Die approbierte Originalversion dieser Dissertation ist an der Hauptbibliothek der Technischen Universität Wien aufgestellt (http://www.ub.tuwien.ac.at).

The approved original version of this thesis is available at the main library of the Vienna University of Technology (http://www.ub.tuwien.ac.at/englweb/).

DISSERTATION

Modeling and Control of Biomass-Fired District Heating Networks

ausgeführt zum Zwecke der Erlangung des akademischen Grades
eines Doktors der technischen Wissenschaften
unter der Leitung von

Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Martin Kozek

am Institut für Mechanik und Mechatronik
Arbeitsgruppe für Regelungstechnik und Prozessautomatisierung

eingereicht an der Technischen Universität Wien,
Fakultät für Maschinenbau

von

Dipl.-Ing. Stefan Grosswindhager
Baumgasse 16/19
1030 Wien

Geboren am 22.02.1984
Matr.Nr.: 0425904

Wien, im April 2013

Meinen Eltern

Kurzfassung

In dieser Dissertation wird ein allgemeines Framework für die Modellierung und Regelung von Biomasse-Fernwärmenetzen vorgestellt. Für die wichtigsten Komponenten eines Fernwärmenetzes (Rohrnetz, Verbraucher und Biomassekraftwerk) werden physikalische und datengetriebene Modelle entwickelt sowie anspruchsvolle Regelalgorithmen entworfen.

Die Berechnung der Strömungsgeschwindigkeiten im Rohrnetz basiert auf einem effizienten graphentheoretischen Ansatz, der mit dem iterativen Newton Algorithmus kombiniert wird. Für die transient-thermische Modellierung wird QUICKEST, eine Finite-Differenzen-Methode 3. Ordnung, herangezogen. Neben dem Finiten-Differenzen-Modell werden zusätzlich datengetriebene bedingt-parametrische Modelle konstruiert, welche in ein prädiktives Regelkonzept eingebettet werden um die Vorlauftemperaturen an kritischen Knoten im Netz zu regeln. Closed-Loop Simulationen belegen, dass diese Modellstruktur die inhärente Nichtlinearität in Fernwärmenetzen auf Grund variabler Strömungsgeschwindigkeiten abbilden kann.

Es wird ferner gezeigt, dass der zyklische Wärmebedearf der Verbraucher mittels sogenannten SARIMA Prozessen und Struktur-Modellen in geeigneter Weise beschrieben werden kann. Beide Modellansätze können in eine Zustandsraumdarstellung überführt werden, in welcher eine Onlineberechnung von Prognosewerten mittels klassischen Kalman-Filters möglich ist. Exogene Einflüsse wie die Umgebungstemperatur werden durch eine stückweise lineare Funktion in beide Modellansätze integriert. Desweiteren wird ein nichtparametrisches Regressionsverfahren verwendet, um die Rücklauftemperatur an den Verbrauchern als Funktion der Umgebungstemperatur und einer sozialen Komponente zu modellieren.

Für die Beschreibung der wesentlichen Prozess-Charakteristika von Biomasse-Rostfeuerungen wird neben einem Greybox-Modell, welches weitgehend auf Massen- und Energiebilanzen basiert, auch ein datengetriebenes Takagi-Sugeno (TS) Fuzzy-Modell vorgestellt. Für die Wahl der Fuzzy-Partitionierung wird auf ein achsen-orthogonales, inkrementelles Partitionierungsschema zurückgegriffen. Die Modellvalidierung anhand von realen Messungen bestätigt die ausgezeichnete Performance beider Modelle. Die lokale Linearität von TS Fuzzy-Modellen erlaubt auch eine Verwendung in einer linear parameter-variierenden Regelung. Konkret wird ein Fuzzy-Modellprädiktiver Regelungsansatz vorgestellt, welcher gegenüber der klassichen modellprädiktiven Regelung signifikant bessere Resultate über den gesamten Betriebsbereich erzielt.

Abstract

In this thesis, a general framework for modeling and control of biomass-fired district heating networks is presented. For the main components of a district heating system (distribution network, consumer stations, and biomass furnace) first-principles models and data-driven models are developed as well as sophisticated control algorithms are designed.

Calculation of the hydraulic flow in the distribution network is accomplished efficiently by means of a graph-theoretical approach in combination with a Newton-based iterative algorithm. For modeling the thermal transients, the third order accurate QUICKEST finite-difference method is successfully applied. In addition to the finite-difference model, a data-driven approach using conditional parametric models is pursued. These models are embedded into an advanced predictive control scheme for the supply temperature at critical points in the network. Closed-loop simulations prove that this modeling framework is capable of handling the inherent nonlinearity in the response characteristics of district heating networks caused by varying flow rates.

It is further shown that the multi-periodic heat-demand of consumer stations can appropriately be described by a SARIMA process and by a structural model. Both models can be incorporated into the versatile state-space framework where classical Kalman recursion allows convenient calculation of on-line forecasting values. Moreover, exogenous influences such as ambient temperature are accounted for by a piece-wise linear function. For modeling the required return temperature at consumer stations a nonlinear static relationship between return temperature and ambient temperature plus social load is identified using a nonparametric regression framework.

To describe the process characteristics of the moving grate biomass furnace a simple grey-box model, largely based on mass and energy balances, and a data-driven Takagi-Sugeno (TS) fuzzy model are constructed. The fuzzy partitions of the individual TS fuzzy models are found by a sophisticated axis-orthogonal, incremental partitioning scheme. Validations with real measurements demonstrate good performance of the two proposed models. The local linearity of TS fuzzy models also allows the use as a linear parameter-varying model in model predictive control. In particular, a fuzzy model predictive control framework is introduced which provides increased performance on the operating envelope compared to classical control schemes.

Danksagung

Diese Dissertation entstand im Rahmen meiner Anstellung als Projektassistent am Institut für Mechanik und Mechatronik, Abteilung für Regelungstechnik und Prozessautomatisierung unter der Leitung von Prof. Dr. Martin Kozek in Zusammenarbeit mit VOIGT+WIPP Engineers GmbH und wurde großteils durch ein FFG-Projekt finanziert.

An erster Stelle möchte ich mich bei Prof. Dr. Stefan Jakubek für die Möglichkeit diese Arbeit zu verfassen und für die ausgesprochen angenehme und kollegiale Atmospähre am Institut bedanken.

Insbesondere gilt mein Dank Dr. Martin Kozek für die kompetente Leitung dieses Projekts und der fachlichen und menschlichen Unterstützung in jeglicher Hinsicht. Weiters möchte ich mich bei Em. O. Prof. Deistler Manfred bedanken, der sich freundlicherweise als Zweitleser und Begutachter zur Verfügung gestellt hat.

Mein Dank gilt auch dem Projektpartner VOIGT+WIPP, die in das Projekt entweder fachlich oder organisatorisch involviert waren. Allen voran möchte ich an dieser Stelle Dipl.-Ing. Andreas Voigt und Dr. Lukas Haffner erwähnen, die immer mit Rat und Tat zur Seite standen.

Zu guter Letzt stehe ich noch in der Schuld meiner Kollegen und Freunde am Institut - insbesondere Raphael Hofstädter, Andreas Fleck und Johannes Unger - die das Arbeitsklima positiv und motivierend mitgestaltet haben und nicht nur für interessante fachliche Gespräche, sondern auch für private Unternehmungen zur Verfügung standen. In der Danksagung sollte auch Em. O. Prof. Hanns-Peter Jörgl nicht unerwähnt bleiben, der die Verantwortung für die tägliche Verköstigung in der Mensa übernahm.

Der allergrößte Dank gilt jedoch meiner Familie, die mich während meiner gesamten Ausbildung immer unterstützte und mir diese erst ermöglichte.

Stefan Grosswindhager

Contents

Glossary **viii**

1 Introduction **1**
1.1 Motivation . . . 1
1.2 Literature Review, State-of-the-Art . . . 2
1.3 Main Contributions . . . 4
1.4 Structure of Work . . . 6

2 Main Components of a District Heating Network **7**
2.1 Distribution Network . . . 7
2.2 Consumer Stations . . . 8
2.3 Biomass District Heating Plant . . . 9

3 Modeling of the Distribution Network **11**
3.1 Physical Models . . . 11
3.1.1 The Hydraulic Model . . . 11
3.1.2 The Thermal Model . . . 16
3.1.3 Simulations . . . 22
3.2 Conditional Parametric Models . . . 27
3.2.1 General Model Structure . . . 27
3.2.2 Estimation . . . 28
3.2.3 Local Linear Model Tree (LOLIMOT) . . . 32
3.2.4 Identification Results . . . 33

4 Modeling of Heat Consumption **44**
4.1 Modeling and Prediction of Heat-Load . . . 44
4.1.1 Heat-Load Characteristics . . . 45
4.1.2 Statistical Methods . . . 47
4.1.3 Modeling and Identification . . . 51
4.1.4 Validation . . . 56
4.2 Modeling of Return Temperature . . . 60
4.2.1 Multivariate Adaptive Regression Splines (MARS) . . . 60
4.2.2 Estimation Results . . . 62

5 Modeling of the Biomass Combustion Plant **64**
5.1 First-Principles Model . . . 64
5.1.1 Introduction . . . 64
5.1.2 Process Description of a Moving Grate Biomass Furnace . . . 64
5.1.3 First-Principles Modeling . . . 65
5.1.4 Model Verification . . . 67

5.2 Data-Driven Model . . . 67
5.2.1 Introduction to Fuzzy Models . . . 67
5.2.2 Takagi-Sugeno (TS) Fuzzy Models . . . 70
5.2.3 Identification and Validation . . . 75

6 Model Predictive Control 83
6.1 Basic Concept . . . 83
6.2 State-Space MPC . . . 85
6.2.1 Augmented Model Design . . . 86
6.2.2 Prediction of Process Output . . . 88
6.2.3 Constraints . . . 90
6.2.4 Optimization . . . 91
6.3 Fuzzy Models in MPC . . . 92
6.3.1 Introduction . . . 92
6.3.2 Linear Model Extraction . . . 92
6.3.3 Internal Model Control . . . 94
6.3.4 Stability . . . 95

7 Model Predictive Control Implementations 97
7.1 Predictive Control of Supply Temperature . . . 97
7.1.1 Problem Formulation . . . 97
7.1.2 Set-Point Strategies . . . 98
7.1.3 Controller Setup . . . 101
7.1.4 Simulation Results . . . 103
7.2 Predictive Control of Moving Grate Biomass Furnace . . . 109
7.2.1 Problem Formulation . . . 109
7.2.2 Process Description . . . 110
7.2.3 Controller Specification . . . 111
7.2.4 Simulation Results . . . 112

8 Conclusion 121
8.1 Summary of the Main Results . . . 121
8.2 Outlook . . . 122

Glossary

Physical Symbols

Notation	Description	Unit
A	pipe cross section	m^2
a	overall heat transfer coefficient	W/m^2K
c	courant number	–
C_{pos}	combustion position	%
c_p	specific heat capacity of water	J/kgK
$c_{p,b}$	specific heat capacity of biomass boiler	J/kgK
$c_{p,f}$	specific heat capacity of flue gas	J/kgK
D	pipe diameter	m
Δp	pressure differential	Pa
f	Darcy friction factor	–
g	acceleration of gravity	m/s^2
H	head difference between the pipe ends	m
q	piezometric head	m
k_c	proportional factor rate of combustion	s/mkg
k_{in}	proportional factor between biomass input feed and grate velocity	m/kg
$k_{L_{min}}$	factor for the required air demand for complete combustion	–
k_{pos}	proportional factor combustion position	s/mkg
k_{th}	heat value of biomass input	kJ/kg
L	pipe length	m
λ	stoichiometric air to fuel ratio	–
τ_{O_2}	time constant of oxygen concentration	s
τ_{pos}	time constant of combustion position	s
$\dot{m}$	mass flow rate	kg/s
$\dot{m}_{in}$	biomass input feed (fuel feed)	kg/s
$\dot{m}_{PA}$	mass flow rate of the primary air supply	kg/s
$\dot{m}_{rec}$	mass flow rate of recycle gas	kg/s
$\dot{m}_{SA}$	mass flow rate of the secondary air supply	kg/s
m_b	mass of biomass boiler	kg

Notation	Description	Unit
$\dot{m}_\mathrm{c}$	rate of combustion	$\mathrm{kg/s}$
$\dot{m}_\mathrm{f}$	mass flow of flue gas	$\mathrm{kg/s}$
m_g	fuel mass on the grate	kg
$\mathrm{O_2}$	oxygen concentration of flue gas	vol %
Q	heat	W
q	volume flow rate	$\mathrm{m^3/s}$
Q_th	thermal power	kW
ρ	density of water	$\mathrm{kg/m^3}$
T_a	ambient temperature	°C
T_f	flue gas temperature	°C
T_FA	fresh air temperature	°C
T_g	ground temperature	°C
T_r	return temperature	°C
T_s	supply temperature	°C
v	flow velocity	$\mathrm{m/s}$
v_g	velocity of moving grate	$\mathrm{m/s}$
$\bar{v}_\mathrm{g}$	offset grate velocity	$\mathrm{m/s}$

Acronyms

Notation	Description
AC	aggregated consumer
ACF	autocorrelation function
AR(I)MA	autoregressive (integrated) moving average
ARX	autoregressive with exogenous input
cARX	conditional parametric ARX
cFIR	conditional parametric FIR
CP	critical point
CV	control volume
DB	downstream boundary
DHN	district heating network
FIR	finite impulse response
LLM	local linear model network
LOLIMOT	local linear model tree
LPV	linear parameter-varying
MAE	mean absolute error
MAPE	mean absolute percentage error
MARS	multivariate adaptive regression splines
MIMO	multiple-input multiple-output
MISO	multiple-input single-output
MPC	model predictive control
MSE	mean squared error
NMSS	non-minimal state-space representation
PACF	partial autocorrelation function
QUICKEST	quadratic upstream interpolation for convective kinematics with estimated streaming terms
RMSE	root mean squared error
SARIMAX	seasonal autoregressive integrated moving average with exogenous input
SISO	single-input single-output
UB	upstream boundary
WLS	weighted least squares
WRLS	weighted recursive least squares

Chapter 1

Introduction

1.1 Motivation

The worldwide concern about global warming due to the emission of CO_2 and other greenhouse gases and the limited availability of fossil fuels have increased the importance of renewables such as biomass in energy production [218, 217]. It can be expected that power producers will have to cope with an increasing number of EU-level regulations concerning emission levels in the future [204]. Such regulatory actions and the general willingness to reduce emissions on the one hand and increasing world energy consumption [41] on the other hand favour the use of biomass.

In particular, biomass-fired district heating has become more and more important over the past decades [15]. The purpose of district heating networks is to provide adequate thermal power (temperature and flow) to consumers from central or distributed heat sources through a network of pipes containing hot water or steam. Such centralized biomass-fired district heating plants provide more efficient combustion and better pollution control than localized installations at the consumers [79]. Moreover, it was analyzed that district heating plants equipped with a combined heat and power unit (CHP) are one of of most effective and efficient forms of power generation [175].

In Austria the number of district heating plants increased from 266 in 1995 [161] to 1185 in 2005 [124]. However, it can be expected, that this rapid growth will not be seen in the near future. Plant operators will have to face economic challenges such as increased biomass prices which will also have a negative impact on the plants' profitability. Biomass prices will likely rise further due to the increasing demand for biomass resources on the market [54]. By 2020 forecasts predict a biomass supply gap of around 24-40 Mt within the EU [44]. Therefore, potential plant operators have to carefully observe current biomass prices and their change and have to consider if, under these market conditions, a biomass plant can be operated profitably.

One main problem is that small and medium-sized decentralized biomass power plants are characterised by a low degree of automation, high losses in the furnace and an inefficient operation of the district heating networks. These plants are mostly equipped with very limited control concepts, which are usually delivered by the plant suppliers themselves. This reduces initial investment costs and commissioning costs at the expense of increased operational costs. Plant operators are often lacking the know-how for an integrated economic and ecological controller optimization for the power plant and the distribution network. This motivates the use of more advanced control schemes for optimization of biomass power plant in connection with district heating networks. The aim is to reduce operational costs by increasing the plant efficiency and thus to stay profitable in a more and more competitive market.

The optimal operation of such complex systems calls for consideration of the entire technological string "production, transport & distribution, and consumption" [10]. The optimization problem on the production side is twofold: First, it consists of determining how the future heat

load is allocated between the different heating plants in order to minimize the operational costs [158]. Second, the heat production in each individual plant has to be done in such way to ensure efficient combustion and to minimize emissions. The operational costs for the heat transport are mainly related to pumping costs and heat loss costs. Therefore, the reduction of heat energy losses and pump energy consumption is one of the most important tasks to reduce costs and improve the efficiency of a district heating systems. On top of this, the varying heat consumption of consumers introduces a time dependency to all these optimization problems. Hence, optimal operation of such highly complex systems requires accurate modeling tools as well as advanced control concepts to fully address the efficiency potential.

1.2 Literature Review, State-of-the-Art

The following literature overview is first focused on modeling and control of the distribution network, then methods for prediction of heat-load are presented, and finally modeling and control concepts for the biomass plant are reviewed.

For modeling hydraulic and thermal behaviour of district heating networks the quasi-dynamic approach is typically applied [65, 193]. Quasi-dynamic means that a steady-state water flow model and a dynamic thermal model is used. One of the first methods for calculating the flow conditions in pipe networks iteratively is the Hardy Cross method. Nowadays, more advanced approaches which utilize iterative solvers such as Newton-Raphson combined with Graph-Theory are chosen [183, 192, 157]. The steady-state solution of the flow condition is then used as a basis for prediction of the thermal transients.

The transient thermal behavior in district heating networks can be described by the convective transport partial differential equation [19, 132]. Due to the complexity of DHN numerical schemes are often applied here [77]. Two numerical solution approaches are presented in [165], namely the *element method* and the *node method*. The former is based on a first order upwind discretization scheme, whereas the node method keeps trace of how long time a water mass element which currently arrives at the outflow node has been on its way from the inflow node by addressing the mass flow time histories between inflow and outflow nodes. The element method was found to be inferior to the node method, both with respect to accuracy and computational cost. The main reason for the poor performance of the element method was found to be problems regarding artificial diffusion, which in turn could result in abnormal smoothing of sharp temperature profiles. In [65], the node method approach and the commercial software TERMIS were compared regarding the prediction of the temperature dynamics in the Naestved DHN. Results indicated that discrepancies between the predicted and measured temperatures are pronounced for consumers at distant pipelines containing numerous bends and fittings. Other implementations of the node method for prediction of the temperature transients have been tested in, e.g., [64, 191]. Recently, a new thermal transient approximation was presented in [193], which is based on a numerical scheme of third-order accuracy in space using Lagrange polynomials for interpolation. The comparison of the calculated and measured temperature fronts at three consumer substations showed good agreement. A finite-difference method, which is regarded to be one of the best methods for modeling convective transport, is the third order accurate QUICKEST (Quadratic Upstream Interpolation for Convective Kinematics with Estimated Streaming Terms) method derived by [129]. In [165] this scheme was briefly introduced and the possible application to modeling thermal transients in district heatings systems was discussed. It was found that slight modifications to the original algorithm are necessary in order get correct solutions at pipe junctions.

A novel data-driven approach for predicting the supply temperature at critical nodes in the networks was introduced in [169, 158]. As modeling framework conditional parametric models were used. These are classical linear regression models in which the coefficients are allowed to

vary as smooth functions of some explanatory variables. It was shown that within this framework it is possible to account for varying transport time delays in the distribution network.

The relevant control variable in district heating systems is considered to be the supply temperature [76]. Today, in most DHN, open-loop control strategies are implemented which conservatively guarantee that the supply temperatures at the consumer stations are maintained within the contractual limits. In an early work on optimizing the supply temperature trajectories an iterative solution was proposed [15]. First, the transport time delays were determined and the temperature propagation was simulated using the node-method, and second a linear programming (LP) problem was solved. A good overview of simple operational models and several operating strategies is given in [21]. Without quantitative analysis, it was already pointed out that for optimal control the trade-off between pumping and heat loss costs has to be taken into account. In [5] the optimal operations management of a DHN with a combined heat and power plant of an Italian company was solved by means of Mixed Integer Linear Programming (MILP). A custom simulation tool for DHN was developed in [211]. The software was written in Java and was used to evaluate real-time control strategies. A concept for controlling the supply temperature in district heating systems using stochastic modeling and prediction was presented in [159]. An Extended Generalized Predictive Controller (EGPC) for predicting the temperature at critical nodes in the network was derived there. Prediction uncertainties were also considered in the control scheme. Robust predictive control of a DHN with two production sites, valves, pumps and several consumers was addressed in [181]. Due to computational feasibility issues, transport delays were held constant. Treating DHN as multi-agent based system was done in [104]. Distributed policies for operational planning were implemented and evaluated with respect to demand side management and resource management.

In general, the approach to model the heat demand may be classified into two categories: physically motivated models based on the thermal causes for heat demand, and data-driven models based on the analysis of historic heat-load data collected by a monitoring system [170]. A purely physical, deterministic model (white box model) for the heat demand based on ordinary differential equations was studied in [95]. A sophisticated grey-box model for the heat consumption was proposed in [155]. The physical knowledge about heat transfer has been used to select an initial model structure. Then data on heat consumption and climate is applied in combination with statistical methods to establish an actual mathematical model of the heat consumption. Simple time series models were presented in [52] and [106]. Here heat-load was split into a temperature dependent part and social behavior of the consumer. In [52] the temperature dependent part was assumed to vary as a piecewise linear function and the social part was represented by a set of dummy variables comprising one week. A sophisticated recurrent neural network model were proposed in [112]. They used (amongst others) the lowest and highest ambient temperature of a day as explanatory variables. Much work was also done on the popular Box-Jenkins methodology [34, 33, 111]. Prediction of heat demand is treated in [34] and [33] by first filtering off exogenous effects such as temperature from the original load series and then identifying a SARIMA (seasonal autoregressive integrated moving average) model for the filtrated series. A SARIMA model was also applied for electricity demand forecasting in [111]. It is obvious that analogies between heat demand and electricity demand can be drawn. For instance, a very comprehensive space-state approach for modeling hourly French national electricity load, taking into account different levels of seasonality, calendar events and weather dependence have been developed in [51, 49]. An extensive survey of forecast models for energy demand in general was presented in [46]. They also discussed different types of seasonality as well as demand data granularity and the level of data aggregation.

Numerous approaches to modeling biomass plants have been proposed in the literature (e.g., [163, 13, 184, 116, 178, 214]. Computational Fluid Dynamics (CFD) simulations for air staging and flue gas recirculation in biomass grate furnaces was performed in [184]. Based on the

simulation results flue gas burnout, furnace temperature distribution as well as combustion efficiency could be improved. A comprehensive CFD investigation of the combustion process in a 50 MW grate fired furnace was also done in [116]. The comparison of the CFD results with available experimental data at the furnace outlet showed good agreement. Very detailed particle size and species modeling was performed in [178, 214]. They investigated the combustion of four biomass materials having different fuel properties and found that (amongst other things) burning rate is mostly influenced by fuel size and smaller fuels result in higher combustion rate. Focusing on real-time applications, a grate-firing biomass furnace was modeled in [163]. This model combines analytical equations for the combustion, a grey-box model for oxygen concentration in the exhaust gases, and a black-box model for steam generation in the boiler. The same furnace model after linearization and reduction was utilized by the author for controlling the combustion load and combustion position in [164]. Based on physical considerations a simple model for a biomass grate furnace was developed in [13]. It only consists of two ordinary differential equations, which makes it particulary suitable for control purposes. The same authors designed a model based controller including an extended Kalman filter and a reference shaping filter in [72]. They claimed that this is the first time ever that a model based control strategy was developed and successfully implemented in a biomass grate furnace.

1.3 Main Contributions

The main contributions of the author were published in the following papers which are expanded and linked together in this thesis.

Paper A

S. Grosswindhager, A. Voigt, and M. Kozek: *Efficient physical modeling of district heating networks.* In modeling and Simulation. ACTA Press, 2011.

Paper B

S. Grosswindhager, A. Voigt, and M. Kozek: *Linear Finite-Difference Schemes for Energy Transport in District Heating Networks.* In Proceedings of the 2nd International Conference on Computer modeling and Simulation (CSSim2011), pp. 35–24, 2011.

Paper C

S. Grosswindhager, A. Voigt, and M. Kozek: *Online Short-Term Forecast of System Heat Load in District Heating Networks.* In Proceedings of the 31nd Annual International Symposium on Forecasting, 2011.

Paper D

S. Grosswindhager, A. Voigt, L. Haffner and M. Kozek: *Predictive control of district heating network using fuzzy DMC.* In Proceedings of International Conference on modeling, Identification & Control (ICMIC), pp. 241–346, 2012

Paper E

S. Grosswindhager, A. Voigt, L. Haffner and M. Kozek: *Fuzzy Predictive Control of District Heating Network.* International Journal of modeling, Identification and Control, accepted

Paper F

S. Grosswindhager, A. Voigt, L. Haffner and M. Kozek: *Fuzzy modeling of Moving Grate Biomass Furnace for Simulation and Control Purposes.* Mathematical and Computer Modelling of Dynamical Systems, in review

In paper A an efficient physical model for simulation of flow and temperature transients in district heating networks is proposed. For the hydraulic part a graph-theoretical approach is utilized. In addition, proper orthogonal decomposition (POD) is applied for obtaining a simplified, reduced hydraulic model. For calculating the thermal propagation a lumped parameter approach is considered where each pipe element consists of ordinary differential equation describing its total heat capacity.

Paper B presents an evaluation of linear explicit finite-difference schemes with different order for solving the energy transport equation in DHN. All methods were put into the versatile state space representation for the purpose of simulating the propagation of two benchmark temperature profiles from the DHN plant to the distant consumers. In particular, it is shown that the third order accurate QUICKEST method can be successfully applied for thermal modeling DHN if algorithm modifications at pipe junctions or when pipe diameters change are performed. It is further proposed to use velocity weighted right wall values to get a robust solution for various flow conditions.

In paper C a seasonal autoregressive integrated moving average (SARIMA) processes is constructed for modeling the system heat load of the DHN Tannheim (Austria). In addition, exogenous influences such as temperature are accounted for by first filtering off these effects from the original load series and then identifying a SARIMA process for the the filtrated series. Furthermore, for the purpose of forecasting the estimated SARIMA model was incorporated into the versatile state space framework and classical Kalman recursion were applied.

Paper D and paper E considers optimal operation of district heating systems by controlling the supply temperature at the plant using a Fuzzy DMC. The local models were found by means of an axis-orthogonal, incremental partitioning scheme. It is shown that the Fuzzy DMC is capable of handling the inherent nonlinearity in the response characteristics of DHN by taking the volume flow rate at the plant as fuzzy variable. Set-point strategies for the critical nodes are carried out and evaluated with respect to operational costs such as pumping and heat loss costs.

In paper F Takagi-Sugeno (TS) fuzzy models are developed for a moving grate biomass furnace for the purpose of simulating and predicting the main process output variables, which are the heat output, the oxygen concentration of flue gas and the temperature of the flue gas. To approximate the process behaviour local linear ARX models were estimated at specific operating regimes using linear and non-linear estimation techniques. The local validity of the linear models was determined by a sophisticated axis-orthogonal, incremental scheme using Gaussian membership functions. The comparison with real measurements obtained from the biomass plant in Grossarl (Salzburg) showed good agreement. In particular, for the heat output minimal simulation error on the validation dataset can be observed.

In addition to the above contributions, structural models in state-space form for modeling the heat-load and the ambient temperature are presented in this thesis. The SARIMA(X) model developed in paper C and the structural model are compared with respect to their prediction performance. It is shown that the latter performs better for longer prediction horizons. Furthermore, to model the consumer stations a static mapping between return temperature and ambient temperature plus social load is developed. The static mapping is estimated using a sophisticated nonparametric regression framework. Moreover, for the control of the supply temperature at

critical nodes in the network a novel supervisory predictive on-line set-point optimizer is introduced, which tries to minimize the operational costs. A simple grey-box model for a biomass grate furnace which is largely based on mass and energy balances is designed in this thesis as well. The validation with real measurements shows good performance. This grey-box model and the fuzzy model developed in paper F are utilized in a fuzzy model predictive control implementation. Various simulation runs are performed and the results clearly demonstrate the benefit of using advanced control schemes such as fuzzy MPC for control of biomass furnaces.

1.4 Structure of Work

This work is structured as follows: Chapter 2 gives a brief introduction to the main components of a district heating network. Chapter 3 addresses the modeling of the distribution network including simulation and validation results using real measurements. A finite-difference scheme is introduced for physical modeling of the thermal transients as well as conditional parametric models for estimating the supply temperature at critical nodes in the network. The following Chapter 4 deals with the modeling of the heat consumption. In particular, statistical methods are applied for predicting the heat-load and modeling the return temperature at consumer stations. The first part of Chapter 5 covers first-principles modeling of biomass combustion plant. The second part gives a general introduction to the fuzzy methodology and demonstrates its effectiveness by designing a fuzzy model for the biomass combustion plant using real measurements. The main concept of state-space model predictive control as well as the extension to fuzzy model predictive control is provided in Chapter 6. Simulation results concerning the predictive control of the supply temperature and of the moving grate biomass furnace are presented in Chapter 7. A brief conclusion is drawn in Chapter 8.

