i-1}), \ldots, \boldsymbol{T}_r(t_1), \boldsymbol{T}_r(t_0)]^{\mathrm{T}} \tag{A.13}$$

$\boldsymbol{\theta}$ is a vector of all unknown parameters in $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{R}_w$ and $\boldsymbol{R}_e$.

The assumption of both $\boldsymbol{w}(t)$ and $\boldsymbol{e}(t_i)$ being normal distributed implies that the joint density of all observations conditional on the unobserved initial value $\boldsymbol{T}_r(t_0)$ is normal as well. In order to parametrize the conditional density, the one-step conditional mean and conditional variance are introduced as

$$\hat{\boldsymbol{T}}_r(t_i|t_{i-1}) = E\left[\boldsymbol{T}_r(t_i)|\mathbf{T}_r(t_{i-1}), \boldsymbol{\theta}\right] \tag{A.14}$$

and

$$\boldsymbol{R}(t_i|t_{i-1}) = V\left[\boldsymbol{T}_r(t_i)|\mathbf{T}_r(t_{i-1}), \boldsymbol{\theta}\right], \tag{A.15}$$

respectively. Hence, the joint probability density conditional on $\boldsymbol{T}_r(t_0)$ is written as

$$L(\boldsymbol{\theta}; \mathbf{T}_r(t_n)) = \prod_{i=1}^{n} (2\pi)^{-\frac{m}{2}} (\det \boldsymbol{R}(t_i|t_{i-1}))^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\varepsilon}^{\mathrm{T}}(t_i)\boldsymbol{R}(t_i|t_{i-1})^{-1}\boldsymbol{\varepsilon}(t_i)\right), \tag{A.16}$$

where m is the dimension of $\boldsymbol{T}_r(t_i)$ and $\boldsymbol{\varepsilon}(t_i)$ is the one-step prediction error given by

$$\boldsymbol{\varepsilon}(t_i) = \boldsymbol{T}_r(t_i) - \hat{\boldsymbol{T}}_r(t_i|t_{i-1}). \tag{A.17}$$

The logarithm of the conditional likelihood function is given by

$$\begin{aligned} \log L(\boldsymbol{\theta}; \mathbf{T}_r(t_n)) \;=\; & -\frac{1}{2}\left(\sum_{i=1}^{n} \log\det \boldsymbol{R}(t_i|t_{i-1}) + \boldsymbol{\varepsilon}^{\mathrm{T}}(t_i)\boldsymbol{R}(t_i|t_{i-1})^{-1}\boldsymbol{\varepsilon}(t_i)\right) \\ & + c(n, m), \end{aligned} \tag{A.18}$$

where $c(n, m)$ is a component depending alone on n and m.

A.3 THE KALMAN FILTER

The conditional mean and conditional variance entering the log-likelihood function (A.18) are calculated recursively by using the Kalman filter consisting of formulas for prediction and reconstruction of mean and variance. The prediction formulas evolve simply from the model assumption, and the reconstruction formulas are obtained by using linear projection rules on normal distributed random variables.

PREDICTION

$$\hat{T}(t_{i+1}|t_i) = \Phi(t_i, t_{i+1})\hat{T}(t_i|t_i) + \Gamma(t_i, t_{i+1})U(t_i) \tag{A.19}$$

$$\hat{T}_r(t_{i+1}|t_i) = C\hat{T}(t_i|t_{i+1}) \tag{A.20}$$

$$P(t_{i+1}|t_i) = \Phi(t_i, t_{i+1})P(t_i|t_i)\Phi^{\mathrm{T}}(t_i, t_{i+1}) + R_v(t_i, t_{i+1}) \tag{A.21}$$

$$R(t_{i+1}|t_i) = CP(t_{i+1}|t_i)C^{\mathrm{T}} + R_e, \tag{A.22}$$

where $P(t_{i+1}|t_i)$ is the covariance of the state prediction error. These formulas rely on the assumption that the input is constant and equal to $U(t_i)$ throughout the sampling interval $t_i \rightarrow t_{i+1}$.