Chapter 2

Main Components of a District Heating Network

District heating systems are basically composed of heat production facilities, consumers stations and a distribution network (see Fig. 2) The former produces heat which is carried via a network of insulated pipes to each individual consumer where heat exchangers are used to transfer heat from the primary side of the distribution pipes to the secondary side of the building. Consumer installations then cause some return temperature and mass flow rates $\dot{m}$ in the network dependent on their space heating and tap-water heating demands. The cooled water is then transported back to the plant where is is heated again.

2.1 Distribution Network

In general the distribution network consists of a finite set of edges representing insulated pipes or valves and a set of nodes as well as pumps and pipe intersections [76, 182]. Note that due to the imperfect insulation of the pipes and friction in the pipes there will always be heat losses as well as pressure losses in the network. Any modeling approach for the distribution network has to take account of this. Another important issue for modeling is the topology of the network. According to the network topology the following groups may be distinguished [76]:

1. pure loop networks
2. tree networks with built-in closed loop structures

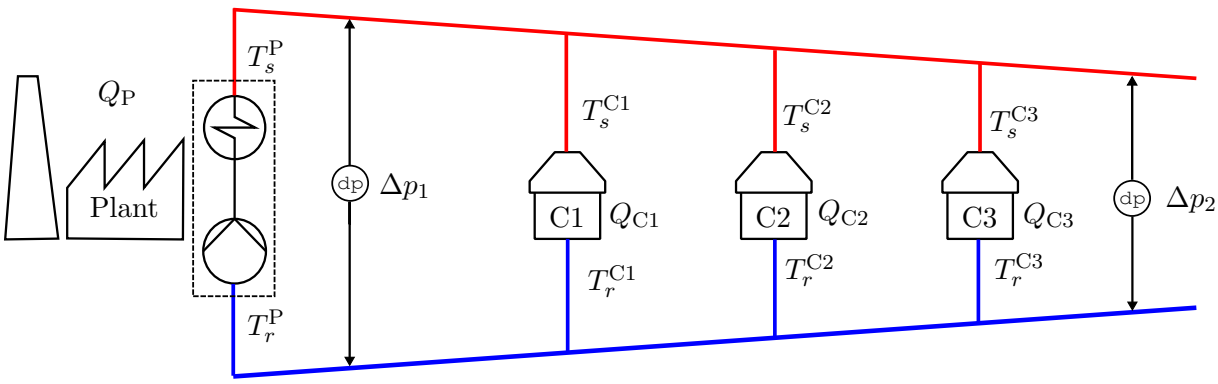


Figure 2.1: Schematic drawing of a district heating network, with supply temperatures T_s, return temperatures T_r, pressure differentials Δp and heat Q.

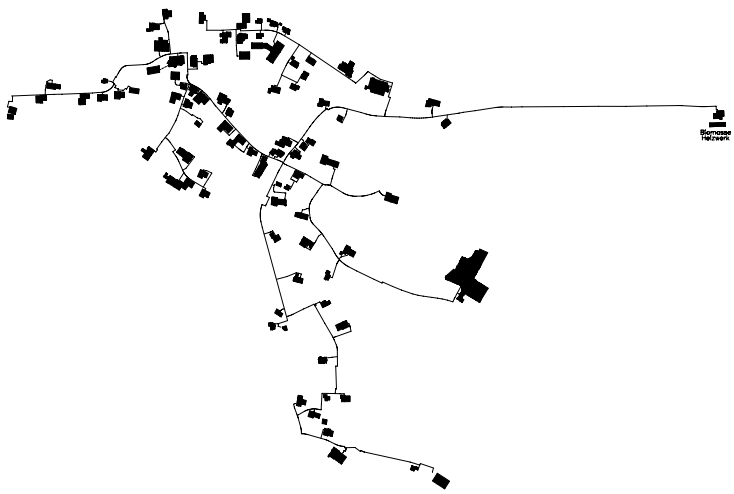


Figure 2.2: District heating network in Tannheim (Austria)

3. tree networks without loops

Pure loop networks consider both the supply line to the consumers and the return line back to the plant. In tree -structured networks the plant represents the source node with consumers acting as sinks. Here additional built-in loops should increase the security of supply. For networks with loops hydraulic calculations are only possible if the flow rates or pressure distributions are predicted simultaneously for the whole network [192]. In particular, iterative procedures such as the Hardy-Cross [43] method or Newton-Raphson method have to be applied [183, 157]. For illustration purposes the distribution network of Tannheim is shown in Fig. 2.1. For such networks which have tree structure the flow rates in the networks are simply calculated by solving a set of linear equations (see Sec. 3.1.1). Although, due to computational reasons, a physically aggregated network is often considered instead of the original network. The equations describing the flow and heat transport are then only constructed for the reduced network.

2.2 Consumer Stations

Consumers connected to the distribution network can either be modeled as heat exchangers or radiator systems [91, 202, 181, 166]. The classical approach is to build the heat energy balance between secondary and primary circuit and solve this for the primary return temperature according to heat exchanger theory [9]. A schematic drawing of a heat exchanger is presented in Fig. 2.2 with the main quantities which affect its design. The heat flow Q transported from the primary side to the secondary side satisfies, if heat losses are neglected, [9, 29]

$$Q = \underbrace{c_p \dot{m}_1 (T_s^1 - T_r^1)}_{\text{Primary Side}} = \underbrace{c_p \dot{m}_2 (T_s^2 - T_r^2)}_{\text{Secondary Side}} \tag{2.1}$$

where c_p is the specific heat capacity of water, $\dot{m}$ is the mass flow rate and T_s and T_r are the supply and return temperature, respectively. Thus, the heat transfer is proportional to the temperature difference between supply line and return line and to the mass flow rate. The secondary supply temperature is often required to be at some predefined level and can therefore assumed to be constant [188]. The primary return temperature, given some supply temperature and heat flow, is then calculated using heat exchanger theory [166]. However, classical heat

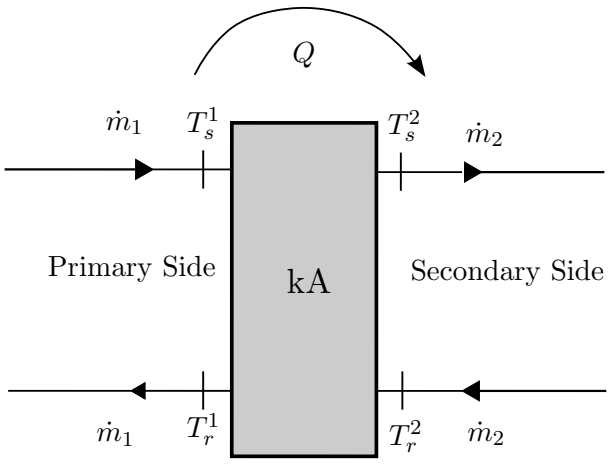


Figure 2.3: Heat exchanger scheme, with the mass flow rate $\dot{m}_i$, supply temperatures T_r^i, return temperatures T_r^i, heat flow Q and characteristic transfer capability kA ($i = 1, 2$)

exchanger theory might not be valid for district heating applications. In [179] it was concluded that the main factors which have an influence on the return temperature are (see also Sec. 4.2):

1. ambient temperature
2. social pattern affecting the heat load.

For instance, for cold days when the space heating dominates the heat-load, higher return temperatures can be anticipated. This is due the fact that the return temperature from the heat exchangers to the radiator circuits is higher than the return temperature from tap water preparation because the water in the radiator circuit is never as cold as the cold water supplying the tap water. On the other hand, when the ambient temperature increases, the tap water preparation becomes a bigger part of the total heat load, and hence decreases the return temperature of the network. For really warm days, however, low consumer heat demand leads to very low flow rates in the network. In order to keep the flow up, supply water might be by-passed to the return line. Other factors mentioned in [179] which influence the return temperature are malfunctions of heat exchangers as well as their efficiency. Heat exchangers are less efficient in case of low supply temperature which result in a higher return temperature.

2.3 Biomass District Heating Plant

Basically, the biomass plants consist of biomass boilers supplying heating and hot water, conventional gas- or oil-fired boilers for peak load demand and supply pumps to guarantee a certain differential pressure in the network. The purpose of the plant is to provide sufficient thermal power to meet consumer demands by maintaining the supply temperature and differential pressure within the contractual limits irrespective of the return temperature and flow rates in the network [147]. The energy conservation law for the district heating network can be formulated as following [202]

$$\underbrace{Q_{\text{bio}} + Q_{\text{foss}} + P_{\text{Pump}}}_{\text{Production}} = \underbrace{c_p \dot{m}(T_s - T_r) + Q_{\text{loss}}}_{\text{Consumption \& Transportation}} \tag{2.2}$$

where Q_{bio} is the heat input by biomass combustion, Q_{foss} heat input by existent conventional furnaces, P_{Pump} is the required pump power to keep up the flow in the network and Q_{loss} are

the heat losses to the surroundings. From an operational point of view it is important that the costs related to heat production are minimized. This also includes, for instance, optimal pump management, when more than one pump is installed, or ensuring efficient combustion, reducing emissions and boiler maintenance time. In the general form this can be considered as a classical unit commitment problem [11, 195]. Note that the dual problem to minimizing production costs (cf. Eq. (2.2)) is to reduce operational costs such as heat loss costs [22]. As concluded in [39] heat loss costs can be reduced by lowering the supply temperature from the plant or increasing the thermal insulation thickness of pipes. Lowering the supply temperature, however, could result in unacceptable supply temperature levels at consumers and also increases the flow rates in the network, resulting in higher pumping costs.

Chapter 3

Modeling of the Distribution Network

3.1 Physical Models

Generally, the governing equations for describing fluid and heat flow in pipes are three coupled partial differential equations (PDE) for the basic conservation laws of *mass*, *momentum*, and *energy* [113]. Conservation of Mass (or the continuity equation) expresses the idea that mass can neither be created nor destroyed, the momentum equation relates to Newton's law of force balances and the energy equation refers to the $1st$ law of thermodynamics. For modelling hydraulic and thermal behaviour of district heating networks the following assumptions are made which allows us to use a set of less complex governing equations:

- Constant density, viscosity, specific heat, and thermal coefficient of the media
- Quasi-dynamic approach

The first assumption formally allows us to restate the problem in an uncoupled thermo-hydraulic form, where the mechanical and thermal properties of the fluid can be considered independently [66]. Compressible media (i.e., nonconstant density) are important if the focus lies on modeling fast transients such as waterhammer effects. On the other hand, if flow conditions in a pipe are only changing slowly, it is reasonable to assume constant density [6]. The term quasi-dynamic relates to a steady-state water flow model and a dynamic thermal model. This means that each time instant the steady-state flow condition in the distribution network is calculated and used as a basis for prediction of the temperature transients [65, 193]. Especially from an operational optimization point of view this approach is appealing, since the thermal and hydraulic dynamics have highly different response times. The flow is driven by pressure waves and reaches steady state typically in a matter of seconds, whereas temperature changes at the plant may need hours to reach consumer stations [76]. Hence, for long term control strategies the dynamics of the flow in the network are of minor importance compared with the dynamics of the temperature changes [193].

In the following sections the conservation of momentum, mass and energy for hydraulic and thermal modelling of district heating pipes are presented.

3.1.1 The Hydraulic Model

The conservation of momentum can be derived by applying Newton's second law of motion

$$\text{Force} = \text{Mass} \times \text{Acceleration} \tag{3.1}$$

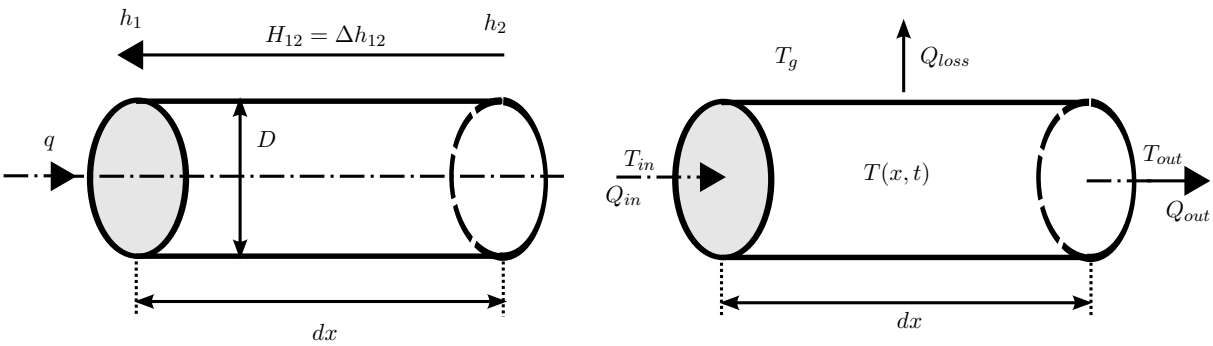


Figure 3.1: Pipe differential elements showing the hydraulic (left) and thermal (right) physical properties; H_{12} is the head difference between the pipe ends and T_g is the ground temperature

to a horizontal differential pipe element (see Fig. 3.1.1 left) which yields [30]

$$\frac{\partial q}{\partial t} + gA\frac{\partial h}{\partial x} + \frac{f}{2DA}q|q| = 0. \tag{3.2}$$

Here h (m) denotes the piezometric head, q (m^3/s) is the volume flow rate , g ($\mathrm{m/s}^2$) is the acceleration of gravity, A (m^2) and D (m) are the cross section and diameter of the pipe element, respectively, and f is the so called *Darcy friction factor*. The latter depends on the Reynold number and the pipe roughness and can be determined by the Law of Hagen-Pousseuille (in case of laminar flows) or the Prandtl-Colebrook formula (in case of turbulent flows) for example [212].

It is also interesting to note the analogy to electrical circuit theory, where the head difference between two points corresponds to voltage, the electrical current is equivalent to volume flow rate or mass flow rate and also all wires have some resistance to current like pipes have some resistance to the flow.

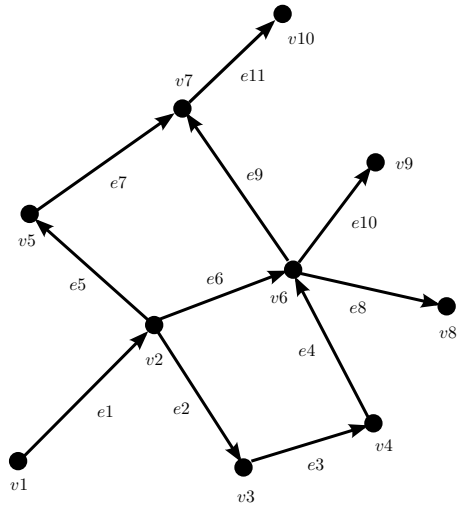
The entire district heating network can basically be discretized into pipe elements as illustrated in Fig. 3.1.1. Hence, the conservation of momentum for the network can be written as a set of ordinary differential equations (ODE) of form (3.2) with appropriate boundary conditions at pipe inflow and outflow nodes. Graph theory [48, 86] allows us to formulate these boundary conditions in an elegant way by utilizing the so called incidence matrix. This matrix will also be used to describe the conservation of mass in the network, which is, due to the assumption of incompressibility, based on the Kirchhoff laws. Before going any further, let us review some graph-theoretic definitions:

Definition 3.1.1 (Diagraph)**.** *Given non-empty sets V and E with $V \cap E = \emptyset$ together with a map $H\colon E \to V \times V$. Then the triple (V, E, H) is called digraph (or directed graph) $D = (V, E, H)$. The elements of V are the vertices (or nodes, or points) of digraph D, the elements of E are its edges.*

Hence, the map H assigns to every edge $e_j \in E$ an initial vertex (tail) $v_i \in V$ and a terminal vertex (head) $v_k \in V$ ($h(e_j) = (v_i, v_k)$). An appropriate representation for this relation H is the so called incidence matrix.

Definition 3.1.2 (Incidence Matrix)**.** *Given a diagraph $D = (V, E, H)$. The matrix*

$$\boldsymbol{G}_{i,j} = \begin{cases} -1, & \textit{if vertex } j \textit{ is initial vertex of edge } i, \\ 1, & \textit{if vertex } j \textit{ is terminal vertex of edge } i, \\ 0, & \textit{otherwise}, \end{cases} \tag{3.3}$$



$D = (V, E, H)$,
$V = \{v_1, v_2, \ldots, v_{10}\}$,
$E = \{e_1, e_2, \ldots, e_{11}\}$,
$H = \{h(e_1) = (v_1, v_2),$
$h(e_2) = (v_2, v_3), \ldots,$
$h(e_{11}) = (v_7, v_{10})\}$.

Figure 3.2: Example diagraph

with $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$ is called incidence matrix of the diagraph D.

Note that each row of the incidence matrix can have maximum two non-zero entries, -1 and 1, since each edge has exactly one initial vertex and one terminal vertex. Therefore, the row sums of the matrix will always be zero. The columns, on the other hand, can have any number of non-zero entries greater than one, since in a directed graph many edges can be connected to a single vertex. For instance, for the case of the directed graph as illustrated in Fig. 3.2 the incidence matrix is populated as follows:

$$\boldsymbol{G} = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \end{bmatrix}. \tag{3.4}$$

The incidence matrix is a comfortable way to describe the topology of any district heating network. If the network structure is once put into this matrix form, the solution to the flow distribution can easily be obtained. Moreover, in [187] a specific composition of the incidence matrix according to certain node boundary conditions was proposed. The following categorization was suggested:

1. Discharge boundary nodes have an outflow (consumption) vector specified with time, $\boldsymbol{Q}_1(t)$, and an unknown head vector $\boldsymbol{h}_1$. These nodes are further called interior nodes.

2. Head boundary nodes have a head vector specified with time, $\boldsymbol{h}_2(t)$, and an unknown outflow vector $\boldsymbol{Q}_2$. These nodes are further called reservoirs nodes.

Note that the outflow vectors are the discharges drawn from the nodes and should not be confused with the discharges through the pipes. Additional boundary conditions such as surge

tanks could be added in a straightforward way (see [187] for details). Corresponding to the type of nodes, a given $(m \times n)$ incidence matrix $\boldsymbol{G}$, a given $(n \times 1)$ head vector $\boldsymbol{h}$ and a given $(n \times 1)$ outflow vector can be arranged as follows:

$$\boldsymbol{G} = [\underset{n-r}{\boldsymbol{G}_1} \; \underset{r}{\boldsymbol{G}_2}], \quad \boldsymbol{h} = \begin{bmatrix} \boldsymbol{h}_1 \\ \boldsymbol{h}_2 \end{bmatrix}_{r}^{n-r}, \quad \boldsymbol{Q} = \begin{bmatrix} \boldsymbol{Q}_1 \\ \boldsymbol{Q}_2 \end{bmatrix}_{r}^{n-r}. \tag{3.5}$$

where r $(r \ll m)$ is the number of reservoir nodes.

In the next step the conservation of momentum for the entire network is formulated using the incidence matrix. First assume that Eq. (3.2) has been set up and spatially discretized for each pipe element such that the overall network can be described by an assembly of interconnected equations. The stationary momentum equation can then be written in vector form as

$$\boldsymbol{H} = \boldsymbol{R}(\boldsymbol{q})\boldsymbol{q} = \boldsymbol{G}_1\boldsymbol{h}_1 + \boldsymbol{G}_2\boldsymbol{h}_2 \tag{3.6}$$

with

$$\boldsymbol{R}(\boldsymbol{q}) = \operatorname{diag}\left\{ \frac{f_1 L_1}{2D_1 A_1^2 g}|q_1|, \frac{f_2 L_2}{2D_2 A_2^2 g}|q_2|, \ldots, \frac{f_m L_m}{2D_m A_m^2 g}|q_m| \right\}. \tag{3.7}$$

and in which $\boldsymbol{q}$ is the discharge vector whose element $q_k,\ k = 1, \ldots, m$ is the discharge through the kth pipe (edge), and H_k, the kth element of $\boldsymbol{H}$ is the head drop in the kth pipe. Note that other hydraulic elements such as pumps or valves can be added in a straightforward way (see[6] for details).

However, problem (3.6) is, from a mathematical point of view, under-determined since there are $m + n - r$ unknowns (discharges trough the pipes plus head at the interior nodes) and only m nonlinear equations. In order to get a well-posed formulation of the system the conservation of mass at the interior nodes has to be taken into account. This can be written as follows

$$\boldsymbol{G}_1^T \boldsymbol{q} = \boldsymbol{Q}_1. \tag{3.8}$$

By introducing a partitioned vector of the unknown variables $\boldsymbol{x} = \begin{bmatrix} \boldsymbol{q}^T & \boldsymbol{h}_1^T \end{bmatrix}^T$, i.e., discharges through the pipes and head at the interior nodes, the solution to the system of equations of (3.8) and (3.6) can be reduced to find the zeros of the following function [157]

$$f(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{R}(\boldsymbol{x}) & \boldsymbol{G}_1 \\ \boldsymbol{G}_1^T & \boldsymbol{0} \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} \boldsymbol{G}_2\boldsymbol{h}_2 \\ -\boldsymbol{Q}_1 \end{bmatrix} = 0. \tag{3.9}$$

A reduction of the $m+n-r$ primary unknowns can further be achieved by either expressing $\boldsymbol{R}$ in terms of $\boldsymbol{q}$ or in terms of $\boldsymbol{H}$ resp. $\boldsymbol{h}$ (i.e., rearranging first part of Eq. (3.6)) which yields the well-known loop and node equations, respectively [157]. For instance, the loop equations can be derived by writing the complete solution to the continuity equation (3.8) in the form

$$\boldsymbol{q} = \boldsymbol{q}_c + \boldsymbol{C}\boldsymbol{u} \tag{3.10}$$

where $\boldsymbol{q}_c$ is any $(m \times 1)$ flow vector that satisfies (3.8), $\boldsymbol{u}$ is any $(m - n + r \times 1)$ vector, and the $m \times (m - n + r)$ matrix $\boldsymbol{C}$ satisfies $\boldsymbol{G}_1^T\boldsymbol{C} = 0$ and $\boldsymbol{C} \neq 0$, i.e., $\boldsymbol{C}$ is the kernel or null space of $\boldsymbol{G}_1^T$. The matrix $\boldsymbol{C}$ is also denoted as loop matrix and $\boldsymbol{u}$ as the loop flow vector. By multiplying Eq. (3.6) by $\boldsymbol{C}^T$ and substituting (3.10) for $\boldsymbol{q}$ one obtains

$$\boldsymbol{C}^T\boldsymbol{R}(\boldsymbol{q})\boldsymbol{C}\boldsymbol{u} + \boldsymbol{C}^T\boldsymbol{R}(\boldsymbol{q})\boldsymbol{q}_c + \boldsymbol{C}^T\boldsymbol{A}_2\boldsymbol{h}_2 = 0. \tag{3.11}$$

This simply expresses the second Kirchhoff Law, that the accumulated head loss around each closed loop is zero.

The Hardy Cross method [43] was the first method for solving the nonlinear loop equations (3.11) iteratively. This method determines the loop flow rate corrections $\boldsymbol{u}$ for each loop independently and than calculates the corrected flow for each pipe. More efficient methods, such as the Newton-Raphson algorithm, calculate the flow corrections simultaneously for each loop in the whole network [183, 192]. Here the iterative update scheme can be written in the following way [157]

$$\begin{aligned} \left[\boldsymbol{C}^T \boldsymbol{R}(\boldsymbol{q}_k)\boldsymbol{C}\right] \Delta \boldsymbol{u}_{k+1} &= -\boldsymbol{C}^T \left[\boldsymbol{R}(\boldsymbol{q}_k)\boldsymbol{q}_k + \boldsymbol{A}_2\boldsymbol{h_2}\right] \\ \boldsymbol{q}_{k+1} &= \boldsymbol{q}_k + \frac{1}{2}\boldsymbol{C}\Delta \boldsymbol{u}_{k+1}. \end{aligned} \tag{3.12}$$

In a similar way the node equations can be derived. First, $\boldsymbol{R}(\boldsymbol{x})$ is expressed in terms of the head $\boldsymbol{h}_1$, i.e., $\boldsymbol{R}(\boldsymbol{h}_1) = \boldsymbol{R}(\boldsymbol{x})$. Then the upper part of Eq. (3.9) can be solved for $\boldsymbol{q}$ under the assumption that $\boldsymbol{R}(\boldsymbol{h}_1)$ is non-singular and be substituted into the lower part of (3.9) which yields

$$\boldsymbol{G}_1^T \boldsymbol{R}(\boldsymbol{h}_1)^{-1}\boldsymbol{G}_1\boldsymbol{h}_1 + \boldsymbol{G}_1^T \boldsymbol{R}(\boldsymbol{h}_1)^{-1}\boldsymbol{G}_2\boldsymbol{h}_2 = -\boldsymbol{Q}_1. \tag{3.13}$$

This is a nonlinear system of $n - r$ equations in the unknown head vector $\boldsymbol{h}_1$ which can again be solved by applying the Newton-Raphson algorithm. The iterative update is given by [157]

$$\boldsymbol{h}_{1,k+1} = \boldsymbol{h}_{1,k} - 2\left[\boldsymbol{G}_1^T \boldsymbol{R}_k^{-1}\boldsymbol{G}_1\right]^{-1}\left[\boldsymbol{Q}_1 + \boldsymbol{G}_1^T \boldsymbol{R}_r^{-1}(\boldsymbol{G}_1\boldsymbol{h}_{1,k} + \boldsymbol{G}_2\boldsymbol{h}_2)\right] \tag{3.14}$$

where $\boldsymbol{R}_k = \boldsymbol{R}(\boldsymbol{h}_{1,k})$.

Example 3.1.1:

For exemplary purposes let us consider the diagraph shown in Fig. 3.2 with corresponding incidence matrix (3.4). Node v_1 should act as reservoir node (r=1) and the remaining nodes are treated as interior nodes. The outflow vector of the interior nodes is given by

$$\boldsymbol{Q}_1^T = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1.5 & 2 \end{bmatrix}$$

and the resistive values for the pipes are randomly chosen as

$$\boldsymbol{R} = \text{diag}\begin{bmatrix} 0.2 & 0.1 & 0.6 & 0.1 & 0.2 & 0.7 & 0.3 & 0.5 & 0.4 & 0.3 & 0.2. \end{bmatrix}$$

The loop matrix for this network (i.e., null space of $\boldsymbol{G}_1^T$) is given by

$$\boldsymbol{C}^T = \begin{bmatrix} 0 & 0 & 0 & 0 & -1 & 1 & -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & -1 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The non-zero entries in the loop matrix are exactly the nodes which are arranged in the two closed loops as shown in Fig. 3.2. The volume flow rate of the first 5 steps of the Newton-Raphson algorithm is presented in the following table:

k	**Volume Flow Rate**										
0	4.5	0	0	0	2	2.5	2	1	0	1.5	2
1	4.5	2.25	2.25	2.25	1	1.25	1	1	1	1.5	2
2	4.5	1.482	1.482	1.482	1.659	1.360	1.659	1	0.342	1.5	2
3	4.5	1.348	1.348	1.348	1.716	1.436	1.716	1	0.284	1.5	2
4	4.5	1.344	1.344	1.344	1.719	1.437	1.719	1	0.281	1.5	2
5	4.5	1.344	1.344	1.344	1.719	1.437	1.719	1	0.281	1.5	2

Note that an initial solution q_0 satisfying the conservation of mass condition (3.8) was obtained by LU-factorization. As one can see, after four iterations the Newton-Raphson algorithm converges to the steady-state flow distribution solution in the network.

Remark. *In district heating systems the supply network often has a tree structure (i.e., no closed loops) with one reservoir node, the biomass plant. In that case, the matrix* $\boldsymbol{A}_1$ *has full rank and the continuity equation* (3.8) *is sufficient to calculate the flow situation in the entire network.*

3.1.2 The Thermal Model

Energy Transport Equation

The modeling of energy transport in district heating pipes belongs to the class of convective (or advective) transport models. A general formulation of the one-dimensional energy transport equation for a scalar $\phi(x,t)$ with source term is given by the following partial differential equation (PDE) [19, 132]

$$\begin{aligned} &\frac{\partial \phi}{\partial t} + v\frac{\partial \phi}{\partial x} + a\phi = 0 \\ &\phi(x,0) = \phi_0(x),\ \ \phi(0,t) = \phi_i(t) \end{aligned} \tag{3.15}$$

where a is a well defined source term (heat loss in case of DHN pipes), $v = v(x)$ is the velocity, and $\phi_0(x), \phi_i(t)$ represent some initial and boundary conditions, respectively. It should be noted that the velocity v does not necessarily represent the physical velocity of the transport medium as discussed in [165]. In particular, for district heating networks the thermal energy propagation can be written as [65, 181]

$$\begin{aligned} &\frac{\partial T}{\partial t} + \frac{4\dot{m}(t)}{\rho D^2 \pi}\frac{\partial T}{\partial x} + \frac{4a}{c_p \rho D}(T - T_g) = 0 \\ &T(x,0) = T_0(x), \quad T(0,t) = T_i(t) \end{aligned} \tag{3.16}$$

where $T(x,t)$ (K) describes the temperature distribution along the pipe, $\dot{m}$ (kg/s) ($\dot{m} = \rho q$) denotes the mass flow through the specific pipe segment, D (m) is the pipe diameter, ρ (kg/m^3) is the relative density of water, c_p (J/kgK) is the specific heat of water, and $T_0(x)$ and $T_i(t)$ are appropriate initial and inflow conditions, respectively. Heat losses to the surroundings are proportional to the temperature difference between fluid temperature and ground temperature T_g (K), and are affected by an overall heat transfer coefficient a (W/m^{2}K). The latter is composed of the convective heat transfer coefficient between the flow and the pipe and the thermal conductivities of the pipe, the insulation and the soil, respectively [191].

An analytic solution to the transport problem (3.16) is easily obtained and has the form [181]:

$$T_{out}(t) = T_g + (T_{in}(t - \tau(t)) - T_g)\exp^{-\frac{4a}{c_p \rho D}\tau(t)} \tag{3.17}$$

where T_{out} and T_{in} denote temperature of the fluid at the pipe outlet and inlet, respectively and where the varying transport time delay $\tau(t)$ is defined by

$$\int_{t-\tau(t)}^{t} \frac{4\dot{m}(\zeta)}{\rho D^2 \pi} d\zeta = L. \tag{3.18}$$

As can be seen in (3.17) the temperature difference between pipe inlet and outlet depends on the transport time delays. This means that the longer it takes the fluid particles to pass through the pipe the more heat is emitted to the surroundings. In practical implementations the calculation of the varying transport time delays is time consuming and a limiting factor for real time simulation and optimization [181]. Numerical solution methods can therefore be used which try to numerically approximate these transport time delays.

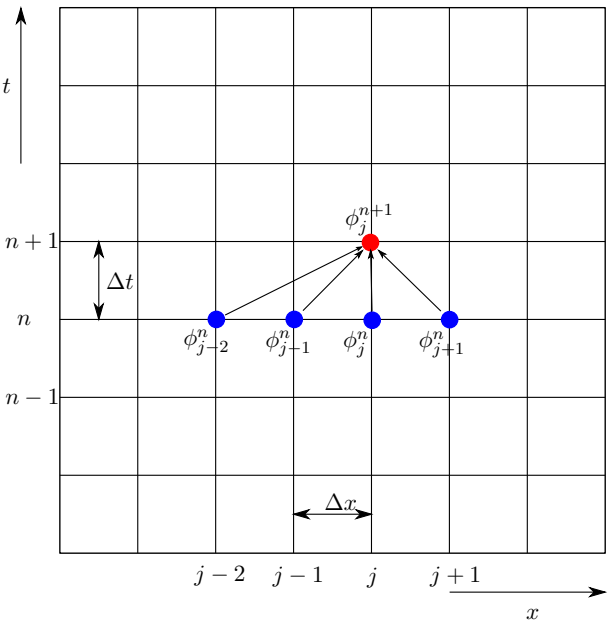


Figure 3.3: Schematic representation of the uniform grid structure in space and time (Quickest stencil is shown)

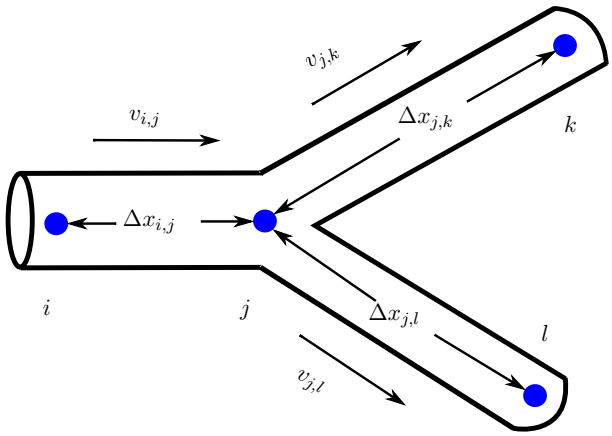


Figure 3.4: Representation of a pipe junction in district heating networks assuming non-uniform grid

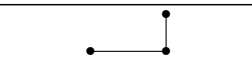
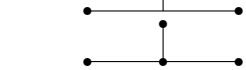
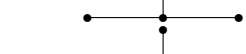
Numerical Solution Methods

In general, two categories of numerical solution methods can be applied to solve partial differential equations of type (3.15): *Finite-Difference Methods* [133, 132, 206, 122] and *Galerkin Methods* [56, 99]. The implementation of finite-difference methods for the adjective transport problem is demonstrated in [154, 215, 67]. The success of the finite-difference approach is mainly due to its conceptual simplicity. This work also focuses on finite-differece schemes. One of these schemes is dicussed in greater detail below.

Quickest. In this work the third order accurate QUICKEST (Quadratic Upstream Interpolation for Convective Kinematics with Estimated Streaming Terms) method derived by [129] was applied due to its excellent performance in terms of convective transport problems [165]. In its original form Quickest used an explicit, Leith-type differencing and third-order upwinding on the convective derivatives to yield a four-point upwinded scheme (cp. Fig. 3.3 and Tab. 3.1) [190]. Numerical tests for pure advection in one-dimensional flow problems have shown that this scheme gives in general better results (reduces the numerical diffusion) than second-order schemes, e.g., Lax - Wendroff or Beam-Warming [67, 77]. However, as demonstrated in [130] both second order schemes, third order schemes such as Quickest, and even higher order methods are prone to small oscillations and/or overshooting, especially near sharp gradients. This limitation is predicted by Godunov's theorem [71], which states that any linear convection scheme with the property of not generating new extrema (monotone scheme) can be at most first-order accurate. To overcome this nonlinear schemes were proposed, which adjust themselves according to the local solution. For an overview see [210].

Generally, finite-difference schemes are constructed by defining a discrete computational grid in the $x - t$ plane with time increments Δt and spatial increments $\Delta x_{l,m}$ between grid points l and m. In the following the notation ϕ_j^n is used where the index j and n denote the discrete mesh points of space and time, respectively (see Fig. 3.3 and Fig. 3.4). A conservative control-volume

Table 3.1: Overview of Finite-Difference Schemes

Name	**Time Update**	**Stencil Representation**
First Order Upwind	$\phi_j^{n+1} = f(\phi_j^n, \phi_{j-1}^n)$	
Lax - Friedrich	$\phi_j^{n+1} = f(\phi_{j+1}^n, \phi_{j-1}^n)$	
Lax - Wendroff	$\phi_j^{n+1} = f(\phi_{j+1}^n, \phi_j^n, \phi_{j-1}^n)$	
Beam Warming	$\phi_j^{n+1} = f(\phi_j^n, \phi_{j-1}^n, \phi_{j-2}^n)$	
Quickest	$\phi_j^{n+1} = f(\phi_{j+1}^n, \phi_j^n, \phi_{j-1}^n, \phi_{j-2}^n)$	

form of (3.15) can be written as (cp. Fig. 3.5 and Fig. 3.6) [130, 200]

$$\phi_j^{n+1} = \phi_j^n + (c_{j-1}\phi_{j-1}^* - c_j\phi_j^*) + \Delta t\bar{S} \tag{3.19}$$

where ϕ_j^* and ϕ_{j-1}^* are the right- and left-face values of the transported scalar for the ith control cell, and c_r and c_l are the right- and left-face values of the Courant number, and $\Delta t\bar{S}$ is the time averaged source term. The Courant number plays also a crucial role in stability analysis and is defined as [42]

$$c = v\frac{\Delta t}{\Delta x}. \tag{3.20}$$

It indicates how fast temperature information travels on the computational grid. Specifically in the Quickest scheme, as the name suggests, the right- and left-face values are calculated using quadratic upstream interpolation. In mathematical terms this can be written as [129]

$$\phi_j^* = \frac{\phi_j^n + \phi_{j+1}^n}{2} - \frac{\Delta x_j}{2}c_j\mathrm{GRAD}_j - \frac{\Delta x_j^2}{6}(1 - c_j^2)\mathrm{CURV}_j \tag{3.21}$$

with

$$\mathrm{GRAD}_j = \frac{\phi_{j+1}^n - \phi_j^n}{\Delta x_j}, \qquad \mathrm{CURV}_j = \frac{\mathrm{GRAD}_j - \mathrm{GRAD}_{j-1}}{(\Delta x_j + \Delta x_{j-1})/2}$$

and $\Delta x_j = \Delta x_{(j,j+1)}$, i.e., the distance between nodes $(j+1, n)$ and (j, n). Note that the original Quickest algorithm as formulated in [129] also allows for a non-uniform computational grid, which is very convenient in case of district heating networks. An overview of different finite-difference schemes including their stencil representation is presented in Tab. 3.1. By means of the stencil one can easily determine the spatial order of the method. See [133, 77] for details.

Modified Quickest. Equation (3.19) represents a conservative form, which means that changes in the flow velocity along the control volume will affect the temperature profile. However, for district heating applications, this produces incorrect results at pipe junctions or when diameters are changed [165]. To overcome this the problem can be formulated non-conservatively:

$$\phi_j^{n+1} = \phi_j^n + \frac{2\Delta t}{\Delta x_j + \Delta x_{j-1}}v_{j-1}(\phi_{j-1}^* - \phi_j^*) + \Delta t\bar{S}. \tag{3.22}$$

Also in district heating networks the time averaged source term $\bar{S}$ can be approximated by $\frac{a}{2}(\phi_j^{n+1} + \phi_j^n)$ [165]. Moreover, in [77] it was recommended to use velocity weighted right wall values in the case of pipe junctions (see Fig. 3.4) to get a more robust scheme. More precisely, to calculate ϕ_4^{n+1} in Fig. 3.5 one could either use $\phi_{4,1}^*$ or $\phi_{4,2}^*$ as right-face value. A natural choice would be to use the node which is connected to the "main" distribution pipe. For instance,

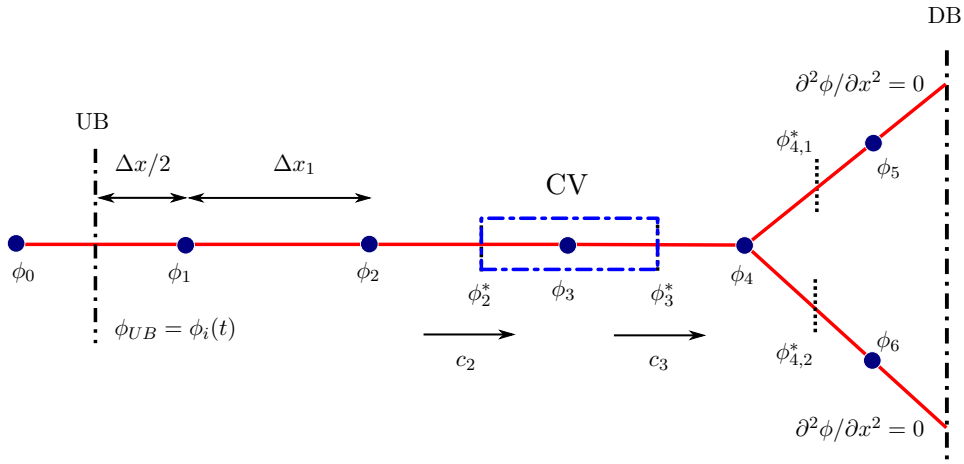


Figure 3.5: Schematic representation of the finite-difference grid for the supply line with imposed boundary conditions: Dirichlet boundary condition plus mirror node upstream (UB); Zero curvature condition downstream (DB)

some mass flow weighting of the potential right wall values would be appropriate. In [77] weights corresponding to the velocity of the fluid in the pipe have been suggested, which worked reasonably well in practice. Here ϕ^*_j and Δx_j in Eq. (3.22) are replaced by

$$\tilde{\phi}^*_j = \frac{\sum_{l=1}^N v_{j,l}\phi^*_{j,l}}{\sum_{l=1}^N v_{j,l}}, \qquad \Delta\tilde{x}_j = \frac{\sum_{l=1}^N v_{j,l}\Delta x_{j,l}}{\sum_{l=1}^N v_{j,l}}. \tag{3.23}$$

where N is the number of pipes connected to the junction, $v_{j,l}$ is the velocity of the fluid and $x_{j,l}$ is the spatial discretization in the lth pipe connected to the junction, respectively.

Similar arguments can be applied for the return line as illustrated in Fig. 3.6. Again special treatment of pipes in junctions is necessary. For instance, the correct return mixing temperature has to be calculated using steady-state energy balances [193]. It is suggested to not compute the mixing temperature directly at the junction (node ϕ_3 in 3.6) but rather to interpolate the left-face values ϕ^*_1 and ϕ^*_2. Hence, ϕ^*_{j-1} and Δx_{j-1} in Eq. (3.22) may be replaced by the following expressions

$$\tilde{\phi}^*_{j-1} = \frac{\sum_{l\in\mathcal{I}} \dot{m}_{l,j}\phi^*_l}{\sum_{l\in\mathcal{I}} \dot{m}_{l,j}}, \qquad \Delta\tilde{x}_{j-1} = \frac{\sum_{l\in\mathcal{I}} \dot{m}_{l,j}\Delta x_{l,j}}{\sum_{l\in\mathcal{I}} \dot{m}_{l,j}}. \tag{3.24}$$

where $\mathcal{I}$ is the index set of nodes affected by the junction and $\dot{m}_{l,j}$ is the mass flow rate through the lth pipe. The Quickest scheme, due to its third order accuracy, also utilizes the far left node ϕ_{j-2} for update of ϕ_j. Hence, as shown in Fig. 3.6 also ϕ_4 is affected by the pipe junction. A simple solution to this problem would be to use a linearly extrapolated hypothetical node as far left node. This means that the curvature term is zero in formula (3.19) for approximating the left-face value ϕ^*_3.

Remark. *Note that the proposed modifications to the classical Quickest scheme are merely ad hoc solutions to handle pipe junctions in district heating systems in practice. It is not claimed that this approach is optimal from a fluid dynamics point of view.*

Boundary Conditions. For the energy transport problem considered it is essential to impose appropriate boundary conditions. In particular, in the case of the supply line where hot fluid is transported from the plant to the consumers, the former acts as an upstream boundary condition

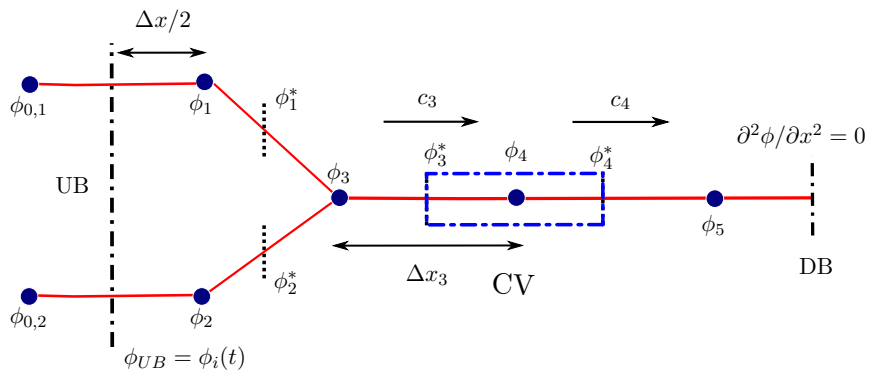


Figure 3.6: Schematic representation of the finite-difference grid for the return line with imposed boundary conditions: Dirichlet boundary condition plus mirror node upstream (UB); Zero curvature condition downstream (DB)

and the latter as downstream boundary conditions. For the case of cold fluid flowing back to the plant this situation is exactly reversed.

Leonard [129] suggested to treat upstream boundaries by introducing an imaginary node value (i.e., a value at a hypothetical node $\Delta x/2$ upstream of the boundary condition; cp. Fig. 3.5 and Fig. 3.6), and to specify the boundary condition to be a wall value rather than a node. The value of the image node is then $\phi_0^{n+1} = -\phi_1^n$ at each time instant of the computation. Furthermore, for the outflow (downstream) numerical boundary conditions zero-curvature was suggested.

Taking into account these boundary conditions and the modifications proposed for pipe junctions, the time-update of an entire network with m interior nodes and r reservoir nodes (plant in the case of supply line and consumers in the case of return line) can be written in the matricial form as follows

$$\boldsymbol{\phi}^{n+1} = \boldsymbol{A}\boldsymbol{\phi}^n + \boldsymbol{B}\boldsymbol{\phi}_i^n, \tag{3.25}$$

with $\boldsymbol{\phi}^n = [\phi_1^n, \ldots, \phi_m^n]$, $(m \times m)$ interior node matrix $\boldsymbol{A}$ and $(m \times r)$ reservoir node matrix $\boldsymbol{B}$. The initial condition is represented by $\boldsymbol{\phi}^0$ and the inflow boundary condition by $\boldsymbol{\phi}_i^n$. The matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ contain the coefficients of the difference formulas (3.22). Note that with varying flow in the network these matrices have to be updated at each time step.

Stability Analysis. A common method to prove stability of finite-differences schemes is the *von Neumann* stability analysis which uses tools from Fourier analysis [189, 131]. Here the solution to Eq. (3.22) is sought in the form

$$\phi_j^n = g^n \exp(\imath j\theta) \tag{3.26}$$

where $\theta = K\Delta x$ is the phase angle with K denoting the wave number, g^n is the amplitude of the Fourier component at $n\Delta t$, and $\imath$ is the imaginary unit. A numerical scheme is said to be von Neumann stable if there is a constant $c > 0$ (independent of θ, Δt, and Δx) such that

$$|G(\theta)| \leq c\Delta t, \qquad \text{for } 0 \leq \theta < 2\pi \tag{3.27}$$

where $|G(\theta)| = |g^{n+1}/g^n|$ is the modulus of the Fourier amplification factor. In practice, the inequality (3.27) in the von Neumann condition will be replaced by the stronger condition [189]

$$|G(\theta)| \leq 1, \qquad \text{for } 0 \leq \theta < 2\pi. \tag{3.28}$$

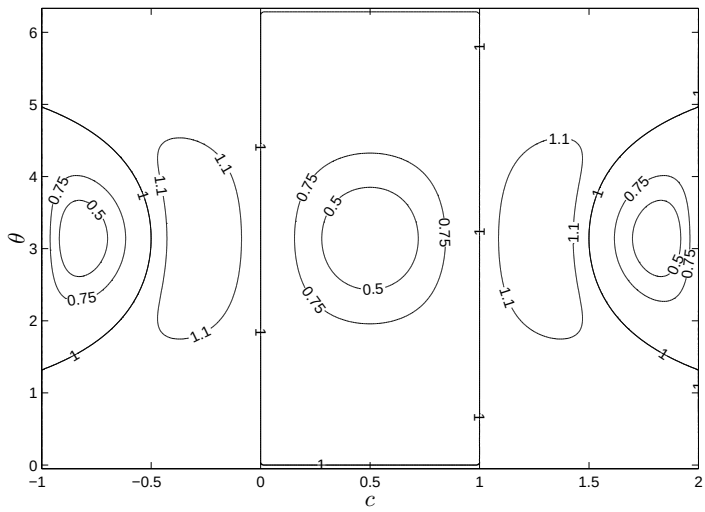


Figure 3.7: Two-dimensional plot of the amplification factor modulus, $G(\theta, c)$, for the Quickest scheme

The von Neumann criterion can now be applied to specify the stability region of the Quickest scheme. First, for reasons of simplicity, assume uniform grid spacing and zero source term. Then the time update for the jth node (Eq. 3.22) can be rewritten to

$$\begin{aligned}\phi_j^{n+1} = &-\left[\frac{1}{6}c(1-c)(2-c)\right]\phi_{j+1}^n + \left[\frac{1}{2}(1-c^2)(2-c)\right]\phi_j^n \\ &+\left[\frac{1}{2}c(1+c)(2-c)\right]\phi_{j-1}^n - \left[\frac{1}{6}c(1-c^2)\right]\phi_{j-2}^n\end{aligned} \tag{3.29}$$

where c is the Courant number as defined in (3.20). Substituting (3.26) into (3.29) yields the following amplification factor

$$\begin{aligned}G(\theta, c) = 1 &+ \left[\frac{1}{2} - \frac{1}{6}\exp(\imath\theta) - \frac{1}{2}\exp(-\imath\theta) + \frac{1}{6}\exp(-2\imath\theta)\right]c^3 \\ &+ \left[-1 + \frac{1}{2}\exp(\imath\theta) + \frac{1}{2}\exp(-\imath\theta)\right]c^2 \\ &+ \left[-\frac{1}{2} - \frac{1}{3}\exp(\imath\theta) - \frac{1}{6}\exp(-2\imath\theta) + \exp(-\imath\theta)\right]c.\end{aligned}$$

The modulus for different values of c and θ is depicted in Fig. 3.7. Is is evident that the Quickest scheme satisfies the von Neumann stability condition (3.28) for $0 \leq c \leq 1$. This restriction is also well known as the Courant-Friedrichs-Lewy or CFL condition [42]. Although being difficult to detect in Fig. 3.7, the Quickest scheme is also von Neumann stable for $c = 2$. The reader is referred to the original work [129] for a detailed analysis of stability.

Besides the scalar case the von Neumann condition can also be stated for finite-difference representation in matricial form:

Definition 3.1.3 (von Neumann stability)**.** *The finite-difference method* (3.25) *is said to satisfy the von Neumann condition if there exists a constant $c > 0$ independent of Δt, Δx and n such that*

$$\rho(\boldsymbol{A}) \leq 1 + c\Delta t \tag{3.30}$$

where $\rho(\boldsymbol{A})$ denotes the spectral radius of matrix $\boldsymbol{A}$.

Theorem 3.1.1. *The von Neumann criterion is necessary for stability of the finite-difference method* (3.25)*, and sufficient, if the following condition holds*

a all Elements of $\boldsymbol{A}$ *are bounded and all but one of the eigenvalues of* $\boldsymbol{A}$ *lie in a circle inside the unit circle, that is, all but one eigenvalue* λ_i *are such that* $|\lambda_i| \leq r < 1$.

Proof. See Theorem 2.2 and Theorem 2.3 in [189]. □

For practical reasons, as mentioned above, the condition (3.30) is considered for the case $t \to 0$. In particular the method (3.25) is stable for $\|A\| \leq 1$ where $\|A\|$ is the L_2-norm of the matrix $\boldsymbol{A}$. This can be seen from the following inequality (see [98, Ch. 5.6])

$$\rho(\boldsymbol{A}) \leq \|A^k\|^{1/k} \leq \|A\|, \quad \forall k \in \mathbb{N}. \tag{3.31}$$

However, this condition is sufficient for stability but not necessary. In fact, as demonstrated in [190] the norm of the matrix $\boldsymbol{A}$ is never less than one for the classical Quickest scheme with imposed boundary conditions as suggested in the original work [129]. Moreover, it was shown that the region enclosed by the condition $\|\boldsymbol{A}^{48}\| = 1$ is approximately the same as the von Neumann stability region.

3.1.3 Simulations

System Description

The proposed hydraulic and thermal methods are applied to the test case of the Tannheim DHN. Tannheim is located in Tyrol, Austria and is a typical tourist centre with about 1100 inhabitants. In 2009, 84 building objects, mainly consisting of private houses, a few hotels and some guest houses, were connected to this system. The length of the entire distribution network (supply plus return line) is about 8 km. For heat production a biomass boiler with a heat power of 2.5 MW is installed. During times of peak heat demand, an oil-fired boiler with 3 MW can be operated.

The distribution network itself, depicted in Fig. 3.8, has a tree structure with one root node, the biomass plant. Thus, the flow condition in the network can be computed by solving the linear system of equations (3.8) on page 14. The available data set consist of measurements of supply temperatures at the plant and consumer stations, as well as heat-load and flow rates with a resolution of 15 minutes. Furthermore, as the heat transport medium is water, a specific heat capacity $c_p = 4184\,\mathrm{J/kgK}$ and a density $\rho = 983\,\mathrm{kg/m^3}$ was chosen (see [212] and [29]). The ground temperature has a seasonal cycle and depends on the depth below the surface [91, p. 156]. In this work a seasonal cycle with an amplitude of $10\,^\circ\mathrm{C}$, i.e., $T_q = 5\,^\circ\mathrm{C}$ in the winter months and $T_q = -5\,^\circ\mathrm{C}$ in the summer months was used.

Network Aggregation. To reduce the complexity of the process model a physically simplified, aggregated version of the real network is used (see Fig. 3.8). More precisely, the number of pipes was reduced from 163 to 12 pipes, and only 7 aggregated consumers (AC) were used. For the aggregation task it is essential that the most important physical properties of the network are preserved, i.e.,

- Total heat-load, mass flow, water volume, and heat loss in the aggregated and original network should be equal.

In [126] a sophisticated aggregation method is developed in which the complexity of the district heating network is gradually reduced by changing a tree structure into a chain structure with no branches. They transformed the various parameters which define the branches from the real

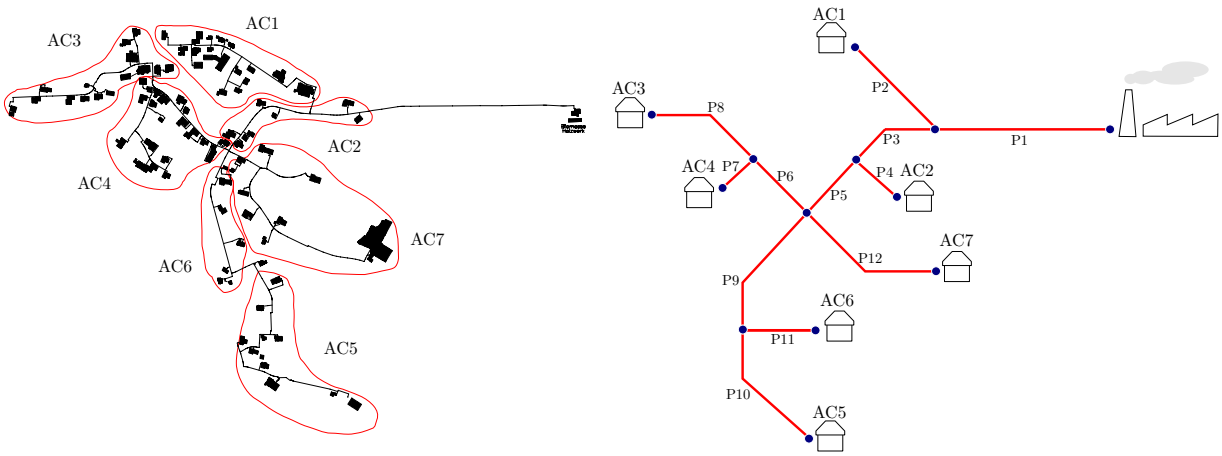


Figure 3.8: Original and aggregated district heating network in Tannheim

network to equivalent parameters in the corresponding aggregated network using conservation laws. Two main assumptions were made: The mass flows vary proportionally, i.e., there is a constant ratio between mass flows, and the return temperatures on the primary side of all heat exchangers are equal. Despite these very strong assumptions simulations have shown that the original network can be reduced from 1079 to approximately 10 branches without affecting the accuracy. Verification on the performance of aggregated networks is also done in [125]. They compared the method proposed in [126] with another systematic method developed in [142]. They concluded that both aggregation methods work well.

For all these methods it is required that a total description of the network exist. This means that the topology of the network and the physical properties of the pipe such as lengths, diameter and insulation have to be known. In this work no exact information about the individual heat loss coefficients of the pipes was available. Therefore, it was decided to estimate the heat loss coefficient by minimizing the error between the simulation output of the model and the process measurements. Theoretically, one could also estimate the pipe length and/or the pipe diameters simultaneously as, for instance, done in [166]. For minimizing the error the Levenberg-Marquardt least-squares algorithm [141] was attached (`lsqnonlin` command in MATLAB®). The lengths of the aggregated pipes were chosen in such a way to allow a uniform spatial grid spacing for the finite-difference model. The pipe diameters are then calculated to satisfy the conservation of the water volumes. The calculated and estimated parameters are given in Tab. 3.2. Two points should be mentioned here: First, to further reduce the number of unknown parameters equal heat losses were assumed for the pipes forming the main distribution line (P1, P3 and P5), as well as for the pipes P6, P8 and P9, P10. Secondly, it was assumed that the pipes in the return line have the same physical properties (length, diameter and heat loss) as the corresponding pipes in the supply line. For reasons of clarity the performed "aggregation steps" are summarized:

1. Find appropriate clusters by visual inspection and expert knowledge of the distribution network topology. Each cluster is represented by one pipe element.

2. Determine the length for each of these pipes elements.

3. Calculate the diameter for each pipe so that the total water volume is preserved.

4. Build aggregated time series for the mass flow rate, supply and return temperature. For the latter two use mass flow rate weighted averages.

Table 3.2: Physical properties of the pipes of the aggregated district heating network; Length, diameter are in m and heat loss in $\mathrm{W/m^2K}$

Pipe	Length	Diameter	Heat Loss
P1	800	0.180	0.86
P2	400	0.095	0.90
P3	200	0.140	0.86
P4	100	0.095	1.11
P5	100	0.140	0.86
P6	300	0.135	2.15
P7	200	0.085	1.31
P8	300	0.130	2.15
P9	400	0.120	1.08
P10	400	0.115	1.08
P11	200	0.090	0.85
P12	600	0.140	1.21

5. Find the heat loss coefficient for each pipe element by minimizing some quadratic error criterion using Levenberg-Marquardt.