RECONSTRUCTION The update or the reconstruction of mean and variance using the newly recorded observation at time t_i is

$$\hat{T}(t_i|t_i) = \hat{T}(t_i|t_{i-1}) + K(t_i)\varepsilon(t_i) \tag{A.23}$$

$$P(t_i|t_i) = P(t_i|t_{i-1}) - K(t_i)R(t_i|t_{i-1})K^{\mathrm{T}}(t_i), \tag{A.24}$$

where the Kalman gain is given by

$$K(t_i) = P(t_i|t_{i-1})C^{\mathrm{T}}R^{-1}(t_i|t_{i-1}). \tag{A.25}$$

The Kalman filter is the remedy for obtaining the sequence of one-step prediction errors along with the sequence of their conditional variances. These sequences constitute the parameter dependent part of the log-likelihood criterion. It is clear that the log-likelihood depends on the entries of Φ, Γ

and $\boldsymbol{C}$ as well as of $\boldsymbol{R}_w$ and $\boldsymbol{R}_e$. The Kalman gain determines the feedback from the prediction error vector to the state vector.

To start the recursive alternating calculation of predictions and reconstructions it is necessary to supply with a prior mean and variance constituting a complete characterization of the *a priori* probability density $\boldsymbol{P}_r(t_0)$

$$\boldsymbol{\mu}_0 = \hat{\boldsymbol{T}}(t_0|t_0) \tag{A.26}$$

$$\boldsymbol{V}_0 = \boldsymbol{P}(t_0|t_0). \tag{A.27}$$

A.4 STATIONARITY

Subject to the condition that $\boldsymbol{\Phi}$, $\boldsymbol{\Gamma}$, $\boldsymbol{C}$, $\boldsymbol{R}_v$ and $\boldsymbol{R}_e$ are constant supplemented with weak technical conditions (Ljung & Söderström, 1983) the Kalman gain will converge to a limit $\bar{\boldsymbol{K}}$, the entries of which depend on the matrices listed. This stationary gain can be determined from the equations (bar indicating limit values)

$$\bar{\boldsymbol{P}} = \boldsymbol{\Phi}\bar{\boldsymbol{P}}\boldsymbol{\Phi}^{\mathrm{T}} + \boldsymbol{R}_v - \bar{\boldsymbol{K}}\bar{\boldsymbol{R}}\bar{\boldsymbol{K}}^{\mathrm{T}} \tag{A.28}$$

$$\bar{\boldsymbol{R}} = \boldsymbol{C}\left(\boldsymbol{\Phi}\bar{\boldsymbol{P}}\boldsymbol{\Phi}^{\mathrm{T}} + \boldsymbol{R}_v\right)\boldsymbol{C}^{\mathrm{T}} + \boldsymbol{R}_e \tag{A.29}$$

$$\bar{\boldsymbol{K}} = \bar{\boldsymbol{P}}\boldsymbol{C}^{\mathrm{T}}\bar{\boldsymbol{R}}^{-1}, \tag{A.30}$$

where $\bar{\boldsymbol{P}}$ is the limit value of the $\boldsymbol{P}(t_i|t_i)$. With appropriate initial values of $\boldsymbol{\mu}_0$ and $\boldsymbol{V}_0$ the stationary values $\bar{\boldsymbol{P}}$, $\bar{\boldsymbol{R}}$ and $\bar{\boldsymbol{K}}$ apply from the first step of the propagation through the set of data.

In the situation, where stationarity is prevalent, the recursions of the Kalman filter for update of covariances $\boldsymbol{P}$ and $\boldsymbol{R}$ are superfluous. That is, the required prediction recursions of the Kalman filter in stationarity are

$$\hat{\boldsymbol{T}}(t_{i+1}|t_i) = \boldsymbol{\Phi}\hat{\boldsymbol{T}}(t_i|t_i) + \boldsymbol{\Gamma}\boldsymbol{U}(t_i) \tag{A.31}$$

$$\hat{\boldsymbol{T}}_r(t_{i+1}|t_i) = \boldsymbol{C}\hat{\boldsymbol{T}}(t_{i+1}|t_i), \tag{A.32}$$

and the only remaining recursion of the reconstruction is

$$\hat{\boldsymbol{T}}(t_i|t_i) = \hat{\boldsymbol{T}}(t_i|t_{i-1}) + \bar{\boldsymbol{K}}\boldsymbol{\varepsilon}(t_i), \tag{A.33}$$

where $\boldsymbol{\varepsilon}(t_i)$ is the one-step prediction error.

The stationary Kalman gain matrix $\bar{\boldsymbol{K}}$ can be determined from Equations (A.28)-(A.30) for given values of $\boldsymbol{\Phi}$, $\boldsymbol{R}_v$ and $\boldsymbol{R}_e$. This means that also when stationarity is assumed to prevail from the first step, the log-likelihood is function of $\boldsymbol{\Phi}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{C}$ as well as of $\boldsymbol{R}_v$ and $\boldsymbol{R}_e$. However, from the recursions of the Kalman filter, remaining in stationarity, it is clear that the two covariance matrices are determinative only for the feedback from prediction errors, implemented in the Kalman gain. Consequently, the existence of $\boldsymbol{R}_v$ and $\boldsymbol{R}_e$, and the parameters therein, can be neglected if instead the entries of the stationary Kalman gain constitute the parameterization determining the feedback from prediction errors to model states.

The log-likehood function in (A.18) is simplified for stationary prediction error covariance $\bar{\boldsymbol{R}}$, viz.

$$\begin{aligned} \log L(\boldsymbol{\theta}, \bar{\boldsymbol{R}}; \mathbf{T}_r(t_n)) &= -\frac{n}{2}\log\det\bar{\boldsymbol{R}} \\ &\quad -\frac{1}{2}\sum_{i=1}^{n}\boldsymbol{\varepsilon}^{\mathrm{T}}(t_i)\bar{\boldsymbol{R}}^{-1}\boldsymbol{\varepsilon}(t_i) + c(n,m), \end{aligned} \tag{A.34}$$

where $\boldsymbol{\theta}$ in this case is a vector of the parameters to be estimated in $\boldsymbol{A}$, $\boldsymbol{B}$ and $\bar{\boldsymbol{K}}$. The recursions for the Kalman filter in stationarity show that $\bar{\boldsymbol{R}}$ does not influence the predictions, and prediction errors, of the state space model. Taking the derivative of (A.34) with respect to each of the independent entries of $\bar{\boldsymbol{R}}$, and solving the system of equations given by the assumption that for optimum values the derivative is zero give the estimate of $\bar{\boldsymbol{R}}$. This analytic solution can be inserted in (A.34) to give a simplified optimization problem.

The Kalman gain gives precisely the parameterization, with the proper number of degrees of freedom, which is necessary to obtain the feedback from prediction errors, and which can be estimated. Therefore the number of degrees of freedom in the parameterization of $\boldsymbol{R}_v$ and $\boldsymbol{R}_e$ equals

the number of entries in $\boldsymbol{K}$. The number of elements in the two covariance matrices always exceeds the number of elements in the Kalman gain. Therefore when wanting to estimate the parameters of the covariance matrices it is necessary to impose restrictions on their entries. The actual parameterization that gives the full number of degrees of freedom in the feedback from prediction errors is, however, complicated, advocating the estimation of Kalman gain instead of $\boldsymbol{R}_v$ and $\boldsymbol{R}_e$.

A.5 NUMERICAL MAXIMIZATION

The ML estimation is carried out by numerical minimization either of the log-likelihood function (A.18), or of the simplified log-likelihood based on stationary $\boldsymbol{R}$. In the latter case the Kalman gain elements are estimated instead of the covariance matrices $\boldsymbol{R}_w$ and $\boldsymbol{R}_e$. The minimization is implemented in the software program CTLSM, see Melgaard & Madsen (1991).

The inverse covariance matrix of the parameter estimates is estimated by the negative, second order derivative of the log-likelihood function with respect to the parameters. That is, element (l, k) of the inverse covariance matrix is approximated by

$$h_{l,k} \approx -\left(\frac{\partial^2}{\partial\theta_l \partial\theta_k} \log L(\boldsymbol{\theta}; \mathbf{T}_r(t_n))\right)_{|\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \tag{A.35}$$

An approximation of the second order derivative is supplied by CTLSM.

REFERENCES

Abraham, B. & Ledolter, J. (1983). *Statistical Methods for Forecasting.* Applied Probability and Statistics. Wiley & Sons, New York.

Akaike, H. (1974). A New Look at The Statistical Model Identification. *IEEE Transactions on Automatic Control*, **19**, 716–722.

Alfa-Laval (1985). Technical Description of Solded Plate Heat Exchanger, Type CB25..

Allende, H. & Heiler, S. (1992). Recursive Generalized M Estimates for Autoregressive Moving-Average Models. *Journal of Time Series Analysis*, **13**(1), 1–18.

Åström, K. J. (1970). *Introduction to Stochastic Control Theory.* Academic Press, New York.

Auestad, B. & Tjøstheim, D. (1990). Identification of Nonlinear Time Series: First Order Characterization and Order Determination. *Biometrika*, **77**(4), 669–687.

Benonysson, A. (1991). *Dynamic Modelling and Operational Optimization of District Heating Systems.* Ph.D. thesis, Laboratory of Heating and Air Conditioning, Technical University of Denmark, Lyngby.