Finite-Difference Model. For the finite-difference model a uniform spatial discretization with $\Delta x = 100\,m$ was chosen. To guarantee von Neumann stability for all flow conditions the time increments Δt were set to 60 s. Note that software tools used today in practice allow for adaptive mesh refinement (AMR) to guarantee stability and achieve high accuracy [137]. AMR automatically refines grids in regions where high resolution is required, as could be the case for pipes with high flow rates. The resulting finite-difference model for the supply line with the proposed discretization and network given by Fig. 3.1.3 + Tab. 3.2 is written for the supply line as

$$\begin{aligned} \boldsymbol{T}_{s,t+1} &= \boldsymbol{A}_{s,t} + \boldsymbol{B}_{s,t} T_{s,t}^{\mathrm{Plant}} + \boldsymbol{E}_{s,t} T_g \\ T_{s,t}^{\mathrm{AC}} &= \boldsymbol{C}\boldsymbol{T}_{s,t} \end{aligned} \tag{3.32}$$

with $\boldsymbol{A_s} \in \mathbb{R}^{40\times 40}$, $\boldsymbol{B_s} \in \mathbb{R}^{40\times 1}$, $\boldsymbol{C_s} \in \mathbb{R}^{7\times 40}$ and $\boldsymbol{E_s} \in \mathbb{R}^{40\times 1}$. And the model for the return line is given by

$$\begin{aligned} \boldsymbol{T}_{r,t+1} &= \boldsymbol{A}_{r,t} + \boldsymbol{B}_{r,t} \boldsymbol{T}_{r}^{\mathrm{AC}} + \boldsymbol{E}_{r,t} T_g \\ T_{r,t}^{\mathrm{Plant}} &= \boldsymbol{C}\boldsymbol{T}_{r,t} \end{aligned} \tag{3.33}$$

with $\boldsymbol{A}_r \in \mathbb{R}^{41\times 41}$, $\boldsymbol{B}_r \in \mathbb{R}^{41\times 7}$, $\boldsymbol{C}_r \in \mathbb{R}^{1\times 41}$ and $\boldsymbol{E}_r \in \mathbb{R}^{41\times 1}$. For simulation of the return line the measured aggregated return temperatures acts as input to the system. This means that the individual consumer stations (heat exchangers) are not explicitly modeled here. Furthermore, as the time resolution of the measurement data is 15 min and the finite-difference models are time discretized with 1 min, the data were up-sampled using zero-order-hold.

Results

To validate the accuracy of the finite-difference models (3.32) and (3.33) one week (19.-26. Dec. 2008) of real data are considered. More precisely, the simulation output at the boundary nodes representing AC5 and AC7 is compared with the corresponding aggregated supply temperature measurements. For performance assessment of the return line model measured return temperatures at the plant are available. Note that at each simulation time step, first the stationary flow distribution has to be calculated based on the available measurements of the flow rate at

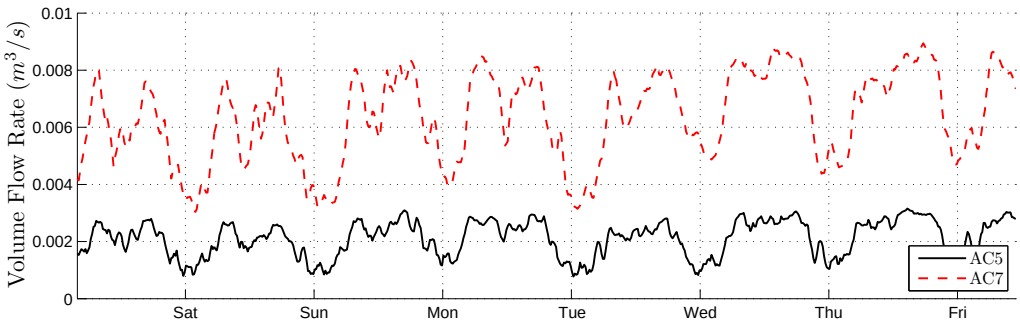


Figure 3.9: Aggregated volume flow rates for AC5 and AC7 between 19.-26. Dec. 2008

the aggregated consumers, and then the matrices in (3.32) + (3.32) have to be updated. Measurements of aggregated flow rates for the validation time period are depicted in Fig. 3.9. It can be observed that the flow rates vary proportionally with the corresponding heat demand of the consumers, with low demand during night and peaks in the morning hours and late afternoon (cf. Sec. 4.1.1).

The simulation results for the supply line are shown in Fig. 3.10. It can be seen that the model is able to capture the transient fluctuations in the supply temperature in spite of the fact that a strongly aggregated network was used. In particular for AC7 a good consistency between measurements and simulation can be observed. One reason for this are the high flow rates of AC7 and thus low transport time delays. This implies that significant discrepancies between simulated and measured temperatures can be anticipated for consumers located at distant pipes having low flow rates (cf. [65]). An interesting point worth mentioning here is that the rising edges of the supply temperature occur shortly after midnight. This can be explained by the implemented control strategy for the supply temperature at the biomass plant. Often plant operators "preheat" the distribution line in order to be able to provide sufficient heat in the peak time for the consumers.

The comparison of the simulation output with measured data for the return line in Fig. 3.11 shows very good agreement. One reason for this is that the return temperature is much smoother in contrast to the supply temperature due to the mass flow rate averaging (see Eq. (3.24)).

Of course, several actions can be taken to improve both the results for the supply and return line. Just to mention a few: aggregate the network to only a modest level, create denser spatial finite-difference grid, use more accurate heat loss coefficients for the pipes and take the heat transfer between supply pipes and return pipes into account, as it is done in [166].

For the purpose of demonstrating the effect of the finite-difference method different temperature profiles at the plant inlet node are considered, namely, a unit step and a single and double Gaussian hill (cf. [215]). The unit step represents a discontinuity in the temperature profile and is a standard benchmark test for every numerical scheme. The Gaussian hills, on the other hand, serve as an indicator on how a numerical scheme can resolve the interference of various peaks and valleys. The results are depicted in Fig. 3.12. Note that constant mass flow rate was given and $T_q = 0$, otherwise the temperature profile would never be zero due to continuous heat transfer. This figure clearly illustrates the transport time delays differences and resulting heat losses for AC5 and AC7. It also gives a good interpretation of the behavior of higher order finite-difference schemes, that is producing unphysical oscillations near discontinuities. First order schemes, on the other hand, are free from under and over estimations, but would produce strong artificial diffusion effects. It is referred to [77] for a detailed comparison of finite-difference schemes with different order.

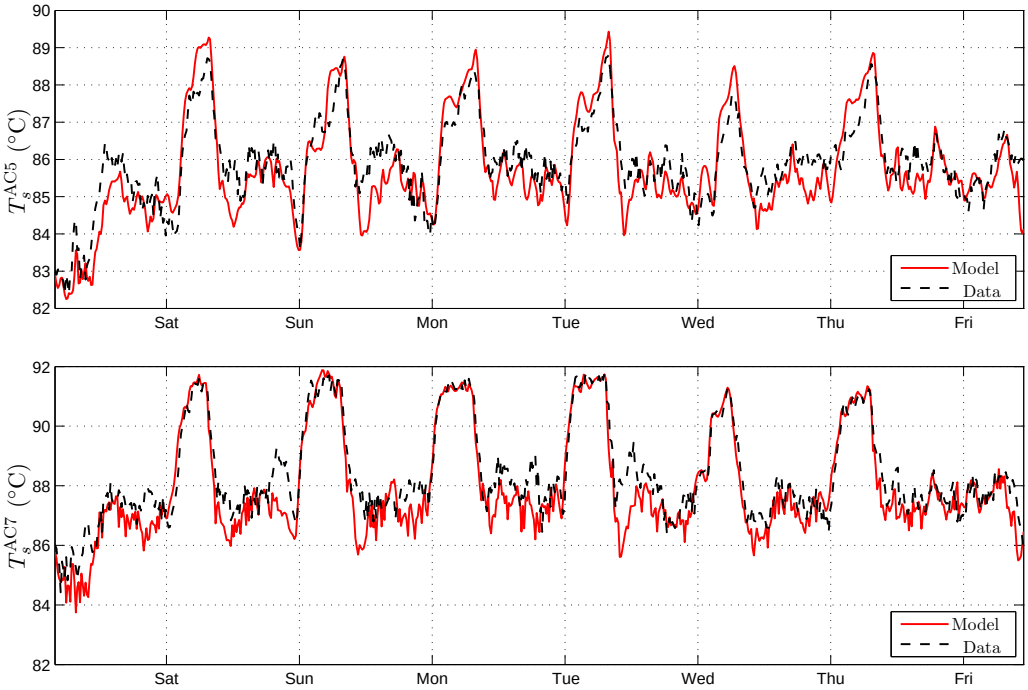


Figure 3.10: Performance comparison of the finite-difference model of the supply temperature at AC5 (top) and at AC7 (bottom) for one week of real data ($\sigma_\epsilon^{AC5} = 0.624\,°C$, $\sigma_\epsilon^{AC7} = 0.65\,°C$)

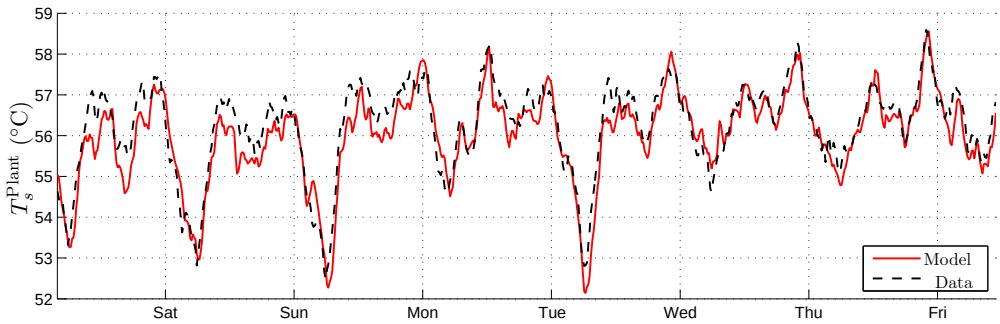


Figure 3.11: Performance comparison of the finite-difference model of the return temperature at at the biomass plant for one week of real data ($\sigma_\epsilon = 0.39\,°C$)

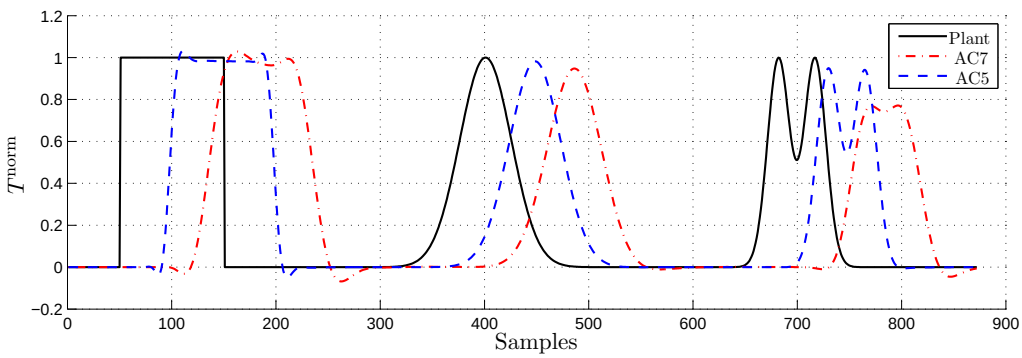


Figure 3.12: Different benchmark temperature profiles

3.2 Conditional Parametric Models

3.2.1 General Model Structure

Conditional parametric models are a special case of the varying-coefficient models which were introduced in [92]. Varying-coefficient models are considered to be linear regression models in which the coefficients are allowed to vary as smooth functions of individual explanatory variables. They are natural extensions of classical parametric models and are becoming more and more popular owing to their good interpretability and flexibility [58]. Parametric statistical inference always necessitates some model assumptions, such as linearity. Although their properties are very well established, linear models are often unrealistic in real applications [153, 57]. Varying-coefficient models, on the other hand, relax the conditions imposed on classical parametric models and try to explore the hidden structure in the data. Basically, varying-coefficient models can be considered as locally parametric models. That is why one also speaks about the class of semiparametric models [58]. In particular, conditional parametric models are varying-coefficient models in which all terms are modeled conditionally on the same argument [36]:

$$y_k = \boldsymbol{x}_k^{\mathrm{T}} \boldsymbol{\beta}(\boldsymbol{z}_k) + \epsilon_k, \quad k = 1, \ldots, N, \tag{3.34}$$

where y_k is the response, $\boldsymbol{x}_k = [x_{k,1}, \ldots x_{k,p}]$ and $\boldsymbol{z}_k = [z_{k,1}, \ldots z_{k,m}]$ are a set of explanatory variables, $\boldsymbol{\beta}(\,\cdot\,)$ is the vector of coefficient functions to be estimated and ϵ_k is some random noise term. In general, m, the dimension of $\boldsymbol{z}_k$, should be chosen low (1 or 2) for practical purposes [156]. Since (3.34) represents a linear model for a given $\boldsymbol{z}_k$, methods like locally weighted least squares (WLS) are a natural way to fit these types of models [93] (see Sec. 3.2.2). Note that in [102] varying-coefficient models are defined as given in Eq. (3.34), which is different from the original work [92] where the coefficients are explicitely allowed to depend on individual explanatory variables.

Specifically, for predicting the supply temperature in district heating networks a conditional parametric ARX (cARX) and a conditional parametric FIR (cFIR) model was proposed in [158] and [169], respectively. It is well-known that the response characteristics in district heating networks are heavily influenced by varying transport delays as a result of varying flow rates [76]. Within the framework of conditional parametric models this nonlinear influence is naturally accounted for by varying-coefficient functions. More precisely, the idea is to predict supply temperature values at particular nodes in the network, the critical nodes, based on supply temperature values at the plant depending on the flow rate. Mathematically, this can be formulated as cARX model as follows [158]

$$y_t = a_1(q_{t-1}) y_{t-1} + \sum_{j=\tau_{\min}}^{\tau_{\max}} \boldsymbol{b}_j(q_{t-1}) x_{t-j} + \epsilon_t, \quad \forall t. \tag{3.35}$$

Here, y_t is the supply temperature at the considered point in the network, q_{t-1} denotes the volumetric flow rate at the plant, x_{t-j} are the lagged values of the supply temperature at the plant, $\tau_{\min}$, $\tau_{\max}$ are minimum and maximum values for the time delay, respectively, and a_1, $b_{\tau_{\min} \ldots \tau_{\max}}$ are coefficient functions, which have to be estimated. Obviously, the coefficients in (3.35) depend on the flow rate in the network. This means that for different flow situations, e.g., high flow rate in winter months and low flow rate in summer months, a different set of parameters is used for modeling the temperature response. In this context the fitting points can also be considered as operating points and for different operating points different process behavior can be assumed.

In both publications, [158] and [169], it was demonstrated that the proposed class of conditional parametric models produce highly accurate results for simulation and prediction of

temperature transients in a district heating system. In [79, 75] the cFIR model was also successfully integrated into a fuzzy predictive control scheme. This topic will further be pursued in Sec. 7.1.

Conceptually similar models to the conditional parametric models are commonly referred to in literature as local linear model networks (LLMs) [151, 153], linear parameter-varying (LPV) systems [1], neuro-fuzzy models [153] or Takagi-Sugeno fuzzy models [177]. The last method will be introduced in Sec. 5.2 and the equivalence to the conditional parametric framework will be discussed.

3.2.2 Estimation

Problem Formulation

A common way to estimate conditional parametric models is by defining *fitting points* $\boldsymbol{z}_{(i)},\ i = 1 \dots r$ and estimate linear models in regions local to these fitting points [93]. Localization is achieved via a weighting kernel K_h, and individual observations $\boldsymbol{z}_k,\ k = 1, \dots, N$ receive weights $K_{h_i}(\boldsymbol{z}_{(i)}, \boldsymbol{z}_k)$. The size of the local neighbourhood around each fitting point is defined by the bandwidth, h, of the weighting kernel. Besides the identification of the set of relevant explanatory variables $\boldsymbol{z}$ and the optimal functional form of the conditional parametric model, the estimation problem can be considered to be threefold:

1. Selecting the fitting points
2. Selecting kernel type and bandwidth
3. Fitting the local linear models

First, the number and the placing of the fitting points has to be decided. In [153] this part is called partitioning of the premise input space, where the premise input space is denoted as the space spanned by $\boldsymbol{z}$. This notation relates to the rule structure of fuzzy models discussed in Sec. 5.2. The second point comprises the choice of the optimal type of kernel function and the selection of its bandwidth [89]. The trade-off in the bandwidth is well known as bias-variance dilemma [57]. For instance, if the bandwidth goes to zero, the estimates approach a piecewise-linear function that interpolates the training data. On the other hand, if the bandwidth gets infinitely large, the fit approaches the global linear least-squares fit of the data [93]. Finally, in step 3 the coefficient functions are estimated using least-squares methods.

Theoretically, these three steps can be combined, i.e., the learning of the parameter set $\{\boldsymbol{z}_{(i)}, h_i, \boldsymbol{\beta}_(i)\}_1^r$ is done simultaneously. This leads to a nonlinear global optimization problem, in which the objective function is in general non-convex with multiple local minima [93, 1]. Therefore, in practice, the first two steps and the third step are often treated separately. Several strategies for proper partitioning of the input space have been proposed in the literature (e.g., [169], [2], [138], [139]). In [169] a method based on the nearest neighbourhood (NN) principle was used. This approach requires knowledge of the empirical distribution function and more fitting points are located where the data are more dense. A comprehensive overview of clustering methods for fuzzy model identification is presented in [2]. They introduced a specific clustering algorithm for obtaining both the centers of the Gaussian membership (kernel) functions, and their standard deviations. Optimal kernel bandwidth selection may also be performed by cross-validation techniques [93, 89]. An approach which combines all three steps in an incremental way is the local linear model tree (LOLIMOT) algorithm proposed in [153].

Local Polynomial Estimates

The coefficient functions $\boldsymbol{\beta}(\cdot)$ in (3.34) are estimated in a nonparametric framework by using local polynomial regression [37]. The only assumption on the coefficient functions is that they are

sufficiently smooth for being locally approximated with polynomials. The estimation problem is reduced to locally fitting linear models at a number of fitting points $\boldsymbol{z}_{(i)},\ i = 1 \dots r$, where the local neighbourhood is determined by using Kernel functions. Following the description in [156] the problem can be formulated as a classical weighted least-squares problem

$$\hat{\boldsymbol{\phi}}_{(i)} = \arg\min_{\boldsymbol{\phi}_{(i)}} \sum_{k=1}^{N} K_{h_i}(\boldsymbol{z}_{(i)}, \boldsymbol{z}_k)(y_k - \boldsymbol{u}_k^T \boldsymbol{\phi}_{(i)})^2, \tag{3.36}$$

with

$$\boldsymbol{\phi}_{(i)} = \boldsymbol{\phi}(\boldsymbol{z}^{(i)}) = \begin{bmatrix} \boldsymbol{\phi}_{1,(i)}^T & \boldsymbol{\phi}_{2,(i)}^{\mathrm{T}} \dots \boldsymbol{\phi}_{p,(i)}^{\mathrm{T}} \end{bmatrix}^T, \tag{3.37}$$

$$\boldsymbol{u}_k^T = \begin{bmatrix} x_{1,k}\boldsymbol{p}_d^T & x_{2,k}\boldsymbol{p}_d^T \dots x_{p,k}\boldsymbol{p}_d^T \end{bmatrix} \tag{3.38}$$

where $\boldsymbol{\phi}_{(i)}$ is the vector of local coefficients at $\boldsymbol{z}_{(i)}$ and the element $\boldsymbol{\phi}_{j,(i)}^T$ is the vector of local coefficients related to the local polynomial approximation of the j-th explanatory variable. Furthermore, $\boldsymbol{p}_d = \boldsymbol{p}_d(\boldsymbol{z}_k)$ is the column vector corresponding to the d-order polynomial evaluated at $\boldsymbol{z}_k$. If, for instance, $\boldsymbol{z}_k = [z_{k,1}\ z_{k,2}]^T$ $(d = 2)$ the polynomial can be obtained as

$$\boldsymbol{p}_2(\boldsymbol{z}_k) = \begin{bmatrix} 1 & z_{k,1} & z_{k,2} & z_{k,1}z_{k,2} & z_{k,1}^2 z_{k,2}^2 \end{bmatrix}^T. \tag{3.39}$$

Moreover, the weighting kernel can be written in the form

$$K_{h_i}(\boldsymbol{z}_{(i)}, \boldsymbol{z}_k) = D(\|\boldsymbol{z}_k - \boldsymbol{z}_{(i)}\|/h_i), \quad \forall k, i, \tag{3.40}$$

where $\|\cdot\|$ is a appropriate norm, $D(\,\cdot\,)$ denotes the kernel function and h_i is the bandwidth for the particular fitting point. Two most commonly used norms are the Euclidean norm and the Mahalonobis norm [153]. Popular kernel functions are, for instance, the non-compact *Gaussian* density function and the compact *tricube* kernel [57]:

$$D(u) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}u^2}, \qquad D(u) = \frac{70}{81}(1 - |u|^3)^3 \mathbb{1}_{|u| \le 1}, \tag{3.41}$$

where $\mathbb{1}_{|u|\le 1}$ denotes the indicator function. For the Gaussian kernel h is equivalent to the standard deviation, whereas for the tricube kernel it determines the radius of the support region [93]. The multidimensional weighting kernel in case of axis-orthogonal Gaussians can also be written in compact form using the product operator [153]

$$K_{\boldsymbol{h}_i}(\boldsymbol{z}_{(i)}, \boldsymbol{z}_k) = \prod_{j=1}^{r} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{z_{k,j} - z_{j,(i)}}{h_{i,j}}\right)^2}. \tag{3.42}$$

Here it was assumed that $\boldsymbol{h}_i$ is a vector of bandwidths $\boldsymbol{h}_i = [h_{i,1}, \dots h_{i,r}]$. This means that a possibly different bandwidth parameter for each dimension of $\boldsymbol{z}_{(i)}$ can be applied. The consideration of different bandwidths for different directions increases the model flexibility and is an essential part of the partition strategy of the LOLIMOT algorithm. LOLIMOT will be further examinded in Sec. 3.2.3. From (3.42) it is easy to see that the fitting point $\boldsymbol{z}_{(i)}$ can be interpreted as the centre of the multidimensional, axis-orthogonal Gaussian kernel function.

The estimates of the coefficient functions obtained as described above are called *local polynomial estimates*. For the special case of constant polynomial approximation the term *local constant estimates* is used [156]. The vector of the coefficient function values $\boldsymbol{\beta}(\boldsymbol{z}_{(i)})$ is finally calculated from

$$\hat{\boldsymbol{\beta}}_{(i)} = \hat{\boldsymbol{\beta}}(\boldsymbol{z}_{(i)}) = \boldsymbol{p}_d^{\mathrm{T}}(\boldsymbol{z}_{(i)})\hat{\boldsymbol{\phi}}_{(i)}, \quad i = 1, \dots, r. \tag{3.43}$$

Coefficient function values evaluated at other points than the fitting points can be obtained by interpolation. Let $\tilde{w}_{(i)}$ be the normalized weights for the ith model, i.e,

$$\tilde{w}_{(i)} = \tilde{w}(\boldsymbol{z}_k, \boldsymbol{z}_{(i)}, h_i) = \frac{K_{h_i}(\boldsymbol{z}_{(i)}, \boldsymbol{z}_k)}{\sum_{j=1}^{r} K_{h_j}(\boldsymbol{z}_{(j)}, \boldsymbol{z}_k)} \tag{3.44}$$

then the interpolated coefficients at point $\boldsymbol{z}_k$ may be calculated from [153]

$$\tilde{\boldsymbol{\beta}}_j(\boldsymbol{z}_k) = \sum_{i=1}^{r} \tilde{w}_{(i)} \hat{\boldsymbol{\beta}}_{j,(i)}, \quad j = 1, \ldots, p. \tag{3.45}$$

Following the same idea, the approximation of the overall model output can be written as the sum of contributions of the individual local model outputs

$$\hat{y_k} = \sum_{i=1}^{r} \tilde{w}_{(i)} \underbrace{\left[x_{k,1}\beta_{1,(i)} + x_{k,2}\beta_{2,(i)} + \cdots + x_{k,p}\beta_{p,(i)}\right]}_{\hat{y}_{k,(i)}}, \quad k = 1, \ldots, N. \tag{3.46}$$

In conclusion, it should be mentioned that although higher order polynomial approximation such as second order might, from a statistical point of view, be beneficial, e.g., reducing bias at the boundaries [93], in this work local constant estimates are considered. This is due to the fact that the conditional parametric models are used for predictive control purposes in a closed-loop setting where parsimonious models are desired.

Adaptive Estimation

On-line estimation of the coefficient function is particular useful if the identified model is utilized for controller purposes [1, 213]. Here a new model will be identified within each sampling instant exploiting the new information contained. Hence, model mismatch due to possible time-varying behaviour of the process can be successfully eliminated. The weighted recursive least-squares method (WRLS) is well suited for this task. Using exponential forgetting and assuming observations at time $t = 1, \ldots, k$ are available, the weighted batch least-squares problem (3.47) for the ith model can be reformulated to obtain the WRLS estimates of $\phi_{(i)}$ as follows [156, 169]

$$\hat{\boldsymbol{\phi}}_{k,(i)} = \arg\min_{\boldsymbol{\phi}_{(i)}} \sum_{t=1}^{k} \alpha^{k-t} K_{h_i}(\boldsymbol{z}_{(i)}, \boldsymbol{z}_t)(y_t - \boldsymbol{u}_t^T \boldsymbol{\phi}_{(i)})^2, \tag{3.47}$$

where $\lambda \in (0, 1]$ is called the forgetting factor. Now, by substituting $\tilde{\boldsymbol{u}}_{k,(i)}^T = \boldsymbol{u}_k^T \sqrt{K_{h_i}(\boldsymbol{z}_{(i)}, \boldsymbol{z}_k)}$ and $\tilde{y}_{k,(i)} = y_k \sqrt{K_{h_i}(\boldsymbol{z}_{(i)}, \boldsymbol{z}_k)}$, the adaptive estimates in (3.47) can be found by a classical recursive least squares scheme. Note that in the remainder of this section the subscript (i) for the ith model is dropped for notational simplicity. The RLS algorithm presented in [219] is given as

$$\begin{aligned}
\hat{\boldsymbol{\phi}}_k &= \hat{\boldsymbol{\phi}}_{k-1} + \boldsymbol{g}_k \left(\tilde{y}_k - \tilde{\boldsymbol{u}}_k^T \hat{\boldsymbol{\phi}}_{k-1}\right) \\
\boldsymbol{g}_k &= \boldsymbol{P}_{k-1} \tilde{\boldsymbol{u}}_k \left(\lambda + \tilde{\boldsymbol{u}}_k^T \boldsymbol{P}_{k-1} \tilde{\boldsymbol{u}}_k\right)^{-1} \\
\boldsymbol{P}_k &= \frac{1}{\lambda} \left(\boldsymbol{P}_{k-1} - \boldsymbol{g}_k \tilde{\boldsymbol{u}}_k^T \boldsymbol{P}_{k-1}\right).
\end{aligned} \tag{3.48}$$

The matrix $\boldsymbol{P}_k$ is proportional to the covariance matrix of the parameter estimates. Hence, if little prior knowledge about the initial parameter values $\boldsymbol{\phi}_0$ is available, the initial value of $\boldsymbol{P}_k$ is usually chosen as $\boldsymbol{P}_0 = \alpha \mathbf{I}$ with large values for α. This reflects the high uncertainty in the initial parameter values. Note that $\mathbf{I}$ represents the identity matrix.

The forgetting factor is in most cases between 0.9 and 1 and is considered to be a tuning parameter. It specifies the trade-off between robustness against disturbances (large λ) and fast tracking capability (small λ) [153]. This trade-off can also be made dynamically by choosing a variable forgetting factor. For instance, in the case of no persistent or low excitation of the ith model the covariance matrix will grow exponentially, since then $\tilde{\boldsymbol{u}}_k^T \boldsymbol{P}_{k-1}$ approaches zero in (3.48), and thus $\boldsymbol{P}_k \approx \frac{1}{\lambda}\boldsymbol{P}_{k-1}$. To overcome this, one can apply an effective forgetting factor for each local model as introduced in [156]:

$$\lambda_k^{\text{eff}} = 1 - (1 - \lambda)\, K_h, \tag{3.49}$$

where K_h denotes the weighting kernel as defined in (3.40). Hence, only the the active, persistently excited local models are adapted. Similar arguments lead to the variable forgetting factor discussed in [153, p. 715 f.]:

$$\lambda_k = 1 - \left(1 - \tilde{\boldsymbol{u}}_k^T \boldsymbol{g}_k\right) \frac{e_k^2}{\Sigma_0} K_h, \tag{3.50}$$

where $e_k = \tilde{y}_k - \tilde{\boldsymbol{u}}_k^T \hat{\boldsymbol{\phi}}_{k-1}$ is the local model error for the corresponding local model and Σ_0 is proportional to the assumed noise variance and is considered for tuning. The idea is that the squared model error e_k in (3.50) should slow down the adaptation if the error is small. Furthermore, the quantity $\left(1 - \tilde{\boldsymbol{u}}_k^T \boldsymbol{g}_k\right)$ is a measure for the current excitation of the process [62]. Additionally, it was proposed to bound the forgetting factor, i.e., $\lambda^{min} < \lambda < \lambda^{max}$ for this approach. In general, one can also apply the rule to only update parameter values if the weight K_h exceeds some predefined threshold. This makes sense if kernels with non-compact support such as Gaussian Kernels are used.