Bittanti, S., Cividini, A., & Scattolini, R. (1982). Identification of a Liquid-Saturated Steam Heat Exchanger. In *Proceedings of the 6th IFAC Symposium on Identification and System Estimation*, pp. 168–173 Arlington, Virginia, USA.

Box, G. & Jenkins, G. (1976). *Time Series Analysis, Forecasting and Control.* Holden-Day, San Francisco.

Campbell, K. (1982). Recursive Estimation of M-estimates for the Parameterts of a Finite Autoregressive Process. *The Annals of Statistics*, **10**(2), 442–453.

Cipra, T. (1992). Robust Exponential Smoothing. *Journal of Forecasting*, **11**, 57–69.

Denby, L. & Martin, R. (1979). Robust Estimation of the First-Order Autoregressive Parameter. *Journal of the American Statistical Association*, **74**(365), 140–146.

Edlund, P.-O. (1989). *Preliminary Estimation of Transfer Function Weights - A Two-Step Regression Approach.* Ph.D. thesis, Stockholm School of Economics, The Economic Research Institute, Stockholm.

Englund, J.-E., Holst, U., & Ruppert, D. (1988). Recursive M-estimators of Location and Scale for Dependent Sequences. *Scandinavian Journal of Statistics*, **15**, 147–159.

Englund, J.-E., Holst, U., & Ruppert, D. (1989). Recursive M-estimators for Stationary Strong Mixing Processes - A representation theorem and asymptotic distributions. *Stochastic Processes and their Applications*, **31**, 203–222.

Englund, J.-E. (1991). Recursive Versions of the Algorithm by Krasker and Welsch. *Sequential Analysis*, **10**(4), 211–234.

Ericsson, O. & Haglund, O. (1991). Load Prediction in Power Network (in Swedish: Belastningsprognosering i Elnät med Hjälp av Neurala Nätverk). Master's thesis, Department of Computer Science, Lund Institute of Technology, Lund.

Gard, T. (1988). *Introduction to Stochastic Differential Equations.* Marcel Dekker, New York.

Goodwin, G. & Salgado, M. (1989). A Stochastic Embedding Approach for Quantifying Uncertainty in the Estimation of Restricted Complexity Models. *International Journal of Adaptive Control and Signal Processing*, **3**, 333–356.

Grimmett, G. & Stirzaker, D. (1982). *Probability and Random Processes.* Clarendon Press, Oxford.

Gummérus, P. (1988). Modelling of Subcentrals in District Heating Systems (in Swedish: Modellering av Abonnentcentraler i Fjärrvärmenät. Department of Energetics, Chalmers Technical University, Gothenburg.

Hampel, F., Ronchetti, E., Rousseuw, P., & Stahel, W. (1986). *Robust Statistics.* Wiley & Sons, New York.

Hampel, F. (1971). A General Qualitative Definition of Robustness. *The Annals of Mathematical Statistics*, **42**(6), 1887–1896.

Härdle, W. & Marron, J. (1985). Optimal Bandwidth Selection in Nonparametric Regression Function Estimation. *The Annals of Statistics*, **13**(4), 1465–1481.

Härdle, W. & Vieu, P. (1992). Kernel Regression Smoothing of Time Series. *Journal of Time Series Analysis*, **13**(3), 209–224.

Härdle, W. (1990). *Applied Nonparametric Regression.* Econometric Society Monographs. Cambridge University Press, Cambridge.

Harvey, A. (1989). *Forecasting, Structural Time Series Models and the Kalman filter.* Cambridge University Press, Cambridge.

Henningsen, A. & Sejling, K. (1990). An Approach to Quantification of Uncertainty in Recursive Estimation of Restricted Complexity Models. In *Proceedings of the Nordic CACE Symposium*, pp. 6.6–6.10 Lyngby.

Hinich, M. (1982). Testing for Gaussianity and Linearity of a Stationary Time Series. *Journal of Time Series Analysis*, **3**(3), 169–176.

Holman, J. P. (1981). *Heat Transfer* (5 edition). McGraw-Hill, Singapore.

Holst, J. & Ekelund, J. (1987). Adaptive Prediction of the Load in a Power Network (in Swedish: Adaptiv Prediktion av Belastningen på et Kraftnät). TFMS-3043, Department of Mathematical Statistics, Lund Institute of Technology, Lund.

Holst, J. & Jonsson, G. (1984). Adaptive Short Term Prediction of Power Load. In *Proceedings of the 9th IFAC World Congress on Automatic Control*, pp. 2103–2108 Budapest.