Regularization. The proposed model architecture will be used to produce multi-step-ahead forecasts in a predictive control setup. Hence, in order to improve the generalization ability, regularized estimates can be considered [169, 93, 141, 103]. This can be done, for instance, by introducing an normalization step after the update of the matrix $\boldsymbol{P}_k$ in (3.48) [83]:

$$\bar{\boldsymbol{P}}_k = \boldsymbol{P}_k(\mathbf{I} + \alpha \boldsymbol{P}_k)^{-1} \tag{3.51}$$

where $\alpha \geq 0$ denotes the regularization parameter. The normalization step (3.51) requires a matrix inversion and the multiplication of two full rank matrices, and therefore simpler algorithms were proposed. For an overview see [83] and the references therein.

Constrained Estimation. Often a priori knowledge about the process to be modeled such as process gains, settling times or open-loop stability is available. A straightforward approach to include this knowledge into the adaptive estimation scheme (3.48) is by defining linear inequalities on the parameter set of the local linear models [1]. To be more specific, the equality and inequality constraints for each local model can be formulated as

$$\begin{aligned} \boldsymbol{M}\boldsymbol{\phi} &= \boldsymbol{k} \\ \boldsymbol{L}\boldsymbol{\phi} &\leq \boldsymbol{c}. \end{aligned} \tag{3.52}$$

For details see [199]. The constrained solution of the WRLS identification can now be obtained by the optimal projection of the unconstrained solution:

$$\hat{\boldsymbol{\phi}}_k^{\text{con}} = \hat{\boldsymbol{\phi}}_k - \boldsymbol{P}_k \boldsymbol{M}^{\mathrm{T}} \boldsymbol{\mu} - \boldsymbol{P}_k \boldsymbol{L}^{\mathrm{T}} \boldsymbol{\nu}, \tag{3.53}$$

where $\hat{\boldsymbol{\phi}}_k^{\text{con}}$ denotes the constrained solution, $\hat{\boldsymbol{\phi}}_k$ and $\boldsymbol{P}_k$ are obtained from (3.48) and $\boldsymbol{\mu}, \boldsymbol{\nu}$ are vectors of Lagrange multipliers associated with equality and inequality constraints, respectively.

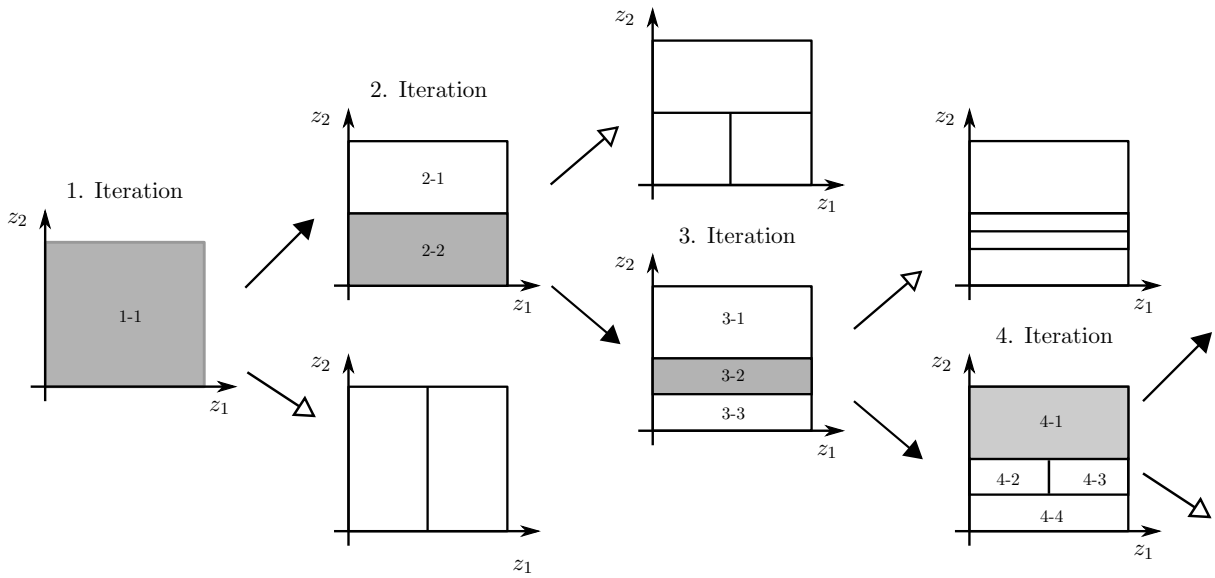


Figure 3.13: First four iterations of LOLIMOT algorithm for a two-dimensional input space ($m = 2$)

These vectors can be determined by quadratic-programming techniques such as complementary linear programming (CLP). Here the CLP to be solved takes the form [199]:

$$\begin{bmatrix} \boldsymbol{\sigma} \\ \boldsymbol{\varsigma} \end{bmatrix} = \begin{bmatrix} \boldsymbol{M}\boldsymbol{P}_k\boldsymbol{M}^T & \boldsymbol{M}\boldsymbol{P}_k\boldsymbol{L}^T \\ \boldsymbol{L}\boldsymbol{P}_k\boldsymbol{M}^T & \boldsymbol{L}\boldsymbol{P}_k\boldsymbol{L}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\nu} \end{bmatrix} + \begin{bmatrix} \boldsymbol{k} - \boldsymbol{M}\hat{\boldsymbol{\phi}}_k \\ \boldsymbol{c} - \boldsymbol{L}\hat{\boldsymbol{\phi}}_k \end{bmatrix} \tag{3.54}$$

with the conditions

$$\begin{gathered} \boldsymbol{\sigma} = 0, \quad \boldsymbol{\varsigma} \geq 0, \quad \boldsymbol{\nu} \geq 0 \\ \boldsymbol{\sigma}^T \mu + \boldsymbol{\varsigma}^T \boldsymbol{\nu} = 0 \end{gathered} \tag{3.55}$$

and where $\boldsymbol{\sigma}$ and $\boldsymbol{\varsigma}$ are the slack variables for the imposed constraints (3.52).

Remark. *It is important to mention that the constraints formulated in* (3.52) *only make sense if local constant estimates are used.*

3.2.3 Local Linear Model Tree (LOLIMOT)

The local linear model tree (LOLIMOT) algorithm combines a heuristic strategy for input space (the space spanned by $\boldsymbol{z}$) decomposition with weighted least squares optimization [153]. In LOLIMOT Gaussian kernel functions are fitted to a rectangular partitioning of the input space performed by a decision tree with axis-orthogonal splits at the internal nodes. Each local model belongs to one hyper-rectangle in which center the fitting point is placed. The standard deviations (i.e., kernel bandwidths) are set proportional to the size of the hyper-rectangle. This allows one model to be valid over a wide operating range of one variable but valid only in a small area of another one [94]. New hyper-rectangles are found by testing the possible splits in all dimensions and taking the one with the highest performance improvement. The algorithm stops when reaching a predetermined modeling error or maximum size of the tree. For a better overview the basic algorithmic concept of LOLIMOT is summarized below (for details see [153, p. 365f]):

1. Fit an initial local model network with given input space decomposition by weighted least squares. If no input space partitioning is available a priori then start with a global linear model.

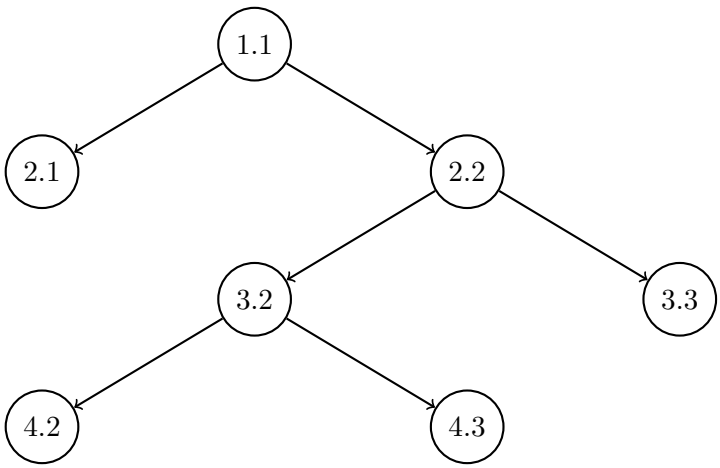


Figure 3.14: Interpretation of Fig. 3.13 as a decision tree

2. Find the worst local linear model by calculating the local loss function for each model.
3. The hyper-rectangle of the worst local linear model is split into two halves by doing an axis-orthogonal split. Divisions in all dimensions are tried.
4. Fit local linear models for all possible alternative splits by weighted least squares.
5. Compare the alternative fits and implement the one split with the highest performance improvement.
6. If the termination criterion is met then stop, else go to Step 2.

The calculation of the error in Step 2 is based on the local sum of squared error loss function and not their mean is utilized. Hence, splits are preferred in regions that contain more data samples. Thus, the local model quality depends on the distribution of the training data. This consequence is desired because more data allow to estimate more parameters at the same accuracy. In particular, as argued in [153], two intrinsic features make LOLIMOT extremely fast: First, at each iteration only the worst local model is considered for division. Second, in Step 3 only the parameters of the local models that are newly generated by the division have to be estimated. Figure 3.13 illustrates four iterations of the LOLIMOT algorithm for a two-dimensional input space. This can also be represented as a decision tree as shown in Fig. 3.2.3. For instance, in the second iteration the local model 2.2 is found to be the worst and is therefore considered for a further split.

As mentioned in [94], the tree construction algorithm of LOLIMOT to determine the centres and standard deviations of the weighting functions exploits ideas from CART [25] and MARS [63, 93]. A method based on decision trees for identification is also proposed in [2]. Here fuzzy clustering is used for the effective partitioning of the input domains of decision trees. The work presented in [121] for identification of radial basis function networks shows also analogies to LOLIMOT. In both methods the incremental building of the trees makes splits, which always halve the rectangular partitioning and which are sensitive to the distribution of the data.

3.2.4 Identification Results

For estimating and validation purposes the aggregated time series of supply temperature and flow of the district heating network in Tannheim was used. The time horizon of the data considered

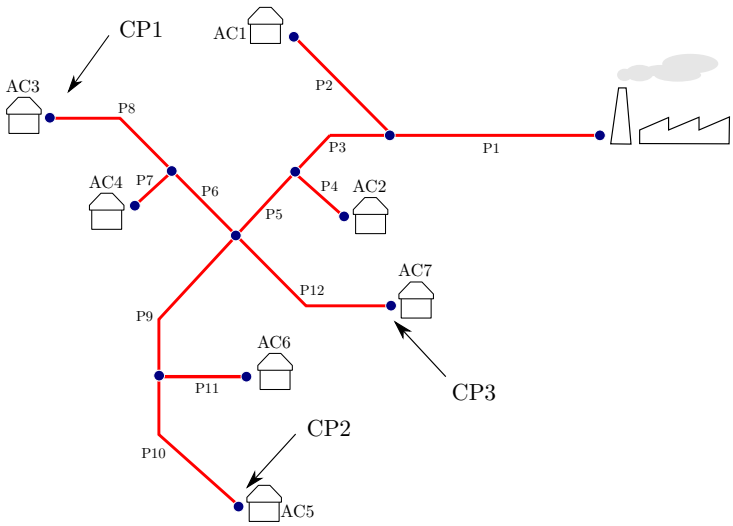


Figure 3.15: Original and aggregated district heating network in Tannheim

for identification is from 1.1.2008 - 31.12.2009 with a resolution of 15 minutes. The first half of the data was used as training data and the second half for validation.

In this work the concept of *critical nodes* is followed, i.e., if supply temperature and pressure is maintained within the contractual limits at this particular node, than it is satisfied at all other points in the network (cf. [169]). On the one hand, of course, this approach reduces the number of models which have to be estimated, since only the supply temperature at the critical node is of interest. On the other hand, this also has practical relevance, as typically only some points in the network are monitored and measurements are not available for all consumers. In this work three points (aggregated consumers) of the reduced distribution network as depicted in Fig. 3.2.4 were considered to be critical, namely CP1 and CP2 which are the most distant consumers, and CP3 which accounts for approximately 35% of the heat consumption in the network.

Data analysis. First, the data set was visually checked for outliers and missing values. Then, in order to obtain a valid data set, all these observations were treated by using robust local regression [37] as suggested in [155]. To be more precise, the time series was first robustly smoothed using local regression (command `smooth` in Matlab) and afterwards the missing values and outliers of the original time series were replaced with the corresponding smoothed observations. Another way to treat missing values is by simple linear interpolation or by excluding the affected data samples from the identification data set. The time series were also down-sampled to half-hourly averages since 30 minutes represents a typical time frame for operational control of DHN [191]. In Sec. 7.1 the conditional parametric models will be incorporated into a predictive control scheme for controlling the supply temperature at the critical nodes.

Table 3.3 summarizes the statistics of the individual supply temperature time series as well as the overall percentage of valid data. As one can see the average supply temperature is much lower for consumers which are connected to distant pipes. Also it seems that there is more volatility in the temperature at aggregated consumer 3 and 5. The minimal supply temperature for these consumers is about 30 °C. This could be explained such that during summer except for tap water no heat is required, and therefore the resulting flow rates are almost zero and thus the transport time delays and heat losses increase. In addition, the density of the volume flow rate at the plant is depicted in Fig. 3.16, since the distribution of the training data has a strong influence on the behavior of the LOLIMOT algorithm.

Table 3.3: Summary of the supply temperature statistics of the identification data set

Data	**Min**	**1. Qu.**	**Median**	**Mean**	**3. Qu.**	**Max**	**NA's**
T_s^{Plant}	76.1833	88.204	90.167	89.8698	91.817	97.433	1.46%
T_s^{AC3} (CP1)	32.95	79.033	83.567	80.4995	85.65	92.5667	3.00%
T_s^{AC5} (CP2)	35.333	77.883	82.55	79.4136	84.633	90.683	1.83%
T_s^{AC7} (CP3)	76.333	85.9	87.2	87.0616	88.433	94.1	2.29%

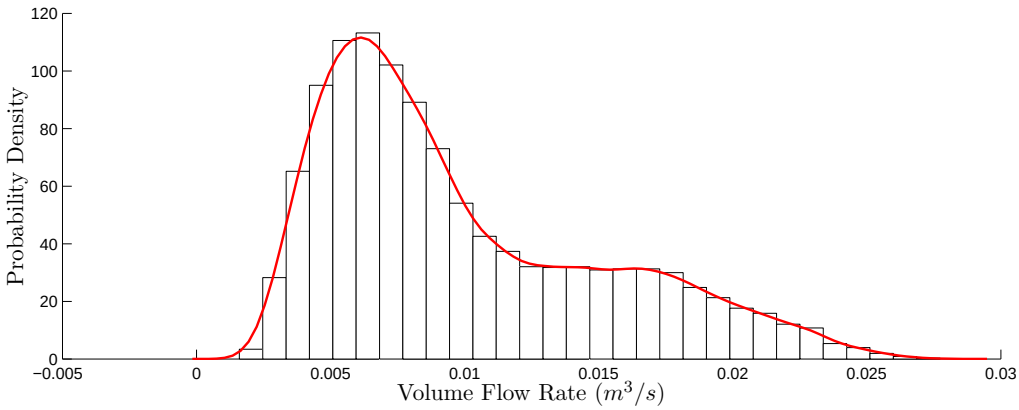


Figure 3.16: Histogram plus estimated density of the volume flow rate

Modeling and Identification. For predicting the supply temperature at the three critical points cARX models and cFIR models were constructed. The following cARX model was chosen for all critical points

$$T_{s,t}^{\text{ACi}} = a_1(q_{t-1})T_{s,t-1}^{\text{ACi}} + \sum_{j=1}^{8} b_j(q_{t-1})T_{s,t-j}^{\text{P}} + c_t(q_{t-1}) + \epsilon_t, \quad t = 1, \ldots, N, \tag{3.56}$$

where T_s^{ACi} is the supply temperature at critical point i and T_s^{P} is the supply temperature at the plant. For the maximal order of the lagged supply temperature at the plant 4 hours was considered to be sufficient. In [76] the mean transport time delays for distant consumers in Tannheim were calculated to be around 70 min. Of course, one can also choose different orders for different critical points. Here expert knowledge may be incorporated. Theoretically, different model types for different flow rates can be constructed. See also [153, p. 694], where it was argued that for high flow rates, transport processes typically have a pure dead time behaviour. The cFIR model for all three critical points is given by

$$T_{s,t}^{\text{ACi}} = \sum_{j=1}^{16} b_j(q_{t-1})T_{s,t-j}^{\text{P}} + c_t(q_{t-1}) + \epsilon_t, \quad t = 1, \ldots, N. \tag{3.57}$$

In both models the term $c_t(q_{t-1})$ represents the offset term in which Fourier harmonics of 24 hours period are embedded in order to take account of the diurnal variation of the heat-load, i.e.,

$$\begin{aligned} c_t(q_{t-1}) = c_0(q_{t-1}) + c_1(q_{t-1}) \sin\left(\frac{2\pi h_t}{48}\right) \\ + c_2(q_{t-1}) \cos\left(\frac{2\pi h_t}{48}\right), \end{aligned} \tag{3.58}$$

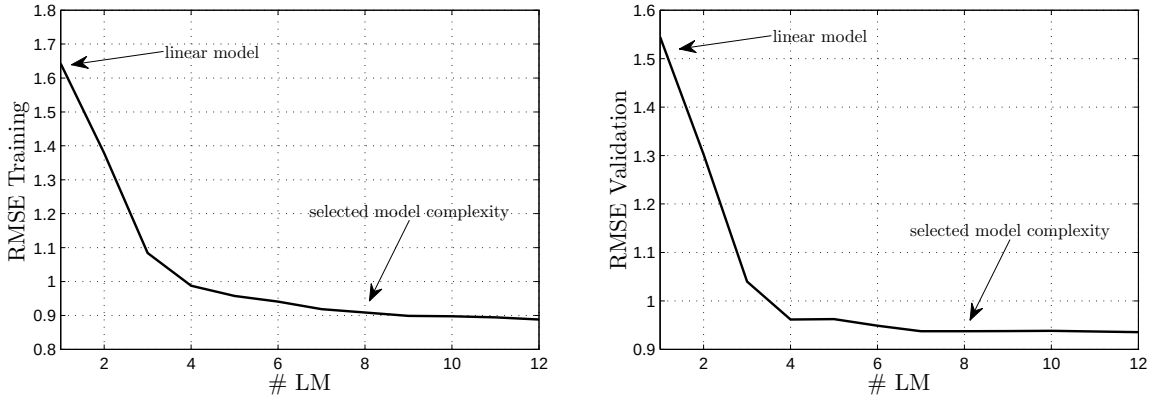


Figure 3.17: Root mean squared simulation error for the training dataset and the validation dataset against the number of local models. Results are shown for the cARX model and CP3

where $h_t = 0, \ldots, 47$ is the actual half-hour of the day. The trigonometric offset term was motivated from the social behavior of consumers (cf. [169]). It can be assumed that each critical point (aggregated consumer) has its own individual daily supply temperature pattern and the offset term (3.58) should take account of this. Similar arguments can be applied to the modeling of the heat-load as it is done in Sec. 4.1. Further note that the coefficient functions of all models were estimated by using local constant polynomials.

Remark. *The volume flow rate q_{t-1} could additionally be filtered by, for instance, exponential smoothing as suggested in [158]. Theoretically, the filter should average over past values with varying horizon owing to the varying transport time delays. The node method mentioned in Sec. 3.1 uses a similar concept. Here the number of volume flow elements filling a pipe is calculated [64].*

To find the optimal number of local linear models r the simulation error can be calculated. In Fig. 3.17 the root mean squared error (RMSE) for the cARX model at critical point 3 (AC7) is plotted against the number of local linear models. Similar results can be expected for the other critical points and the cFIR model. Theoretically, a minimum of the curve of simulation error for the validation data set should indicate the optimal number of local linear models. This implies that the model produces a good fit but does not overfit the training data. In this case it was found that 8 local models are adequate. This number was taken for both cARX and cFIR models and for all critical points. Of course, one could also use information criteria such as AIC to determine the optimal number of local models and to avoid overfitting. For more details on this topic, see, e.g., [153, p. 158 f.].

Figure 3.18 illustrates the partitioning of the volume rate in different regions determined by the LOLIMOT algorithm. On the y-axis the degree of memberships are plotted which are the normalized weights calculated from (3.44). It can be observed that the membership functions are denser for lower volume flow rates. This reflects the fact that first, the nonlinearity of the process is stronger in this operating regime, and second that LOLIMOT utilizes the local sum of squared error loss function for performance comparison. Hence, splits are preferred in regions that contain more data samples (cf. Fig. 3.16).

Validation. One way to validate the estimation results is by calculating the stationary gain for different flow conditions as shown in Fig. 3.19. Recall that the heat losses in the network are proportional to the flow rates. These heat losses can be expressed as the stationary gain of

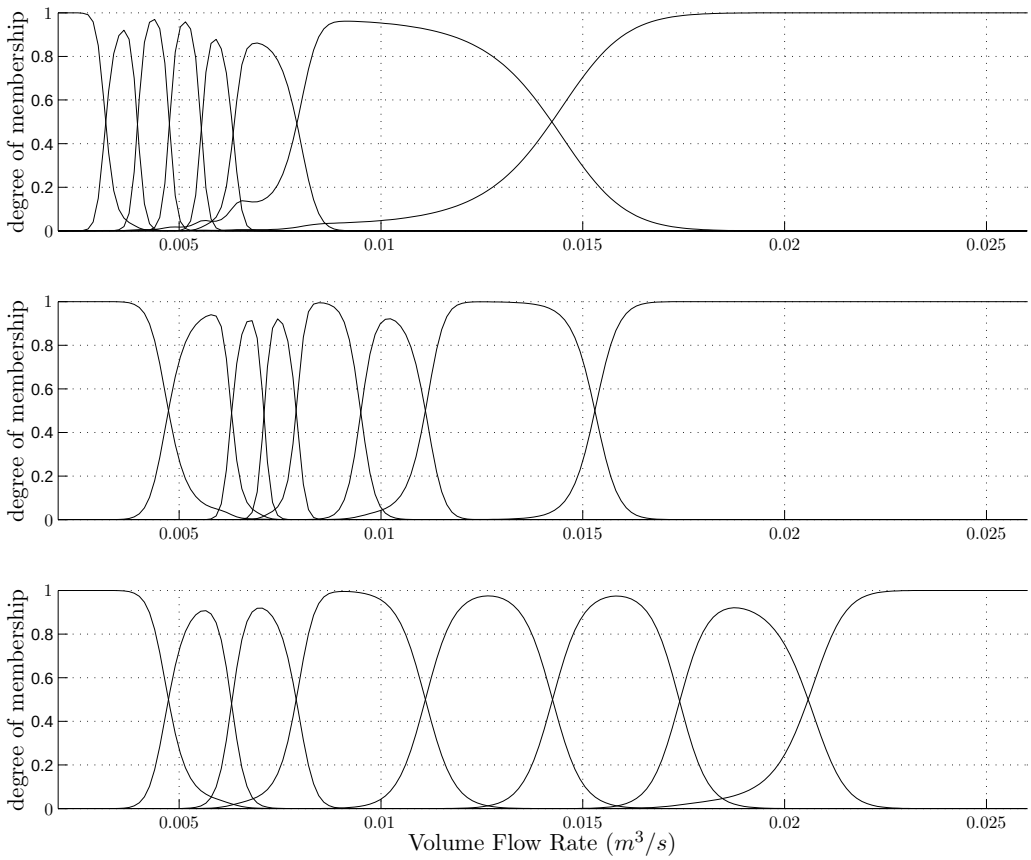


Figure 3.18: Membership functions generated by LOLIMOT of CP1 (AC3), CP2 (AC5) and CP3 (AC7)

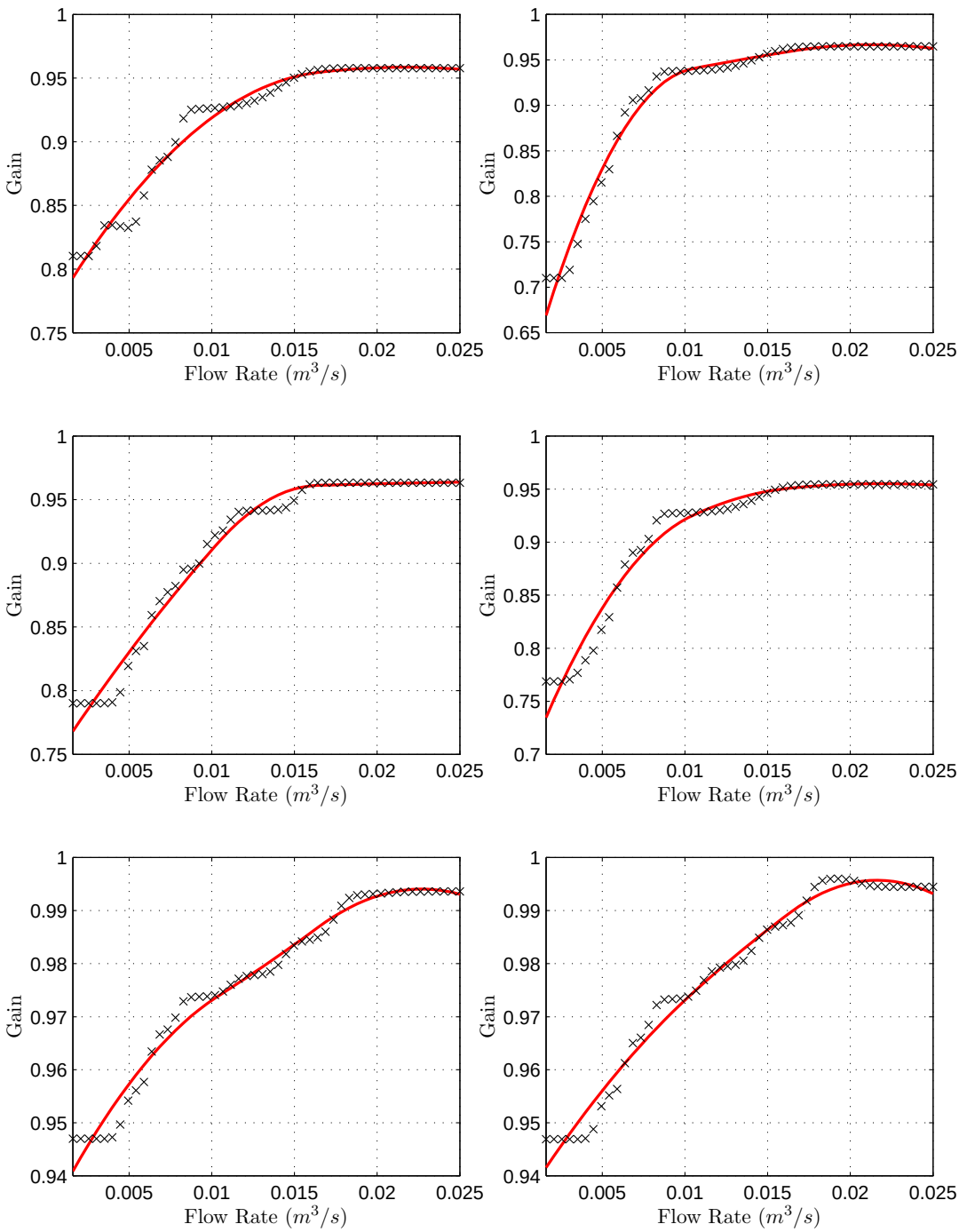


Figure 3.19: Stationary gain against volume flow rate of CP1 (AC3), CP2 (AC5) and CP3 (AC7) for cARX (left) and cFIR models (right)