Holst, J. (1977). *Adaptive Prediction and Recursive Estimation.* Ph.D. thesis, TFRT-1013, Department of Automatic Control, Lund Institute of Technology, Lund.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer Feedforward Networks are Universal Approximators. *Neural Networks*, **2**, 359–366.

Huber, P. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, **35**, 73–101.

Huber, P. (1973). Robust Regression: Asymptotics, Conjectures and Monte Carlo. *The Annals of Statistics*, **1**(5), 799–821.

Huber, P. (1981). *Robust Statistics.* Wiley & Sons, New York.

Humo, E. & Popovic, M. (1982). On an Approach to Dynamic Analysis of Certain Class of Lumped Parameter Systems. In *Proceedings of the 6th IFAC Symposium on Identification and System Estimation*, pp. X–X Arlington, Virginia, USA.

IMSL Inc., Houston (1987a). *IMSL Math/Library - Fortran Subroutines for Mathematical Applications.*

IMSL Inc., Houston (1987b). *IMSL Stat/Library - Fortran Subroutines for Statistical Analysis.*

Ito, A. & Masubuchi, M. (1978). Dynamic Analysis and Model Control of Plate Heat Exchange Systems. In *Proceedings of the 7th IFAC World Congress on Automatic Control*, pp. 343–349 Helsinki.

Jonsson, G. & Holst, J. (1989). Statistical Parameter Estimation of a Counter Flow Heat Exchanger. In *International Symposium on District Heat Simulation* Reykjavik. Nordic Council of Ministers Energy Research Cooperation.

Jonsson, G. & Palsson, O. (1991). Use of Empirical Relations in the Parameters of Heat Exchanger Models. *Industrial & Engineering Chemical Research*, **30**(6), 1193–1199.

Jonsson, G. & Palsson, O. (1993). An Application of Extended Kalman Filtering to Heat Exchanger Models. *Journal of Dynamic Systems, Measurement and Control, to be published*, X.

Jonsson, G., Palsson, O., & Sejling, K. (1992). Modeling and Parameter Estimation of Heat Exchangers - A Statistical Approach. *Journal of Dynamic Systems, Measurement and Control*, **114**(6), 673–679.

Jonsson, G. (1990). *Parameter Estimation in Models of Heat Exchangers and Geothermal Reservoirs.* Ph.D. thesis, TFMS-1006, Department of Mathematical Statistics, Lund Institute of Technology, Lund.

Kakac, S., Shah, R. K., & Bergles, A. E. (1983). *Low Reynolds Number Flow Heat Exchanger.* Hemisphere Publishing Corporation, Washington.

Klimko, L. & Nelson, P. (1978). On Conditional Least Squares Estimation for Stochastic Processes. *The Annals of Statistics*, **6**(3), 629–642.

Krasker, W. & Welsch, R. (1982). Efficient Bounded-Influence Regression Estimation. *Journal of the American Statistical Association*, **77**(379), 595–603.

Kuh, E. & Samarov, A. (1986). Robust Recursive Estimation and Detection of Shifts in Regression. In *Compstat 1986*, pp. 217–222. Physica-Verlag, Heidelberg.

Ljung, L. & Söderström, T. (1983). *Theory and Practice of Recursive Identification.* MIT Press, Cambridge, Massachusetts.

Ljung, L. (1987). *System Identification, Theory for the User.* Prentice-Hall, Englewood Cliffs, New Jersey.

Ljung, G. (1993). On Outlier Detection in Time Series. *Journal of the Royal Statistical Society, Series B*, **55**(2), 559–567.

Madsen, H., Nielsen, A., & Saxhof, B. (1992a). Identification of Models for Heat Dynamics of Buildings. Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby.

Madsen, H., Palsson, O., Sejling, K., & Søgaard, H. (1992b). *Models and Methods for Optimization of District Heating Systems, Part II: Models and Control Methods.* Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby.

Madsen, H. (1985). *Statistically Determined Dynamical Models for Climate Processes.* Ph.D. thesis, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby.

Madsen, H. (1989). *Time Series Analysis* (Preliminary edition). Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby.

Malmström, B. (1991). *Computerized Operation-optimal Load Distribution for Cogeneration Production.* Ph.D. thesis, Department of Energetics, Chalmers Technical University, Gothenburg.

Martin, R. & Masreliez, C. (1975). Robust Estimation via Stochastic Approximation. *IEEE Transactions on Information Theory*, **IT-21**(3), 263–271.