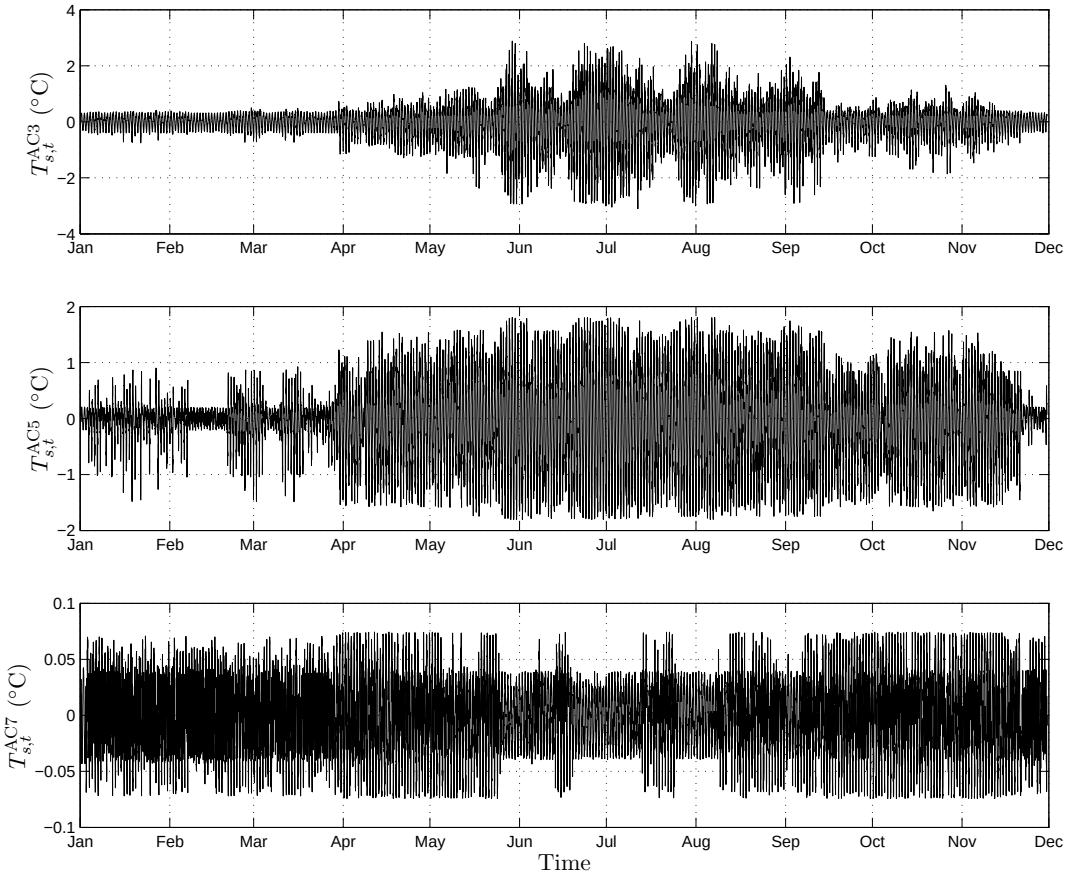


Figure 3.20: Estimated social component for the three critical points

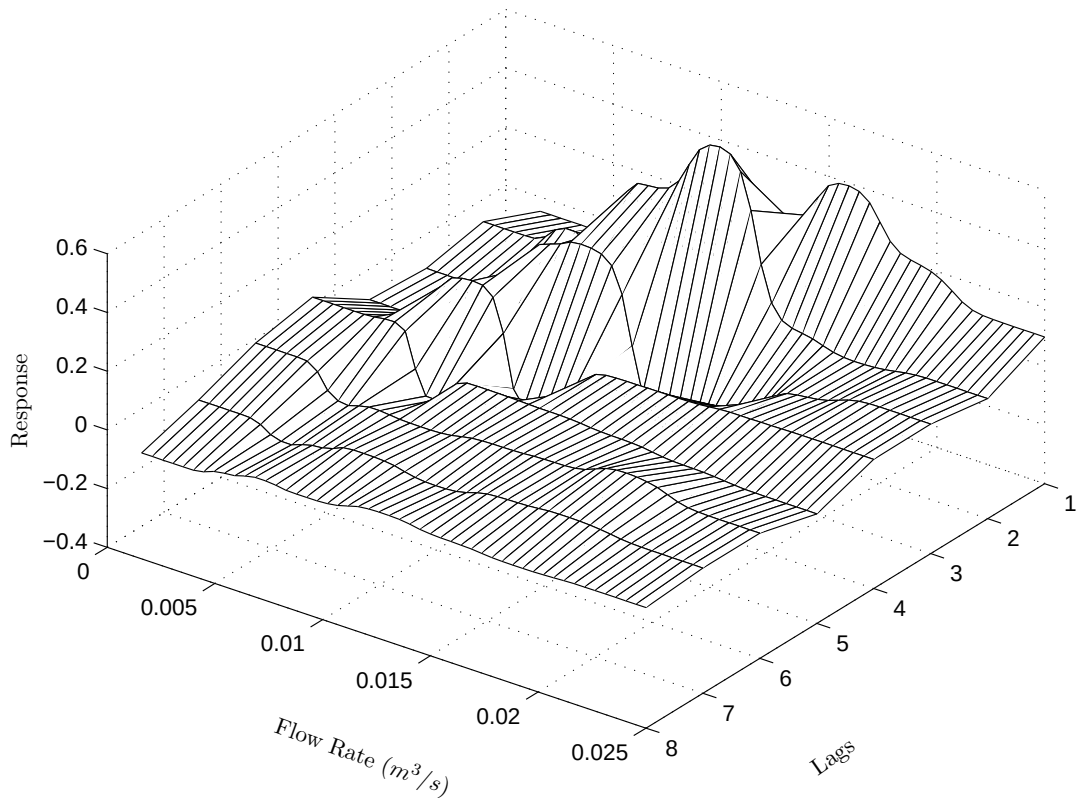


Figure 3.21: Impulse response function of the cARX model for critical point 3 (AC7)

the individual local model. The stationary gain for the cARX model is given by (neglecting the trigonometric offset)

$$K^{\mathrm{cARX}}(q) = \frac{\sum_{j=1}^{8} b_j(q)}{1 - a_1(q)} T_s^{\mathrm{p}} + \frac{c_0(q)}{1 - a_1(q)} \tag{3.59}$$

and for the cFIR model by

$$K^{\mathrm{cFIR}}(q) = \sum_{j=1}^{16} b_j(q) T_s^{\mathrm{p}} + c_0(q). \tag{3.60}$$

In Section 7.1 heat losses are related to operational costs in district heating networks. In this section the gain curves presented in Fig. 3.19 are utilized for generating the optimal set-point for the predictive controller by minimizing the operational costs.

The social component (trigonometric offset) of the three critical points considered is shown in Fig. 3.20. It can be observed that the social influence on the supply temperature is more pronounced for AC3 and AC5. Also, as mentioned above, the amplitude is higher during summer time. This makes it somewhat difficult to produce accurate forecasts, since forecast errors will be dominated by the social component which is considered to be a measurable disturbance acting on the process.

For illustration purposes, the impulse response function of the cARX model for AC7 is plotted in Fig. 3.21. This figure clearly demonstrates the influence of the flow rate on the local parameter estimates. Whereas for low flow rates the "rock mass" occurs near higher lag terms, it is reversed for high flow rates.

Forecasting. To assess the forecasting performance of the cARX and cFIR models multi-step-ahead prediction errors for the period between 1.1.2009 and 31.3.2009 ($\approx$ 4300 samples) were calculated. Here, in particular, only the winter months are used. This was found to be necessary due to the strong social consumer pattern, which is considered to be a disturbance, of AC3 and AC5 in summer (cp. Fig. 3.20). Otherwise a fair performance comparison would not be possible. Another argument for this was that the focus lies on evaluating the predictive performance of the models with regard to future process behaviour and not disturbances. As prediction performance measure, the root mean squared error and the mean absolute error (MAE) are taken which are defined for a k-step-ahead prediction as follows:

$$\mathrm{RMSE}(k) = \sqrt{\frac{1}{N} \sum_{t=1}^{N} \left(\hat{y}_{t+k} - y_{t+k}\right)^2}, \tag{3.61}$$

$$\mathrm{MAE}(k) = \frac{1}{N} \sum_{t=1}^{N} \left|\hat{y}_{t+k} - y_{t+k}\right|. \tag{3.62}$$

The average of all these k-step-ahead prediction errors is then calculated and used to determine the overall k-step-ahead prediction performance. Note that for multi-step forecasts both the flow and future values of the supply temperature and volume flow at the plant have to be treated as unknown. The supply temperature, on the one hand, acts as control input in the predictive control framework introduced in Sec. 7.1 and can be assumed to be known. Future flow rate values, on the other hand, cannot be exactly known and therefore an appropriate model has to be constructed. In [169] an ARX model was suggested:

$$q_t = \sum_{j=1}^{p} a_j q_{t-j} + \epsilon_t, \quad \forall t \tag{3.63}$$

Table 3.4: Prediction results of cFIR models with observed flow rate measurements as input

	CP1 (AC3)		**CP2 (AC5)**		**CP3 (AC7)**	
k	RMSE	MAE	RMSE	MAE	RMSE	MAE
1	1.0381	0.7465	1.0109	0.6963	0.7799	0.5051
4	1.0709	0.7952	1.0507	0.7470	0.8001	0.5294
8	1.0712	0.8014	1.0633	0.7624	0.8038	0.5332
12	1.0728	0.8029	1.0631	0.7620	0.8040	0.5339
16	1.0723	0.8002	1.0656	0.7592	0.8065	0.5354

Table 3.5: Prediction results of cARX models with observed flow rate measurements as input

	CP1 (AC3)		**CP2 (AC5)**		**CP3 (AC7)**	
k	RMSE	MAE	RMSE	MAE	RMSE	MAE
1	0.5558	0.4292	0.4924	0.3776	0.4294	0.3335
4	1.1753	0.8904	1.0859	0.8116	0.7297	0.5600
8	1.2016	0.9123	1.2286	0.9108	0.7136	0.5469
12	1.2637	0.9605	1.3026	0.9659	0.7147	0.5481
16	1.2686	0.9672	1.3431	0.9884	0.7168	0.5486

with order $p = 48$, i.e., 24 h. An interesting alternative approach to this ARX model is utilized in Sec. 7.1. Knowing that the varying flow rate is caused by the varying heat demand of the consumers, one could directly predict the heat load and calculate back to the flow rate using the thermodynamic relationship (2.1) presented on page 8. Such a sophisticated heat-load model is proposed in Sec. 4.1.

The results for the averaged errors of the k-step-ahead prediction with $k = 1, \ldots, 16$ are summarized in Tab. 3.4 - 3.7. Prediction errors for the cARX models and for the cFIR models are calculated using both predicted flow rate and measured flow rate as input. For computing future flow rates the ARX model (3.63) with coefficients $a_1, \ldots a_p$ estimated by least-squares is used. Note that the parameters of the local models were reestimated at each time step using weighted recursive least-squares and the concept of variable forgetting factor (3.49) proposed on page 31 with $\lambda = 0.985$. Additionally, a regularization term of form (3.51) with $\alpha = 0.01$ was imposed.

Comparing the results, one can clearly observe that the prediction performance deteriorates with the length of the prediction horizon. Furthermore, as expected, the errors are smaller when measured volume flow rate values are used as input to the model. For $k = 1$ the error is the same, since in both cases measured flow rates are used. Also, due to the error accumulation caused by the autoregressive component, the performance of the cFIR is better than the performance of the cARX for longer prediction horizons.

It could be argued that, because only three months of validation data are used with only a few local models being in fact "active", the error differences between measured and predicted flow rate scenarios might be bigger. For this reason, the prediction errors for critical point 3 based on the entire year 2009 are presented in Tab. 3.8. Here the prediction performance for the case of measured, predicted and constant future flow rates as inputs are compared. A significant performance difference between the latter two can be observed. This further motivates the use of accurate models to calculate future volume flow rates.

Table 3.6: Prediction results of cFIR models with predicted flow rate measurements as input

	CP1 (AC3)		**CP2 (AC5)**		**CP3 (AC7)**	
k	RMSE	MAE	RMSE	MAE	RMSE	MAE
1	1.0381	0.7465	1.0109	0.6963	0.7799	0.5051
4	1.0869	0.8076	1.0804	0.7686	0.7931	0.5395
8	1.1019	0.8270	1.1266	0.8160	0.8075	0.5634
12	1.1104	0.8363	1.1366	0.8290	0.8229	0.5796
16	1.1090	0.8339	1.1359	0.8246	0.8358	0.5859

Table 3.7: Prediction results of cARX models with predicted flow rate measurements as input

	CP1 (AC3)		**CP2 (AC5)**		**CP3 (AC7)**	
k	RMSE	MAE	RMSE	MAE	RMSE	MAE
1	0.5558	0.4292	0.4924	0.3776	0.4294	0.3335
4	1.1815	0.8974	1.1029	0.8200	0.7346	0.5658
8	1.2259	0.9391	1.3333	0.9680	0.7548	0.5827
12	1.2968	0.9921	1.4598	1.0666	0.7898	0.6024
16	1.3014	0.9991	1.4890	1.1015	0.8093	0.6185

Table 3.8: Prediction results of cARX model for CP3 using observed/predicted/constant future flow rates as input

	cARX			**cFIR**		
k	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE
1	0.3805	0.3805	0.3805	0.7154	0.7154	0.7154
4	0.6628	0.6903	0.7151	0.7344	0.7518	0.7786
8	0.6713	0.7152	0.8132	0.7399	0.7632	0.8542
12	0.6835	0.7503	0.9336	0.7409	0.7736	0.9299
16	0.6926	0.7749	1.0055	0.7427	0.7798	0.9551

Chapter 4

Modeling of Heat Consumption

4.1 Modeling and Prediction of Heat-Load

As outlined in the introductory chapter to this work, due to the large operational costs involved, efficient operation of a district heating system is highly desirable. In particular, the stochastic nature of consumer heat demand has a major impact on the operation, as return temperature and flow rates in the network depend on consumers' space heating and tap-water heating demand [147]. Systems without demand forecasts require constant adjustment of the optimization computations inducing heavy computation load [167]. Hence, the incorporation of prediction methods to the optimization system is necessary for efficient operation and scheduling. Based on the forecasts for heat demand, future operational plans can be constructed as well as optimal resource allocation for energy production can be done by the optimization system.

In general, the heat load, which is excerted by the consumer installation on the network, depends on the following [147]

- meteorology (the weather)
- temperature demand (the behavior) of the occupants
- dynamics of the installation and, above all, of the building itself
- control instrumentation, its settings and programming.

Taking account of this a very reproducible demand pattern can be generated. In general there are many different types of consumer installations (apartments, hotels, school,...) connected to the network, so a separate forecast model has to be constructed for each individual installation. This problem could be tackled by, for instance, using clustering methods [32, 55, 31]. In these works, the electricity consumption patterns of customers were identified and customers exhibiting similar load diagrams were grouped by means of clustering. However, the main problem is the lack of sufficient measurement data for most consumers, thus most of the research is focused around building one global model for the total load at the plant [78, 82]:

$$Q_p(t) = Q_l + \sum_i Q_{c,i}(t + \tau_i(\dot{m})) \tag{4.1}$$

where Q_p stands for the heat produced at the plant, Q_l for the heat loss, $Q_{c,i}$ denotes heat consumption at consumer i and τ_i is the transport delay as a function of the mass flow rates $\dot{m}$ in the network.

In this work a seasonal autoregressive integrated moving average process with exogenous variables (SARIMAX) as well as a structural model in state-space form is presented for modeling

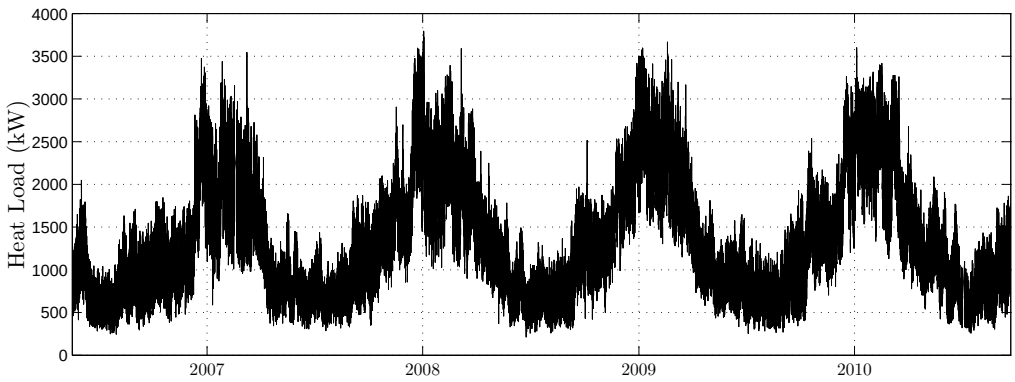


Figure 4.1: Multi-year record of system heat-load of DHN in Tannheim

the system heat-load of DHN Tannheim (Austria). The direct forecast of consumer load is not examined, due to the highly stochastic nature of the available consumer data. The exogenous influences of the ambient temperature on the heat-load are accounted for by a piece-wise linear function.

4.1.1 Heat-Load Characteristics

To give a better understanding of the characteristics of annual heat consumption, 4 years of heat-load data from the DHN Tannheim are shown in Fig. 4.1. One can clearly observe the typical yearly cycles with high values occurring in winter months and low values in summer months. As listed in [95] approximately 80% of the total heat demand can be attributed to ambient temperature. Other exogenous variables that have minor effect on the heat-load are cold-water temperature of the return line, solar radiation, wind speed and humidity [155, 52]. In general, it can be stated that the following four load components shape the total system heat-load [95]

- space heating for buildings
- domestic hot tap water preparation
- heat losses in the network
- additional work-day loads.

Space heating ensures thermal comfort in buildings by maintaining the indoor temperature at a desired level. It is naturally the predominant part of the total load in cold days. Domestic hot tap water preparation relates to satisfying hot-water demands. As pointed out in [95], hot-water demand shows large fluctuations and variations between different consumers and is therefore difficult to model. Heat losses result from the distribution network and are typically in the range of 5% to 20% of the total heat load [95, 85]. Additional work-day loads refer to the weekly pattern of the heat-load. Hence, beside the annual season, also a weekly periodicity as well as a daily of the heat load exist.

Real measurements of one week of heat demand for a cold and warm period are depicted in Fig. 4.1. In general it can be assumed that the shape of the heat demand differs between working days and weekends [52, 198, 46]. However, considering Fig. 4.1 it is difficult to say if a weekly pattern exists in that data snapshot. One possible reason for this is that Tannheim is a typical tourist centre. Hence, the winter tourism usually with many hotel guests on the weekend has a significant impact on the shape of the heat-load demand curve.

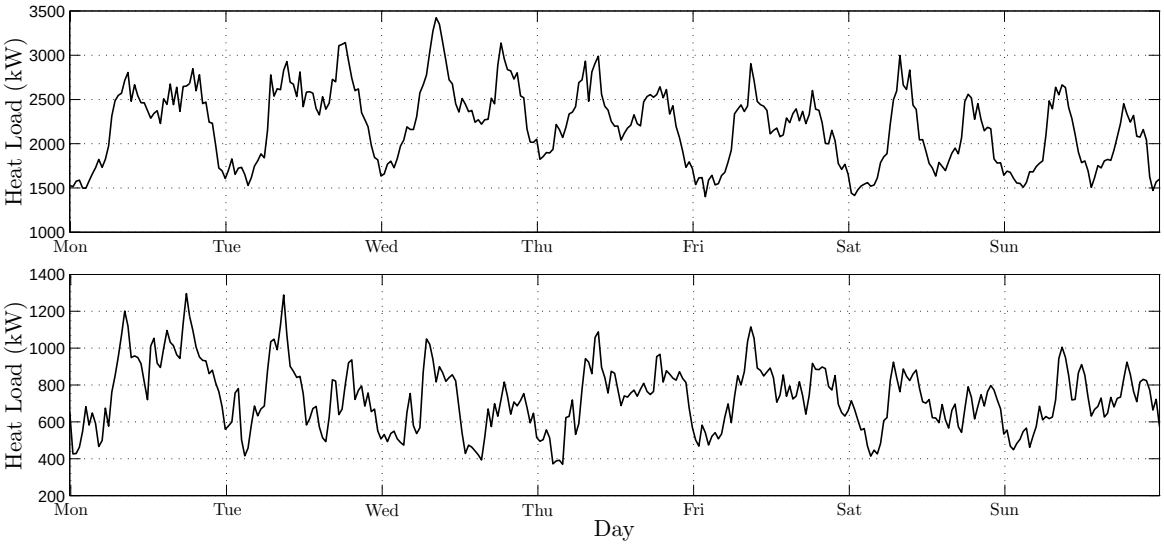


Figure 4.2: Weekly season of system heat-load of DHN in Tannheim for winter (top) and summer (bottom)

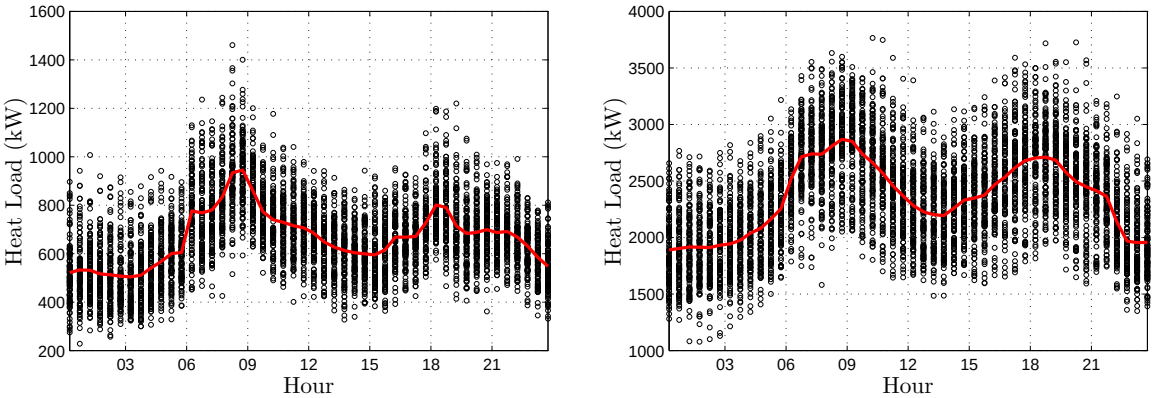


Figure 4.3: Daily heat-load profile for summer (left) and winter (right)

In addition, the intra-day periodicity is depicted in Fig. 4.3. Here clear differences between summer and winter days can be observed. Note that the red lines are constructed by building the temperature averages for all days where the mean daily temperature was higher than 15 °C for summer and lower than -2 °C for winter. For the years 2008 and 2009 this summed up to 107 and 110 days, respectively. Both heat profiles have in common that the demand abruptly increases between 6am and 8am, decreases afterwards slowly and increases again in the later afternoon with a peak at about 7pm. It seems that for summer days the two peaks are more narrow than for winter days and that the heat demand increase in the afternoon starts at 4pm, whereas for winter days it begins alread at 2pm.

To sum up, heat-load has in general a yearly, weekly, and daily pattern. This pattern is highly correlated with ambient temperature and the social behavior of the consumers. In conclusion, any forecast technique applied has to deal with this multi-periodicity.

4.1.2 Statistical Methods

SARIMA(X)

First some important concepts related to time series analysis that are used throughout this section are recalled. The following definitions are given from Chapter 2 and 3 of [27].

Definition 4.1.1 (Stationarity)**.** *The time series $\{Y_t, t \in \mathbb{Z}\}$, with index set $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$ is said to be stationary if, for $t \in \mathbb{Z}$, $\mathrm{E}[Y_t^2] < \infty$, $\mathrm{E}[Y_t] = m$ and, for all, s, $t \in \mathbb{Z}$, $\mathrm{E}[Y_t Y_s] = \mathrm{E}[Y_{t-s} Y_0]$.*

Remark. *Stationarity as just defined is also referred to as weak stationarity [26, 27]. For the remainder the term stationarity always refers to the properties specified by Definition 4.1.1.*

Definition 4.1.2 (ACF)**.** *Let $\{Y_t\}$ be a stationary time series. The autocorrelation function (ACF) ρ associated with $\{Y_t\}$ is defined, for all $t \in \mathbb{Z}$, as*

$$\rho(t) = \frac{\gamma(t)}{\gamma(0)} \tag{4.2}$$

where the autocovariance function $\gamma(t) = \mathrm{E}[Y_t Y_0] - \mathrm{E}[Y_t]\mathrm{E}[Y_0]$.

The autocorrelation function is extremely useful for assessing the degree of dependence in the data and to select a model for the data that reflects this. As it will be discussed later, the autocorrelation function together with the partial autocorrelation function (PACF) can be used to identify the appropriate order of the polynomials of an autoregressive moving average (ARMA) process. Note that the PACF at lag k may be regarded as the correlation between Y_1 and Y_{k+1} adjusted for the intervening observations $Y_2, \ldots, Y_k$ (see [27, p. 98] for details).

Definition 4.1.3 (ARMA(p,q) Process)**.** *Let $\{Y_t, t = 0, \pm 1, \pm 2, \ldots\}$ be a stationary time series with zero mean. It is said to be an ARMA(p,q) process if, for every $t \in \mathbb{Z}$,*

$$Y_t - \sum_{k=1}^{p} \phi_k Y_{t-k} = \epsilon_t + \sum_{k=1}^{q} \theta_k \epsilon_{t-k} \tag{4.3}$$

where $\{\epsilon_t\}$ is a white noise process with variance $\sigma^2 > 0$.

Equation 4.3 can be written symbolically in the more compact form

$$\phi_p(B) Y_t = \theta_q(B) \epsilon_t, \quad t = 0, \pm 1, \pm 2, \ldots \tag{4.4}$$

where ϕ and θ are the pth and qth degree polynomial

$$\phi_p(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p \tag{4.5}$$

and

$$\theta_q(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q, \tag{4.6}$$

and B is the backward shift operator defined by

$$B^j Y_t = Y_{t-j}, \quad j = 0, \pm 1, \pm 2, \ \ldots \ . \tag{4.7}$$

The polynomials $\phi(\,\cdot\,)$ and $\theta(\,\cdot\,)$ are referred to as the autoregressive and moving average polynomials, respectively. Note that if $\phi(\,\cdot\,)$ and $\theta(\,\cdot\,)$ have no common zeros and the characteristic equation (4.5) has all its roots outside the unit circle then the ARMA(p,q) process (4.3) is said to be causal. Similarly, the roots of $\theta(\,\cdot\,)$ in (4.6) must lie outside the unit circle if the process is to be invertible [27, 45, 23].

Remark. *It can be shown that a causal and invertible ARMA(p,q) has both the infinite moving average and infinite autoregressive representation. See [27, 23] for details.*

To model nonstationary time series with trends and seasonal patterns the seasonal autoregressive integrated moving average processes have been introduced [27, 26, 23]. In particular, $\{Y_t\}$ is said to follow a SARIMA$(p, d, q) \times (P, D, Q)_s$ process with period s if, for every $t \in \mathbb{Z}$, the differenced process $X_t := \nabla^d \nabla_s^D Y_t$ is a causal ARMA process,

$$\phi_p(B)\Phi_P(B^s)X_t = \theta_q(B)\Theta_Q(B^s)\epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{4.8}$$

according to the Definition 9.6.1 of [27], where the polynomials are defined, for all $z \in \mathbb{C}$, as

$$\begin{aligned} \phi_p(z) &= 1 - \sum_{k=1}^{p} \phi_k z^k, \qquad & \Phi_P(z) &= 1 - \sum_{k=1}^{P} \Phi_k z^k \\ \theta_q(z) &= 1 + \sum_{k=1}^{q} \theta_k z^k, \qquad & \Theta_Q(z) &= 1 + \sum_{k=1}^{Q} \Theta_k z^k \end{aligned} \tag{4.9}$$

and where

$$\nabla^d = (1 - B)^d, \quad \nabla_s^D = (1 - B^s)^D \tag{4.10}$$

denote the non-seasonal and seasonal differencing operator of order d and D, respectively. In order to make the time series stationary the non-seasonal differencing operator is used to eliminate polynomial trends and the seasonal operator is used to eliminate seasonal (periodic) patterns of length s. Note that the differenced process $\{\nabla^d \nabla_s^D Y_t\}$ is causal if and only if $\phi_p(z) \neq 0$ and $\Phi_P(z) \neq 0$ for $|z| \leq 1$ [27]. To estimate the coefficients of the AR- and MA-polynomials maximum likelihood can be applied [23].

As outlined in the beginning of this section the ambient temperature has a major influence on the heat-load. It is therefore reasonable to also include these exogenous effects in the model representation. For a SARIMA model with exogenous variables the term SARIMAX is generally used [17]. One way to construct SARIMAX models is by first filtering off the exogenous dependent part from the original time series and treat the residuals as SARIMA process [17, 34, 78].

Definition 4.1.4 (SARIMAX)**.** *A time series $\{Y_t\}$ is said to follow a SARIMAX$(p, d, q) \times (P, D, Q)_s$ process if, for every $t \in \mathbb{Z}$, it satisfies*

$$\begin{aligned} Y_t &= \boldsymbol{x}_t^T \boldsymbol{\beta} + V_t \\ \phi_p(B)\Phi_P(B^s)\nabla^d \nabla_s^D V_t &= \theta_q(B)\Theta_Q(B^s)\epsilon_t \end{aligned} \tag{4.11}$$

where $\boldsymbol{x}_t \in \mathbb{R}^m$ *is the vector of explanatory variables with corresponding parameter vector* $\boldsymbol{\beta} \in \mathbb{R}^m$.

Hence, fitting SARIMAX models can basically be separated into two stages (cp. [17]):

1. Filter off the exogenous influence from the original time series by linear regression using ordinary least squares
2. Identify and fit a SARIMA model for the filtered series (i.e., residuals of Step 1) using maximum likelihood estimation

For forecasting of the time series $\{Y_t\}$ both the contributions of the SARIMA model and the exogenous effects have to be considered. This also implies that predicted values of the exogenous variables are needed.