Martin, R. D. & Yohai, V. (1985). Robustness in Time Series and Estimating ARMA Models. In Hannan, E., Krishnaiah, P., & Rao, M. (Eds.), *Handbook of Statistics*, Vol. 5, pp. 119–155. Elsevier Science, Amsterdam.

Martin, R. D. (1981). Robust Methods for Time Series. In Findley, D. F. (Ed.), *Applied Time Series Analysis II*, pp. 683–760. Academic Press.

Masada, G. Y. & Wormley, D. N. (1982). Evaluation of Lumped Parameter Heat Exchanger Dynamic Models. *ASME 82-WA/DSC-16.*

Masreliez, C. (1975). Approximate Non-Gaussian Filtering with State and Observation Relations. *IEEE Transactions on Automatic Control*, **20**, 107–110.

Másson, E. & Wang, Y. (1990). Introduction to Computation and Learning in Neural Networks. *European Journal of Operational Research*, **47**, 1–28.

McDougall, A. (1994). Robust Methods for Recursive ARMA Estimation. *Journal of the Royal Statistical Society, Series B*, **56**(1), X–X.

Melgaard, H. & Madsen, H. (1991). The Mathematical and Numerical Methods used in CTLSM. Tech. rep. 7, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby.

Mellentin, G. (1992). Non-Linear Time Series Analysis (in Danish:Ikke-Lineær Tidsrækkeanalyse). Master's thesis, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby.

Palsson, O. (1989). Time Continuous Dynamic Models of Heat Exchangers (in Danish: Tidskontinuerte Dynamiske Modeller for Varmevekslere). Master's thesis, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby.

Palsson, O. (1990). Dynamic Models for Heat Exchangers (in Danish: Dynamiske Modeller for Varmevekslere. Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby.

Parkum, J., Poulsen, N., & Holst, J. (1992). Recursive Forgetting Algorithms. *International Journal of Control*, **55**(1), 109–128.

Parkum, J. (1992). *Recursive Identification of Time-Varying Systems*. Ph.D. thesis, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby.

Poljak, B. & Tsypkin, J. (1980). Robust Identification. *Automatica*, **16**, 53–63.

Poulsen, N. & Holst, J. (1982). Robust Self Tuning Controllers in Nonstationary Situations. *Ricerche di Automatica*, **13**, 197–217.

Priestley, M. (1988). *Non-linear and Non-stationary Time Series Analysis.* Academic Press, London.

Rao, T. & Gabr, M. (1980). A Test for Linearity of Stationary Time Series. *Journal of Time Series Analysis*, 1(1), 145–158.

Ripley, B. (1992). Statistical Aspects of Neural Networks. Department of Statistics, University of Oxford, To appear in the proceedings to be published by Chapman & Hall in January 1993, Oxford.

Robinson, P. (1983). Nonparametric Estimators for Time Series. *Journal of Time Series Analysis*, 4(3), 185–207.

Salgado, M., de Souza, C., & Goodwin, G. (1990). Qualitive Aspects of the Distribution of Errors in Least Squares Estimation. *Automatica*, 26(1), 97–101.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), 461–464.

Seather, S. (1992). The Performance of Six Popular Bandwidth Selection Methods on Some Real Data Sets. *Computational Statistics*, 7(3), 225–250.

Sejling, K., Madsen, H., & Holst, J. (1988). Adaptive k-step Predictions of the Heat Consumption in a District Heating System. Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby.

Sejling, K. (1987). Adaptive Prediction Models for District Heating Systems (in Danish: Adaptive Prognosemodeller for Fjernvarmesystemer). Master's thesis, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby.

Shoureshi, R. & Paynter, H. M. (1983). Simple Methods for Dynamics and Control of Heat Exchangers. In *Proceedings of the American Control Conference*, pp. 1294–1298.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, New York.

Søgaard, H. T. (1993). *Embedded Models in Time Series Analysis.* Ph.D. thesis, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby.

Steiner, M. (1989). Low Order Dynamic Models of Heat Exchangers. In *International Symposium on District Heat Simulation* Reykjavik. Nordic Council of Ministers Energy Research Cooperation.

Tong, H. (1990). *Non-linear Time Series, A Dynamical System Approach.* Clarendon Press, Oxford.

Waldemark, J. (1993). *Neural Network Techniques in District Heating Applications.* Ph.D. thesis, Department of Applied Physics, University of Umeå, Umeå.

Weigend, A., Huberman, B., & Rummelhart, D. (1990). Predicting the Future: A Connectionist Approach. *International Journal of Neural Systems*, **1**(3), 193–209.