State-Space Model

Recall that a representation of a time series $\{Y_t\}$ in classical decomposition form can be written as [53, 90]

$$Y_t = m_t + s_t + \epsilon_t \quad t = 1, \ldots, n. \tag{4.12}$$

where m_t is a slowly changing function known as the trend component, s_t is a function with known period d referred to as the seasonal component, and ϵ_t is random noise or disturbance component.

As discussed in the first part of Sec. 4.1.2, the Box-Jenkins approach tries to eliminate these trend and seasonal components through differencing in order to obtain a stationary time series. The state-space approach, on the other hand, is based on a structural analysis of the problem [90]. Here the different components that make up the series, such as the aforementioned trend and seasonal terms, but also cycle, calender variations, explanatory variables or interventions, are modeled separately before being put together in the state-space model. Following the notation of [53] a linear Gaussian state-space model can be formulated as

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_t, & \boldsymbol{\epsilon}_t &\sim \mathcal{N}(0, \boldsymbol{H}_t), \\ \boldsymbol{\alpha}_{t+1} &= \boldsymbol{T}_t \boldsymbol{\alpha}_t + \boldsymbol{R}_t \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim \mathcal{N}(0, \boldsymbol{Q}_t), \qquad t = 1, \ldots, n, \\ & & \boldsymbol{\alpha}_1 &\sim \mathcal{N}(\boldsymbol{a}_1, \boldsymbol{P}_1), \end{aligned} \tag{4.13}$$

where $\boldsymbol{\alpha}_t$ is the $m \times 1$ unobserved state vector, $\boldsymbol{y}_t$ is the $p \times 1$ observation vector and the matrices $\boldsymbol{Z}_t, \boldsymbol{T}_t, \boldsymbol{R}_t, \boldsymbol{H}_t$ and $\boldsymbol{Q}_t$ are referred to as the state-space system matrices, which are initially assumed to be known. Furthermore, the error terms $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ are assumed to be serially independent and independent of each other at all time points. The initial state vector is denoted as $\boldsymbol{\alpha}_1$ and has mean $\boldsymbol{a}_1$ and variance matrix $\boldsymbol{P}_1$ and is also assumed to be independent of the error terms at all time points. The first equation of (4.13) is referred to as the observation equation and has the structure of a linear regression model. The second equation is called the transition equation and basically has the form of a first order vector autoregressive model.

There has also been much interest in the embedding of (S)AR(I)MA processes into the framework of state-space models for purposes of forecasting, as well as for specification and maximum likelihood estimation of parameters [23]. Classical state-space representations of ARMA models are presented in [26, 23, 105]. With minor adaptations SARIMA models can also be put into state-space form (see, e.g., [97, 78]). The basic idea is to construct an ARMA state-space model for the stationary differenced series $\nabla^d \nabla_s^D \boldsymbol{y}_t$ and augment the state vector by the nonstationary variables such as $\boldsymbol{y}_{t-i}$ and $(1-B)^d \boldsymbol{y}_{t-i}$.

Kalman Recursions. The Kalman filter produces the minimum mean squared linear estimator of the state vector $\boldsymbol{\alpha}_{t+1}$ based on the observations $Y_t = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_t\}$, that is, $\boldsymbol{a}_{t+1} = \mathrm{E}\left[\boldsymbol{\alpha}_{t+1}|Y_t\right]$ and the corresponding variance matrix of the estimator $\boldsymbol{a}_{t+1}$, denoted as $\boldsymbol{P}_{t+1} = \mathrm{Var}(\boldsymbol{\alpha}_{t+1}|Y_t)$ for $t = 1, \ldots, n$ [109]. The Kalman filter for the state-space model given by (4.13) can be formulated as [53, 4]

$$\begin{aligned} \boldsymbol{v}_t &= \boldsymbol{y}_t - \boldsymbol{Z}_t\boldsymbol{a}_t, & \boldsymbol{F}_t &= \boldsymbol{Z}_t\boldsymbol{P}_t\boldsymbol{Z}_t^T + \boldsymbol{H}_t, \\ \boldsymbol{K}_t &= \boldsymbol{T}_t\boldsymbol{P}_t\boldsymbol{Z}_t^T\boldsymbol{F}_t^{-1} & \boldsymbol{L}_t &= \boldsymbol{T}_t - \boldsymbol{K}_t\boldsymbol{Z}_t, & t &= 1, \ldots, n, \\ \boldsymbol{a}_{t+1} &= \boldsymbol{T}_t\boldsymbol{a}_t + \boldsymbol{K}_t\boldsymbol{v}_t, & \boldsymbol{P}_{t+1} &= \boldsymbol{T}_t\boldsymbol{P}_t\boldsymbol{L}_t^T + \boldsymbol{R}_t\boldsymbol{Q}_t\boldsymbol{R}_t^T \end{aligned} \tag{4.14}$$

for a given mean vector $\boldsymbol{a}_1$ and variance matrix $\boldsymbol{P}_1$ of the initial state vector $\boldsymbol{\alpha}_1$. The one-step-ahead prediction error (innovation) of the observation vector is $\boldsymbol{v}_t = \boldsymbol{y}_t - \mathrm{E}[\boldsymbol{y}_t|Y_{t-1}]$ with variance matrix $\boldsymbol{F}_t = \mathrm{Var}(\boldsymbol{y}_t|Y_{t-1})$. The matrix $\boldsymbol{K}_t$ is also called the Kalman gain matrix.

Estimation. The state-space model (4.13) is linear and driven by Gaussian disturbances and can therefore be treated by standard time series methods based on the Kalman filter [53, 90, 4]. Assuming that $\boldsymbol{a}_1$ and $\boldsymbol{P}_1$ are known the log-likelihood function is given by [97]

$$\begin{aligned} \log L = \log p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n; \boldsymbol{\psi}) &= \sum_{t=1}^{n} \log p(\boldsymbol{y}_t|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t-1}; \boldsymbol{\psi}) \\ &- \frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t_1}^{n}\left(\log|\boldsymbol{F}_t| + \boldsymbol{v}_t^T\boldsymbol{F}_t^{-1}\boldsymbol{v}_t\right), \end{aligned} \tag{4.15}$$

where $\boldsymbol{\psi}$ is the vector of (hyper)parameters under consideration and $\boldsymbol{v}_t$ are the innovations with variances $\boldsymbol{F}_t$. For a fixed vector $\boldsymbol{\psi}$ the innovations $\boldsymbol{v}_t$ and their variances $\boldsymbol{F}_t$ are computed via the Kalman filter (4.14). The unknown parameter vector $\boldsymbol{\psi}$ can be estimated numerically by maximizing (4.15) via some quasi-Newton optimization routine. For computational details it is referred to [53].

Diffuse Initialization. In most practical applications some elements of the mean vector $\boldsymbol{a}_1$ and variance matrix $\boldsymbol{P}_1$ of the initial state vector $\boldsymbol{\alpha}_1$ are not known. In such cases the state vector $\boldsymbol{\alpha}_1$ can generally be composed of a constant component, stationary component and nonstationary component as follows [53]

$$\boldsymbol{\alpha}_1 = \boldsymbol{a} + \boldsymbol{A}\boldsymbol{\delta} + \boldsymbol{R}_0\boldsymbol{\eta}_0, \quad \delta \sim \mathcal{N}(0, \kappa\boldsymbol{I}),\ \boldsymbol{\eta}_0 \sim \mathcal{N}(0, \boldsymbol{Q}_0), \tag{4.16}$$

where $\boldsymbol{a}$ is the known constant, $\boldsymbol{\delta}$ is the nonstationary component, $\boldsymbol{\eta}_0$ is a stationary component and $\kappa \to \infty$. The variance matrix $\boldsymbol{P}_1$ is then written as

$$\boldsymbol{P}_1 = \kappa\boldsymbol{P}_\infty + \boldsymbol{P}_* \tag{4.17}$$

with $\boldsymbol{P}_\infty = \boldsymbol{A}\boldsymbol{A}^T$ and $\boldsymbol{P}_* = \boldsymbol{R}_0\boldsymbol{Q}_0\boldsymbol{R}_0^T$. Note that a vector $\boldsymbol{\delta}$ with distribution $\mathcal{N}(0, \kappa\boldsymbol{I})$ as $\kappa \to \infty$ is said to be diffuse. For partially diffuse initial state vectors modifications to the classical Kalman filter recursions are required. The initial phase of the Kalman filter is then called diffuse initialization [47, 118, 53]. For instance, the SARIMA state-space model has a nonstationary component, therefore diffuse initialization techniques are necessary.

Forecasting. Each time a new observation is made the system is updated using the Kalman procedure. Multi-steps-ahead forecasts are then recursively obtained by applying the Kalman filter (4.14) with $\boldsymbol{K}_t = \boldsymbol{0}$ [53]. Hence, for l-step-ahead predictions one gets

$$\boldsymbol{a}_{t+k} = \boldsymbol{T}_{t+k-1}\boldsymbol{a}_{t+k-1}, \quad k = 1, \ldots, l \tag{4.18}$$

Table 4.1: Summary of the statistics of the training data set

Data	**Min**	**1. Qu.**	**Median**	**Mean**	**3. Qu.**	**Max**
Heat Load (kW)	336	946	1474	1606	2264	3631
Ambient Temperature (°C)	-20.785	-2.0852	3.168	3.9283	10.804	29.209

with variance matrix generated from

$$\boldsymbol{P}_{t+k} = \boldsymbol{T}_{t+k-1}\boldsymbol{P}_{t+k-1}\boldsymbol{T}_t^T + \boldsymbol{R}_{t+k-1}\boldsymbol{Q}_{t+k-1}\boldsymbol{R}_{t+k-1}^T, \qquad k = 1, \ldots, l. \tag{4.19}$$

These are equivalent to the classical state-space prediction formulas.

4.1.3 Modeling and Identification

For modeling purposes of the proposed forecasting methods the system heat load of Tannheim DHN was considered. The entire available data consist of measurements of the system heat load between 18.05.2006 and 22.09.2010 with a resolution of 15 minutes. The time series was resampled to 30 min, which is the time discretization used for the design of the predictive controller in Sec. 7.1.

Due to computational issues 6 months of data (1. Jan. - 30. Jun. 2009) were considered to be a sufficiently sized training data set. The second half of the year 2009 is used for validation. A statistical summary of the training data set is presented in Tab. 4.1. Of course, the estimated parameter values depend on the chosen training data. As discussed above, pronounced differences between the heat-load shape in the summer months owing to tap-water heating and winter months, where space-heating dominates, can be expected. To overcome this one might use periodic models, as suggested in [97], or construct some type of varying coefficient models (cf. Sec. 3.2).

First of all, obvious outliers (i.e., extreme values) in the data were identified and the corresponding values are treated as missing values. There is no further need for special treatment of missing values, since the state-space approach can handle missing observations in a straightforward manner (see [53, Sec. 2.7]). Next, the exogenous influence of the ambient temperature was modeled as piecewise linear function $f_p(T_{a,t})$ (cf. [52, 34]). As illustrated in Fig. 4.4 a function with five segments is used, but the number can, of course, be chosen arbitrarily. In general, the basis set for a piecewise-polynomial of order M with knots $\zeta_j,\ j = 1, \ldots, K$ would be [93]

$$\begin{aligned} h_j(X) &= X^{j-1}, \quad j = 1, \ldots, M, \\ h_{M+l}(X) &= (X - \zeta_l)_+^{M-1}, \quad l = 1, \ldots, K. \end{aligned} \tag{4.20}$$

For the SARIMAX model procedure these temperature effects were subtracted from the original time series, i.e., $V_t = y_t - f_p(T_{a,t})\ \forall t$. In the state-space modeling approach the basis functions of the piecewise fit were directly included as regressor variables.

All modeling and identification was done in MATLAB® using the SSM Toolbox [168]. This toolbox also provides procedures for incorporating SARIMA models into the state-space framework for the purpose of estimating (e.g., correct diffuse initialization) and forecasting.

SARIMAX

As illustrated in Fig. 4.5 the autocorrelation of the ambient temperature corrected residuals have local maxima at lag multiples of 24 and 48 which supports the hypothesis of seasonal autocorrelations. The latter indicates high correlation between observations 24 h apart, due to a sampling period of 30 min. This implies that differencing has to be applied, in order to make the time series stationary.

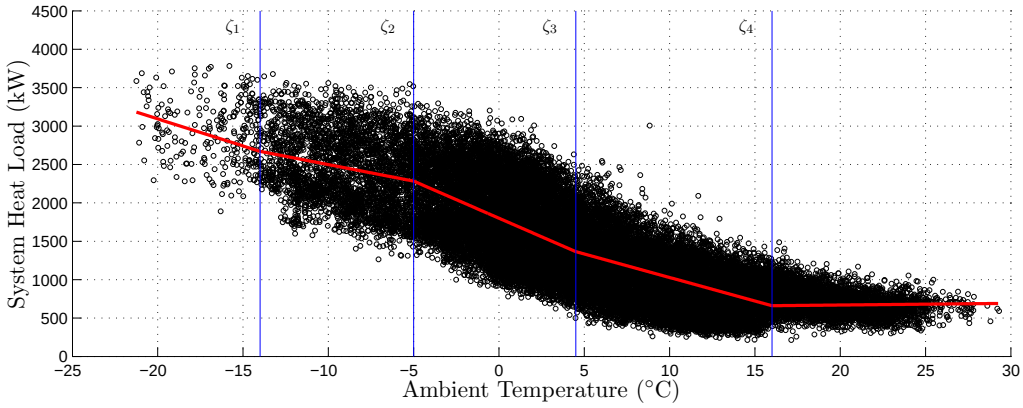


Figure 4.4: System heat load against ambient temperature plus piecewise linear function describing the dependency

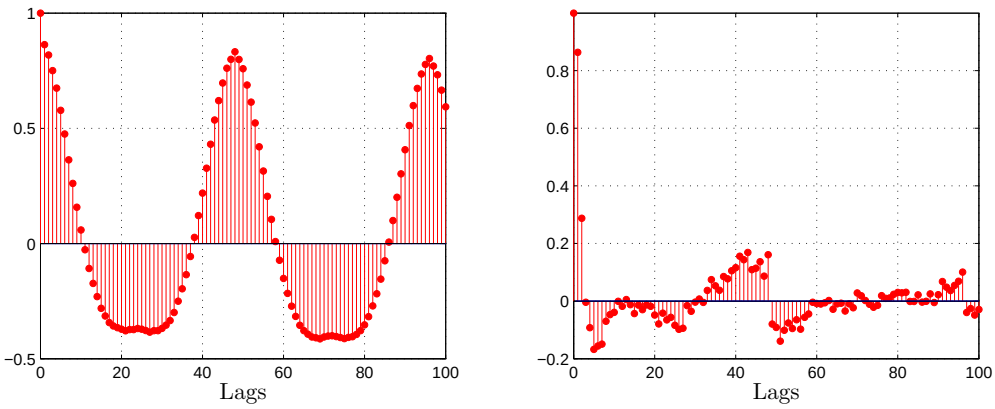


Figure 4.5: ACF (left) and PACF (right) of the residuals $\{\hat{V}_t\}$

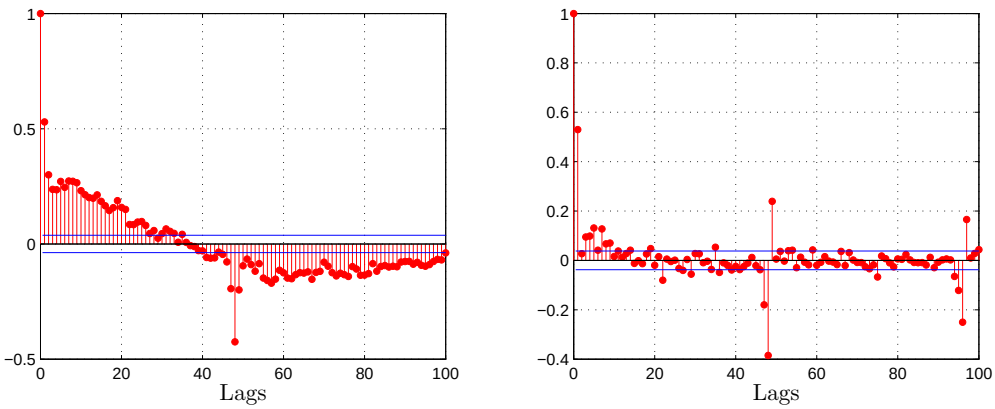


Figure 4.6: ACF (left) and PACF (right) of the seasonal differenced time series $\{\nabla_{48}\hat{V}_t\}$ with 95% confidence interval

Stationarity. To check for stationarity of the residuals $\{V_t\}$ the KPSS test [123] and the ADF test [180] were considered (cf. [17]). The KPSS test (command `kpsstest` in MATLAB®), suggests that, in the six months of training data, $\{\nabla\hat{V}_t\}$, $\{\nabla_{48}\hat{V}_t\}$ and $\{\nabla\nabla_{48}\hat{V}_t\}$ are stationary, whereas the ADF test (`adftest`) does not reject stationarity for all combination of differencing and even for the non-differences series. Note that the ADF test only checks for unit roots and might not be able to correctly identify seasonal unit roots. For further discussion on this see, e.g., [128]. Based on these test results it was difficult to determine the appropriate order of the differencing operator. However, considering Fig. 4.6 and Fig. 4.7 it was decided that the residuals $\{V_t\}$ may be represented by a SARIMA process with seasonal differences of order one ($D = 1$) and non-seasonal differences of order one or zero ($d = 0/1$).

Modeling. Theoretically, the sample ACF and PACF of the differenced series should help to determine the correct order of the seasonal and non-seasonal MA and AR part. According to Fig. 4.6 the sample ACF has the form of a damped sine wave with one spike at lag 48 and the sample PACF seem to have an exponential decay and several significant spikes at multiples of 48. Furthermore, the sample ACF of the doubled differenced series as illustrated in Fig. 4.7 does not show the damped sine wave anymore and the sample PACF indicates similar behaviour with spikes at lags at 48. Following the guideline of [23, Ch. 6] a set of candidate models were specified. For all of these competing models the Bayesian information criterion was calculated (BIC) (see [53, p. 152]:

$$BIC = n^{-1}\left[-2\log L_d(y|\hat{\boldsymbol{\psi}})\right] + (q+w)\log n \tag{4.21}$$

where $\log L_d(y|\hat{\boldsymbol{\psi}})$ is the value of the diffuse log-likelihood, q is the number of diffuse initial elements and w, the dimension of $\hat{\boldsymbol{\psi}}$, is the number of parameters to be estimated. The three models with the lowest scores were (in increasing order): SARIMA$(2,1,1)\times(0,1,1)_{48}$, SARIMA$(1,1,1)\times(0,1,1)_{48}$ and SARIMA$(2,0,0)\times(0,1,1)_{48}$. It was decided to take the best model in terms of BIC value for further analysis.

In particular the estimated coefficients of the SARIMA$(2,1,1)\times(0,1,1)_{48}$ model are as follows

$$(1+0.4142B-0.09673B^2)(1-B)(1-B^{48})V_t = (1-0.8813B)(1-0.9045B^{48})\epsilon_t, \tag{4.22}$$

with $\epsilon_t \sim \mathcal{N}(0, 1.163\cdot 10^4)$. The proposed model also satisfies the causality and invertibility condition. (i.e. all roots of AR and MA characteristic polynomials lie outside the unit circle).

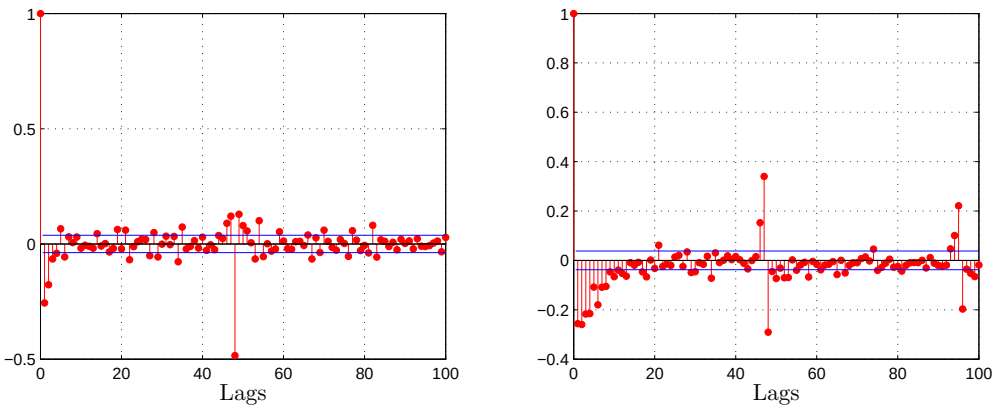


Figure 4.7: ACF (left) and PACF (right) of the double differenced time series $\{\nabla\nabla_{48}\hat{V}_t\}$ with 95% confidence interval

Diagnostic checking. After having specified a tentative model the (standardized) residuals should be inspected for outliers and checked for normality, heteroscedasticity and serial correlation. Some suspiciously large std. residuals could be observed. The corresponding observations were treated with dummy variables (see [40]) and the chosen SARIMA model was reestimated. To investigate normality of the residuals the Jarque-Bera Test [100] (command `jbtest` in MATLAB®) can be used. The test rejects the null hypothesis that the distribution is normal ($p = 0.001$). The reason might be that the residuals have a skewness of 0.12 (i.e., skewed to the right) and kurtosis of 3.8 (i.e, distribution is more outlier-prone than the normal distribution), which can also be confirmed be inspecting the histogram and the QQ-Plot in Fig. 4.8. Heteroskedasticity of the residuals was checked with the test statistic proposed in [53, p. 34] (see also [117]), and for serial correlation the Box-Ljung statistic [140] (`lbqtest`) was applied. The former does not not reject the null hypothesis of homoscedasticity (0.0568 at a 5% significance level). On the other hand, the null hypothesis of no autocorrelation is clearly rejected ($p \approx 0$). Although in Fig. 4.8 the absolute value of the autocorrelation is always smaller than 0.05. Thus, it seems, at least visually, that the model has captured the essential dependencies in the series. Note that the two other SARIMA models produced quantitatively the same test results.

State-Space Model

In the following, a univariate state-space model for the heat load as well as the temperature is constructed. The state-space approach also allows easy extension to the multivariate case as demonstrated in [117, 53, 40]. For instance, in [50] the dependent hourly univariate electricity load is represented as a daily multivariate state-space model with 24 regression equations, i.e., one regression equation for each hour of the day. Furthermore, to reduce the number of unknown parameters a factor model approach was pursued.

For modeling the heat-load, different state-space variants were compared (i.e., different seasonal terms, stochastic vs. deterministic regressors; cf. [117]). Based on the lowest BIC score the following model was chosen:

$$Q_t = \mu_t + \gamma_t + \underbrace{\beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + \beta_4 x_{4,t} + \beta_5 x_{5,t}}_{\text{Temperature effects}} + \underbrace{\beta_6 x_{6,t}}_{\text{Weekend Dummy}} + \epsilon_t \quad \forall t \tag{4.23}$$

Here a weekend dummy was included as it is assumed that heat-load differs between weekdays and weekends. Of course, one might also include dummies for holidays or calender effects such as Christmas and Easter. Furthermore, a trend component μ_t as well as a trigonometric seasonal

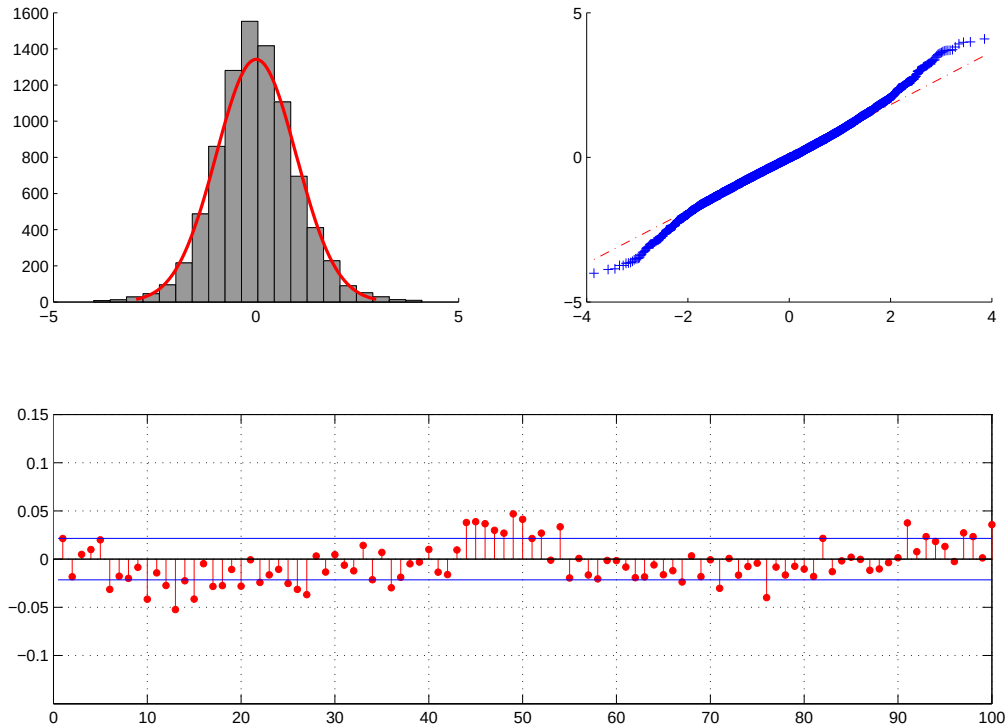


Figure 4.8: Diagnostic plots of standardized prediction error for SARIMA model: (top left) histogram plus density of normal distribution; (top right) QQ-Plot; (bottom) correlogram

Table 4.2: Maximum Likelihood estimation results for the state-space heat load model

Component	**Variance**
$\mu_{t+1} = \mu_t + \zeta_t$	$\zeta_t \sim \mathcal{N}(0, 1443)$
$\gamma_{t+1} = \gamma_t + \omega_t$	$\omega_t \sim \mathcal{N}(0, 0.1087)$
$\beta_{1,t+1} = \beta_{1,t} + \xi_{1,t}$	$\xi_{1,t} \sim \mathcal{N}(0, 0.4137)$
$\beta_{2,t+1} = \beta_{2,t} + \xi_{2,t}$	$\xi_{2,t} \sim \mathcal{N}(0, 1.86 \cdot 10^{-4})$

component γ_t with a period of 48 samples were included. The nonlinear ambient temperature effects are accounted for by the piecewise linear function as depicted in Fig. 4.4. The regressors $x_{1,\ldots,5,t}$ stand for the corresponding basis functions. The maximum likelihood estimate for the variance of the irregular component is $\hat{\sigma}_\epsilon^2 = 8098$ and the other estimates are summarized in Tab. 4.2. Note that in particular the seasonal component and some of the temperature effects are formulated in a stochastic way, i.e., they are allowed to change over time. The state vector for this specific model has dimension 54 and is composed as follows

$$\alpha_t = \begin{pmatrix} \mu_t & \gamma_t & \gamma_{t-1} & \cdots & \gamma_{t-46} & x_{1,t} & x_{2,t} & \cdots & x_{6,t} \end{pmatrix}^T . \tag{4.24}$$

Hence, both the trigonometric offset and the coefficient of the exogenous variable are placed into the state vector.

Diagnostic checking. The same test statistics as for the SARIMA model were applied. The test for homoscedasticity is satisfactory ($p = 0.80$), whereas independence is not satisfied ($p = 0.017$) and also the assumption of normality ($p = 0$) is clearly violated. These test results can also be verified in the diagnostic plots Fig. 4.9. As one can see there is highly significant short term correlation in the prediction errors. To overcome this one might include lagged variables of the output as additional explanatory variables to (4.23).

Temperature Model. As discussed before the temperature has a major influence on the heat-load. Consequently, the accuracy of the prediction will strongly be correlated with the quality of the estimates of the future ambient temperature. In real applications exact weather forecasts can be obtained from meteorological services. In this work a very simple structural model for the temperature series is proposed, which is basically composed of a trend component and a trigonometric seasonal component:

$$\begin{aligned} T_{a,t} &= \mu_t + \gamma_t + \epsilon_t, & \epsilon_t &\sim \mathcal{N}(0, \sigma_\epsilon^2), \\ \mu_{t+1} &= \mu_t \alpha_t + \zeta_t, & \zeta_t &\sim \mathcal{N}(0, \sigma_\zeta^2), \quad t = 1, \ldots, n. \end{aligned} \tag{4.25}$$

This type of model is also called *local level model* [53]. The maximum likelihood estimates for variance terms are $\hat{\sigma}_\epsilon^2 = 1.331 \cdot 10^{-7}$ and $\hat{\sigma}_\zeta^2 = 0.2121$.

4.1.4 Validation

To assess the performance of the proposed models different prediction accuracy measures for observations between 1. Nov. 2009 - 1. Feb. 2010 were computed. More precisely, the RMSE, MAE and the MAPE (mean absolute percentage error [46]) for rolling k-steps-ahead prediction ($k = 1, \ldots, 16$) were considered. The 16-step-ahead prediction equals 8 h due to half hourly measurements. This was also chosen as prediction horizon for the predictive controller in Sec. 7.1.