West, M. (1981). Robust Sequential Approximate Baysian Estimation. *Journal of the Royal Statistical Society, Series B*, **43**(2), 157–166.

Wiklund, H. (1989). *Accumulation and Prediction of Load in District Heating Systems (in Swedish: Ackumulering och Prediktering av Fjärrvärmelast).* Ph.D. thesis, Department of Applied Physics, University of Umeå, Umeå.

IMSOR Ph.D. Theses

1. **Sigvaldason, Helgi.** (1963). *Beslutningsproblemer ved et hydrotermisk elforsyningssystem.* 92 pp.

2. **Nygaard, Jørgen.** (1966). *Behandling af et dimensioneringsproblem i telefonien.* 157 pp.

3. **Krarup, Jakob.** (1967). *Fixed-cost and other network flow problems as related to plant location and to the design of transportation and computer systems.* 159 pp.

4. **Hansen, Niels Herman.** (1967). *Problemer ved forudsigelse af lydhastighed i danske farvande. Analyse af et stokastisk system. Del 1: Tekst. Del 2: Figurer og tabeller.* 104 pp. + 95 pp.

5. **Larsen, Mogens E.** (1968). *Statistisk analyse af elementære kybernetiske systemer.* 210 pp.

6. **Punhani, Amrit Lal.** (1968). *Decision problems in connection with atomic power plants.* 133 pp.

7. **Clausen, Svend.** (1969). *Kybernetik, systemer og modeller.* 205 pp.

8. **Vidal, R.V. Valqui.** (1970.) *Operations research in production planning. Interconnections between production and demand. Volume 1-2.* 321 pp.

9. **Bilde, Ole.** (1970). *Nonlinear and discrete programming in transportation, location and road design. Volumes 1-2.* 291 pp.

10. **Rasmusen, Hans Jørgen.** (1972). *En decentraliseret planlægningsmodel.* 185 pp.

11. **Dyrberg, Christian.** (1973). *Tilbudsgivning i en entreprenørvirksomhed.* 158 pp.

12. **Madsen, Oli B.G.** (1973). *Dekomposition og matematisk programmering.* 271 pp.

13. **Dahlgaard, Peter.** (1973). *Statistical aspects of tide prediction. Volume 1. Volume 2: Figures and tables.* 202 pp. + 170 pp.

14. **Spliid, Henrik.** (1973). *En statistisk model for stormflodsvarsling.* 205 pp.

15. **Pinochet, Mario.** (1973). *Operations research in strategic transportation planning. The decision process in a multiharbour system.* 374 pp.

16. **Christensen, Torben.** (1973). *Om semi-markov processer. Udvidelser og anvendelser inden for den sociale sektor.* 239 pp.

17. **Jacobsen, Søren Kruse.** (1973). *Om lokaliseringsproblemer, modeller og løsninger.* 355 pp.

18. **Marqvardsen, Hans.** (1973). *Skemalægning ved numerisk simulation.* 222 pp.

19. **Mortensen, Jens Hald.** (1974). *Interregionale godstransporter. Teoridannelser og modeller.* 223 pp.

20. **Severin, Juan Melo.** (1974). *Introduction to operations research in systems synthesis. A chemical process design synthesis application.* 249 pp.

21. **Spliid, Iben & Uffe Bundgaard-Jørgensen.** (1974). *Skitse til*
22. *en procedure for kommunalplanlægning.* 544 pp.

23. **Mosgaard, Christian.** (1975). *International planning in disaster situations.* 187 pp.

24. **Holm, Jan.** (1975). *En optimeringsmodel for kollektiv trafik.* 246 pp.

25. **Jesson, Pall.** (1975). *Stokastisk programmering. Del 1: Modeller. Del 2: Metodologiske overvejelser og anvendelser.* 333 pp.

26. **Iversen, Villy Bæk.** (1976). *On the accuracy in measurements of time intervals and traffic intensities with application to teletraffic and simulation.* 202 pp.

27. **Drud Arne.** (1976). *Methods for control of complex dynamic systems. Illustrated by econometric models.* 209 pp.

28. **Togsverd, Tom.** (1976). *Koordinering af kommunernes ressourceforbrug.* 295 pp.

29. **Jensen, Olav Holst.** (1976). *Om planlægning af kollektiv trafik. Operationsanalytiske modeller og løsningsmetoder.* 321 pp.

30. **Beyer, Jan E.** (1976). *Ecosystems. An operational research approach.* 315 pp.

31. **Bille, Thomas Bastholm.** (1977). *Vurdering af Egnsudviklingsprojekter. Samspil mellem benifit-cost analyse og den politiske vurdering i en tid under forandring.* 260 pp.

32. **Holst, Erik.** (1979). *En statistisk undersøgelse af tabletserier.* 316 pp.

33. **Aagaard-Svendsen, Rolf.** (1979). *Econometric methods and Kalman filtering.* 300 pp.

34. **Hansen, Steen.** (1979). *Project control by quantitative methods.* 230 pp.

35. **Scheufens, Ernst Edvard.** (1980). *Statistisk analyse og kontrol af tidsafhængige vandkvalitetsdata.* 152 pp.

36. **Lyngvig, Jytte.** (1981). *Samfundsøkonimisk planægning.* 252 pp.

37. **Troelsgård, Birgitte.** (1981). *Statistisk bestemmelse af modeller for rumlufttemperatur.* 213 pp.

38. **Raft, Ole.** (1981). *Delivery planning by modular algorithms.* 220 pp.

39. **Jensen, Sigrid M.** (1981). *Analyse af interregionale togrejser. + Figurer og appendices.* 212 pp. + 174 pp.

40. **Ravn, Hans.** (1982). *Technology and underdevelopment. The case of Mexico.* 376 pp.

41. **Hansen, Sten.** (1983). *Phase-type distributions in queueing theory.* 209 pp.

42. **Ferreira, Jose A.S.** (1984). *Optimal control of discrete-time systems with applications.* 252 pp.

43. **Behrens, Jens Christian.** (1985). *Mathematical modelling of aquatic ecosystems applied to biological waste water treatment + Appendix 1-2.* 32 pp. + 389 pp. + 180 pp.

44. **Poulsen, Niels Kjølstad.** (1985). *Robust self-tuning controllers.* 240 pp.

45. **Madsen, Henrik.** (1985). *Statistically determined dynamic models for climate processes. Part 1-2.* 428 pp.

46. **Sørensen, Bo.** (1986). *Interactive distribution planning.* 253 pp.

47. **Lethan, Helge B.** (1986). *Løsning af store kombinatoriske problemer.* 173 pp.

48. **Boelskifte, Søren.** (1988). *Dispersion and current measurements. An investigation based on time series analysis and turbulence models.* Risø-M-2566. 154 pp.

49. **Nielsen, Bo Friis.** (1988). *Modelling of multiple access systems with phase type distributions.* 253 pp.

50. **Christensen, John M.** (1988). *Project planning and analysis. Methods for assessment of rural energy projects in deveoping countries.* Risø-M-2706. 158 pp.

51. **Olsen, Klaus Juel.** (1988). *Texture analysis of ultrasound images of livers.* 162 pp.

52. **Holst, Helle.** (1988). *Statistisk behandling af nærinfrarøde reflektionsmålinger.* 309 pp. + app.

53. **Knudsen, Torben.** (1989). *Start/stop strategier for vind-diesel systemer.* 275 pp.

54. **Ersbøll, Bjarne Kjær.** (1989). *Transformations and classifications of remotely sensed data. Theory and geological cases.* 297 pp.

55. **Kragh, Anders Laage.** (1990). *Kø-netværksmodeller til analyse af FMS anlæg.* 205 pp.

56. **Hansen, Christian Kornerup.** (1991). *Statistical methods in the analysis of repairable systems reliability.* 56 pp.

57. **Parkum, Jens Ejnar.** (1992). *Recursive identification of time-varying systems.* 206 pp.

58. **Bilbo, Carl M.** (1992). *Statistical analysis of multivariate degradation models.* 167 pp.

59. **Carstensen, Jens Michael.** (1992). *Description and simulation of visual texture.* 234 pp.

60. **Halse, Karsten.** (1992). *Modeling and solving complex vehicle routing problems.* 372 pp.

61. **Hendricks, Elbert.** (1992). *Identification and estimation of nonlinear systems using physical modelling.* 273 pp.

62. **Windfeld, Kristian.** (1992). *Application of computer intensive data analysis methods to the analysis of digital images and spatial data.* 190 pp.

63. **Iwersen, Jørgen.** (1992). *Statistical control charts – Performance of Shewhart and CUSUM charts.* 320 pp.

64. **Olsson, Carsten Kruse.** (1993). *Image Processing Methods in Materials Science.* 274 pp.

65. **Sejling, Ken.** (1993). *Modelling and Prediction of Load in District Heating Systems.* 283 pp.