The predictive performance for the temperature model are summarized in Tab. 4.3. The average mean absolute error for 16-step-ahead prediction is only about 2.7 °C, despite the very

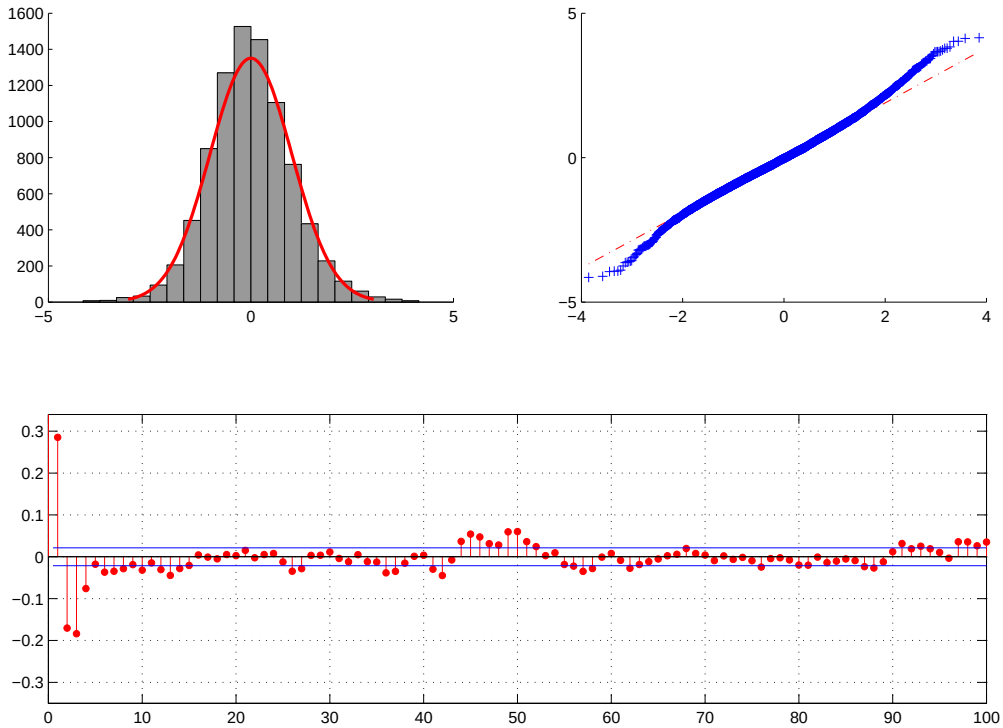


Figure 4.9: Diagnostic plots of standardized prediction error for state-space model: (top left) histogram plus density of normal distribution; (top right) QQ-Plot; (bottom) correlogram

Table 4.3: Prediction performance for the state-space temperature model; prediction accuracy measures are in °C

k	RMSE	MAE	$\hat{\sigma}_\epsilon$
1	0.4433	0.3255	0.4480
4	1.4617	1.1148	1.4630
8	2.3662	1.8531	2.3671
12	2.9546	2.3600	2.9553
16	3.3249	2.6777	3.3254

Table 4.4: Prediction performance comparisons of heat-load with observed ambient temperature; prediction accuracy measures are in kW

	SARIMAX			**State-Space**		
k	RMSE	MAE	MAPE	RMSE	MAE	MAPE
1	96.3910	75.1324	7.7289	112.3568	87.9796	8.8025
4	138.3906	106.7152	10.8636	134.6611	104.1640	10.3422
8	153.1478	118.6062	12.0378	145.4309	111.6266	10.8882
12	162.4168	126.5970	12.9925	150.1913	116.2737	11.4213
16	167.1573	131.0025	13.5311	148.3943	114.3652	11.3284

simple forecast model. At this point a comparison with weather forecasts obtained by professional services would be interesting.

Results for heat-load predictions are presented in Tab. 4.4 and Tab. 4.5. In the former table it was assumed that future ambient temperature values were known and in the latter predicted values were used instead. As one can see there is a significant performance difference between simulations using the observed ambient temperature and predicted ambient temperature. This underlines the fact that accurate future temperature estimates are very important. Furthermore, one can see that the SARIMAX model compared to the state-space models yields better results in terms of the one-step ahead prediction error. Which can be explained by the autocorrelation component of the SARIMAX model. For longer prediction horizons, however, the state-space model clearly outperforms the SARIMAX model. To demonstrate the actual modeling performance 16-step-ahead estimates for the case of higher heat demand is shown in Fig. 4.10 and for the case of lower heat demand in Fig. 4.11, respectively. In-sample, in contrast to out-of-sample, means that at each time step a new observation comes in and the Kalman filter can be applied.

Table 4.5: Prediction performance comparisons of heat-load with predicted ambient temperature; prediction accuracy measures are in kW

	SARIMAX			**State-Space**		
k	RMSE	MAE	MAPE	MSE	MAE	MAPE
1	97.0447	75.3374	7.6981	110.2977	85.9392	8.7821
4	150.9600	116.3761	11.3744	140.8308	108.1882	10.7629
8	191.5365	147.6779	14.0467	167.8475	126.6016	12.1214
12	220.9623	170.5754	16.2791	186.8227	141.5765	13.4487
16	239.9274	184.1929	17.4938	194.8436	147.5003	13.8549

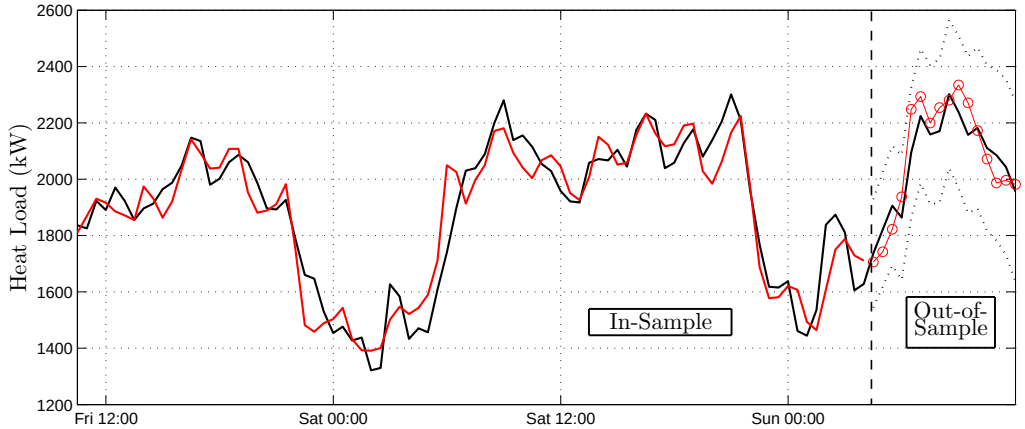


Figure 4.10: 16-step-ahead forecasts of system heat-load during winter, including their 90% confidence interval; the full black line are the heat-load observations

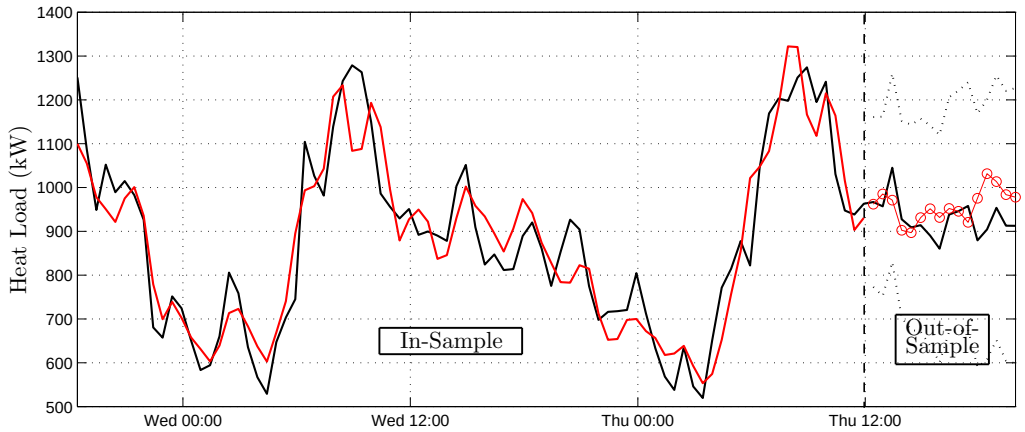


Figure 4.11: 16-step-ahead forecasts of system heat-load during summer, including their 90% confidence interval; the full black line are the heat-load observations

As discussed above, the daily heat-load has in general different pattern during summer where tap water heating dominates as opposed to cold winter days where space heating dominates. Nonparametric or semiparametric modeling frameworks such as conditional ARX models (cf. Sec. 3.2) where coefficients vary with the ambient temperature might be an appealing alternative to global parametric models. This was also motivated in [155] where a grey box model for heat consumption was proposed and the statistical part was described by a varying coefficient model. In the next section a sophisticated nonparametric regression approach is briefly introduced which is then used for modeling the primary return temperature at consumer stations.

4.2 Modeling of Return Temperature

Recall that consumer stations can be represented by heat exchangers which take the necessary heat from the primary supply side by cooling the primary return temperature. Knowledge of the primary return temperature is needed for calculating the volume flow rates in the network which are required for the determining the corresponding transport time delays. Often these return temperatures vary only smoothly and be can assumed to be constant. A physically motivated approach, modeling the consumer as heat exchangers, was mentioned on in the introduction chapter (see page 8). For this approach details about the heat exchanger such as total surface area and heat-transfer coefficient as well as the secondary side supply and return temperature must be known. Although classical heat exchanger theory is not valid for district heating applications as concluded in [179]. It was argued that the return temperature mainly depends on the ambient temperature and the social load. For this reason and because no technical data sheets of the installed heat exchangers and no information about the secondary side setup was available, it was decided to describe the return temperature at the aggregated consumer by the following nonlinear mapping:

$$T_r^{\mathrm{AC}} = f(T_s^{\mathrm{AC}}, \underbrace{T_a, c}_{Q_{\mathrm{AC}}}) \tag{4.26}$$

where c is the trigonometric offset representing the social component. As indicated here, heat-load was modeled in a similar way by utilizing the ambient temperature and a trigonometric seasonal term. From (4.26) the flow rates at the aggregated consumer can now be calculated using the thermodynamic relationship

$$q_{\mathrm{AC}} = \frac{Q_{\mathrm{AC}}}{\boldsymbol{c}_p \rho (T_s^{\mathrm{AC}} - T_r^{\mathrm{AC}})}. \tag{4.27}$$

With these flow rates the flow distribution in the entire network can be determined using the techniques discussed in Sec. 3.1.1.

In Fig. 4.12 the scatterplot of supply temperature and ambient temperature is depicted. There is a break point at an ambient temperature of approximately 1 °C for aggregated consumer 4 and at 18 °C for aggregated consumer 7 where the return temperature has its lowest value. For higher ambient temperatures, the return temperature increases again. As discussed in [179] this can be explained by the fact that for cold days when space-heating dominates the return temperature increases with the ambient temperature. For warm days, however, the total load in the system is small, leading to very low flow rates in the system. To overcome this and keep the flow up, supply water might be by-passed to the return line.

4.2.1 Multivariate Adaptive Regression Splines (MARS)

A sophisticated method for nonlinear regression problems is Multivariate Adaptive Regression Splines technique (MARS) developed by [63]. MARS was, for instance, also successfully applied

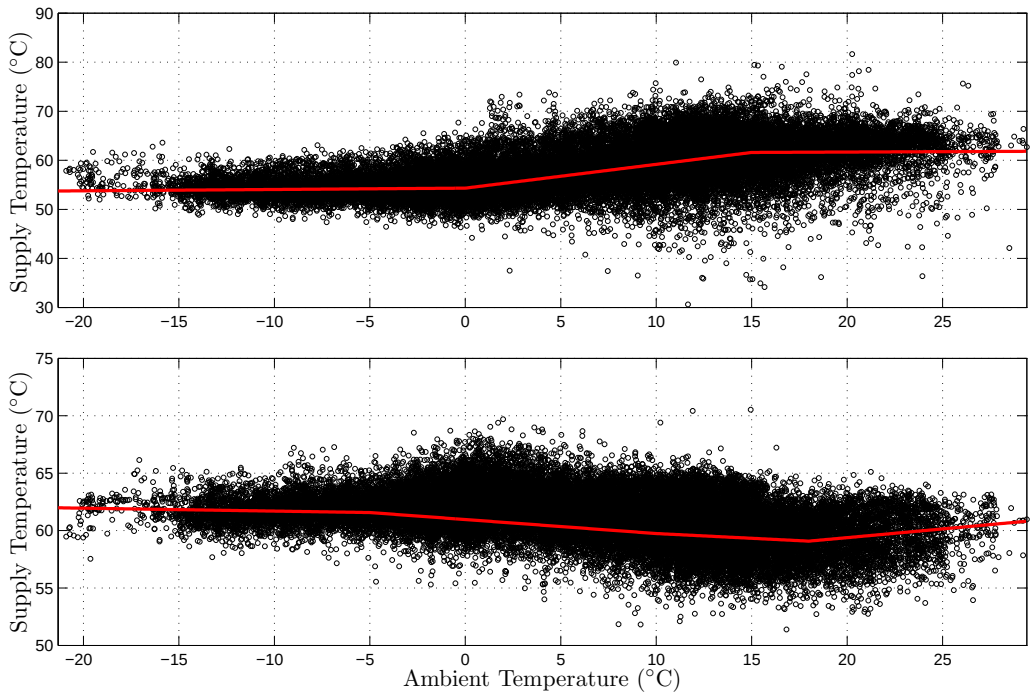


Figure 4.12: Supply temperature against ambient temperature for AC4 (top) and AC7 (bottom) plus piecewise linear fit

for modeling of nonlinear time series phenomena [136, 135]. MARS fits a model in the form of an expansion in product spline basis functions of explanatory variables. The basis functions are chosen during a forward and backward recursive partitioning strategy. More precisely, MARS uses as basis functions two-sided truncated functions of order q of the form $(x-t)_+^q$ and $(t-x)_+^q$ [63]. The idea is to create such pairs for each input variable $\boldsymbol{x}_j$ with knots at each observed value x_{ij} of that input. This gives the following collection of basis functions

$$\mathcal{C} = \left\{(\boldsymbol{x}_j - t)_+^q, (t - \boldsymbol{x}_j)_+^q\right\}, \; t \in \{x_{ij}, \ldots, x_{Nj}\}, \; j = 1, \ldots, p \tag{4.28}$$

The MARS model for an output variable y, and M terms, can be presented in the following equation [93]

$$y = \beta_0 + \sum_{m=1}^{M} \beta_m h_m(\boldsymbol{x}) \tag{4.29}$$

where each $h_m(\boldsymbol{x})$ is a function in $\mathcal{C}$, or a product of two or more such functions. For a given h_m the coefficients β_m in (4.29) can be estimated by standard linear regression. The challenging part, however, is the construction of the functions $h_m(\boldsymbol{x})$. Here a forward and backward recursive partitioning strategy is applied. Basically, the algorithm consists of the following steps:

1. Start with the simplest model involving only the constant function $h_0(\boldsymbol{x}) = 1$
2. Search the set of all candidate functions $\mathcal{C}$, and add those which produce the largest decrease in training error
3. Apply step 2 recursively until a model of pre-determined maximum complexity is derived
4. In the last stage, a pruning procedure is applied where those basis functions are removed that contribute least to the overall goodness of fit.

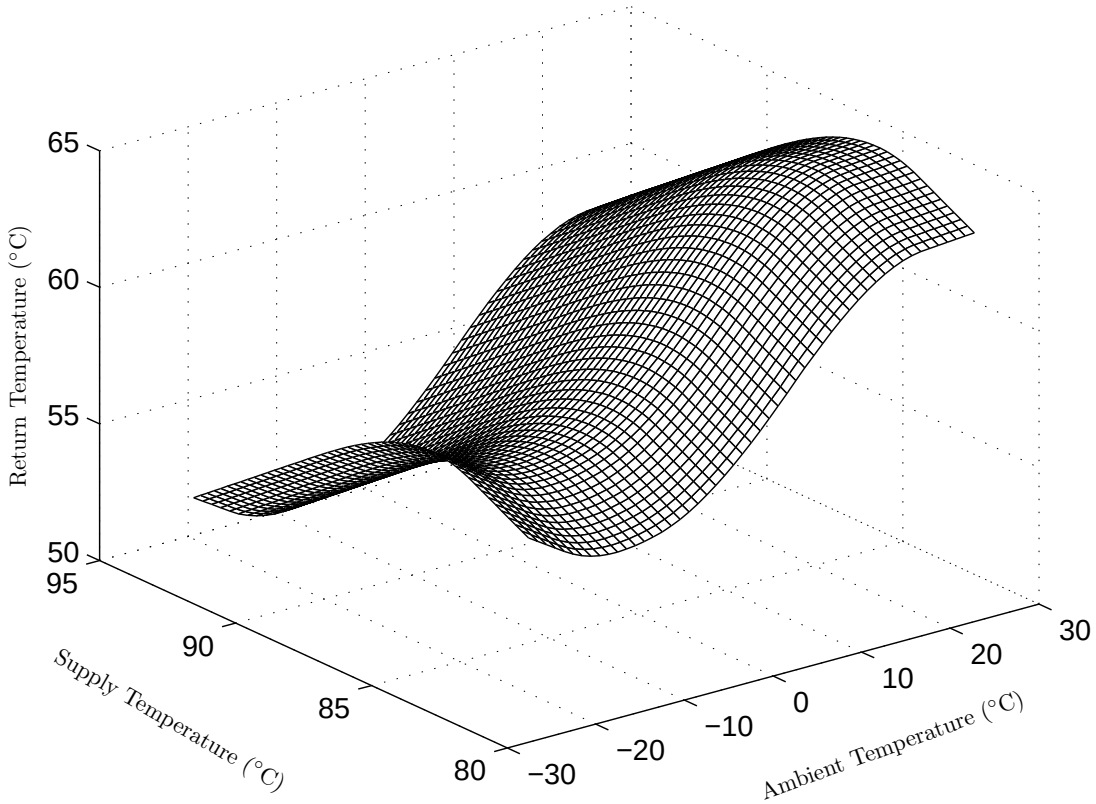


Figure 4.13: Regression surface for the return temperature estimated by MARS for AC4

Note that as measure of the goodness of fit the generalized cross validation-error (cf. Eq. (5.38)) is used taking into account both the residual error and the model complexity. For more algorithmic details the interested reader is referred to [63, 93].

4.2.2 Estimation Results

For estimation purposes the aggregated time series of (primary) supply temperature, return temperature and the ambient temperature at all seven aggregated consumers are considered. The time frame of the data is the year 2009. The nonlinear mapping (4.26) was estimated by MARS using the MATLAB® toolbox `ARESLab` [101]. For the configuration parameters the default values were used (i.e., maximal basis interaction is one, piecewise cubic models, maximal number of basis function in forward phase is 21). The results for aggregated consumer 4 and aggregated consumer 7 are shown in Fig. 4.13 and Fig. 4.14, respectively. The aforementioned effects of the ambient temperature are clearly observable. There is also a recognizable relationship between supply temperature and return temperature. Considering Fig. 4.13 it seems that higher supply temperature implies lower return temperature. This fact is also confirmed by classical heat exchanger theory.

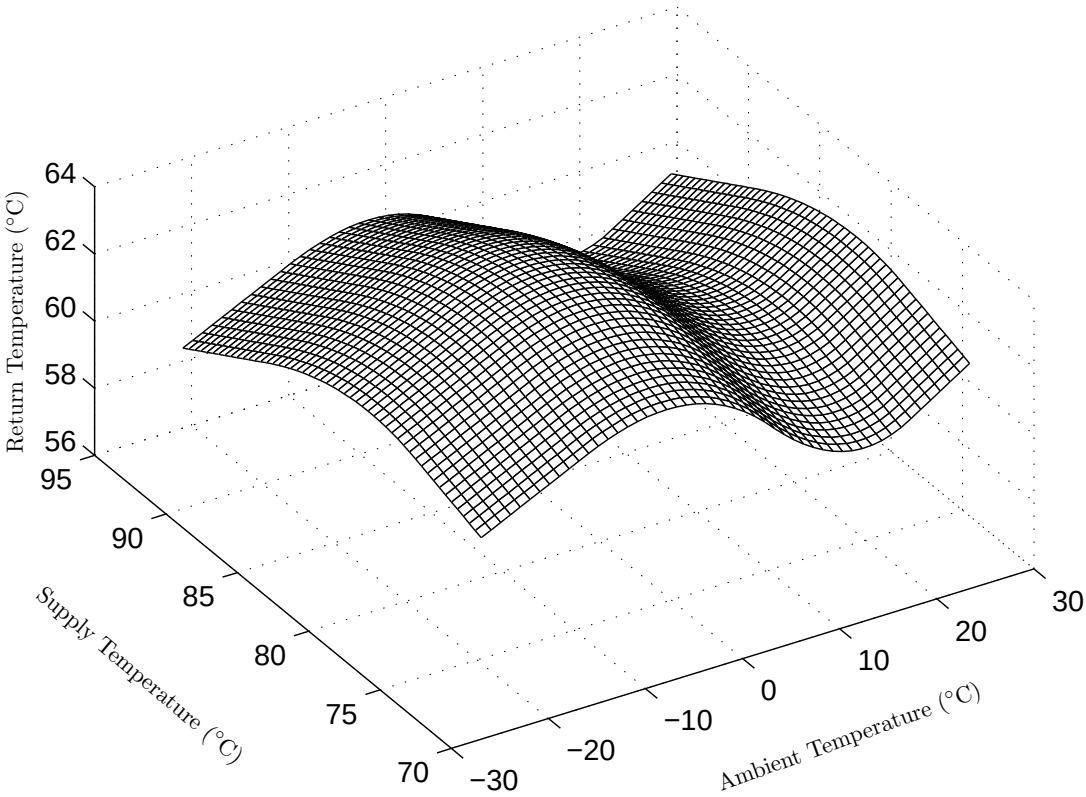


Figure 4.14: Regression surface for the return temperature estimated by MARS for AC7

Chapter 5

Modeling of the Biomass Combustion Plant

5.1 First-Principles Model

5.1.1 Introduction

Due to the worldwide concern about global warming due to the emission of CO_2 and other greenhouse gases and the limited availability of fossil fuels the importance of renewables, such as biomass, in energy production has increased [218, 217]. Grate firing is a state-of-the art technique that is currently used in biomass combustion for heat and power production [217, 163, 119].

In the following a simple grey-box model for a biomass plant is proposed which is largely based on mass and energy balances. Only for the relationship describing the rate of combustion was expert knowledge used. First, a brief process description is given.

5.1.2 Process Description of a Moving Grate Biomass Furnace

Modern grate-fired boilers consist generally of four key elements: a fuel feeding system, a grate assembly, an intelligent air supply system (primary and secondary air) and the freeboard which enables a complete combustion system [163, 161]. For the heat production a steam generator with either natural or forced circulation is chosen. Optionally, a steam turbine can be added for power production. A schematic sketch of a grate-fired boiler is given in Figure 5.1. Biomass is fed into the furnace from the left side and moved onto the moving grate by hydraulically operated pushers. Then the fuel is transported on the moving grate from left to right with grate speed v_{g} and the residual ashes are removed from the rightmost side.

The following input variables were used for the grey-box model which were identified by expert knowledge:

1. Biomass input feed ($\dot{m}_{\mathrm{in}}$ in kg/s)
2. Mass flow rate of the primary air supply ($\dot{m}_{\mathrm{PA}}$ in kg/s)
3. Mass flow rate of the secondary air supply, ($\dot{m}_{\mathrm{SA}}$ in kg/s)
4. Mass flow rate of recycle gas ($\dot{m}_{\mathrm{rec}}$ in kg/s).

It is important to note that the speed of the moving grate is assumed to be proportional to the biomass input feed. According to experts, this is a reasonable assumption and works well in practice.

The thermal power Q_{th} and the flue gas temperature T_{f} are the main operating parameters in biomass furnaces [72]. Monitoring and the ability to simulate the oxygen concentration O_2 of the

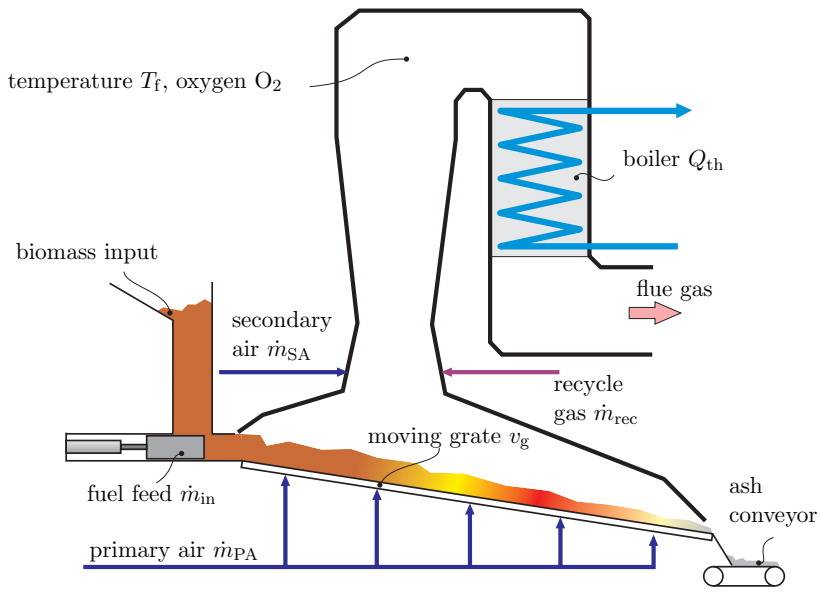


Figure 5.1: Schematic drawing of the moving grate biomass furnace. The fuel is fed to the left side of the grate and is transported to the ash conveyor on the far right side (reproduced with modifications from [164])

flue gas is also considered to be essential, since it has a significant impact on the efficiency of the combustion [119]. For instance, incomplete combustion can lead to high emissions of unburned pollutants such as Carbon monoxide [160]. For safe and secure operation it is also important to determine the position of the highest rate of combustion C_{pos} which should be approximately at the center of the grate [164]. Hence, the relevant output variables of the process are defined as:

1. Combustion position (C_{pos} in %)
2. Heat output (thermal power) (Q_{th} in kW)
3. Oxygen concentration (O_2 in %)
4. Flue gas temperature (T_f in °C).

Relevant disturbances of the process are the moisture content of biomass fuel and the varying grate speed. The latter is a result of the varying heat demand which requires adjustment of the grate speed. Furthermore, the agglomeration and break-up of the fuel particles on the fuel bed occurs in a highly stochastic manner. Due to these characteristics the global behavior of a moving grate biomass furnace within its operational envelope is nonlinear and has to be modeled adequately. For this purpose a simple grey-box model is proposed in the following as well as an alternative data-driven modeling approach in Sec. 5.2.

5.1.3 First-Principles Modeling

Combustion position. The combustion position C_{pos} can be calculated from the fuel mass on the grate m_g and the grate velocity v_g:

$$\tau_{pos}\frac{dC_{pos}}{dt} = k_{pos}v_g m_g - C_{pos}, \tag{5.1}$$

where k_{pos} is some constant and τ_{pos} is a given time constant. Due the assumption of a fixed relationship between input feed and grate speed, the latter is determined by the biomass input

feed $\dot{m}_{\mathrm{in}}$ as follows

$$v_{\mathrm{g}} = k_{\mathrm{in}} \dot{m}_{\mathrm{in}} \tag{5.2}$$

where k_{in} is the proportional factor between fuel input and the grate velocity. Furthermore, the fuel mass on the grate is given by the following mass balance

$$\frac{dm_{\mathrm{g}}}{dt} = \dot{m}_{\mathrm{in}} - \dot{m}_{\mathrm{c}} \tag{5.3}$$

with $\dot{m}_{\mathrm{c}}$ denoting the mass reduction due to drying and combustion. The mass reduction is the amount of a biomass material that undergoes drying or combustion over a period of time and is determined empirically by

$$\dot{m}_{\mathrm{c}} = k_{\mathrm{c}} \left(v_{\mathrm{g}} + \bar{v}_g \right) \dot{m}_{\mathrm{PA}} m_{\mathrm{g}} \tag{5.4}$$

were $\dot{m}_{\mathrm{PA}}$ is the primary air mass flow, k_{c} is an appropriate factor, and $\bar{v}_g$ is an offset term with regard to the grate velocity. The latter guarantees that the combustion rate is not zero in case of a stop of the moving grate.

Thermal Power. The thermal output Q_{th} of the plant is computed from the mass reduction and the heat value of the biomass input k_{th}

$$\tau_{\mathrm{th}} \frac{dQ_{\mathrm{th}}}{dt} = k_{\mathrm{th}} \dot{m}_{\mathrm{c}} - Q_{\mathrm{th}}. \tag{5.5}$$

The time constant τ_{th} is given by

$$\tau_{\mathrm{th}} = \frac{m_{\mathrm{b}} c_{p,\mathrm{b}}}{c_{p,\mathrm{f}} \left(\dot{m}_{\mathrm{PA}} + \dot{m}_{\mathrm{SA}} + \dot{m}_{\mathrm{rec}} \right)} \tag{5.6}$$

where m_{b} denotes the mass of the biomass boiler, $\dot{m}_{\mathrm{SA}}$ is the mass flow rate of secondary air supply, $\dot{m}_{\mathrm{rec}}$ is the mass flow rate of recycle gas, $c_{p,\mathrm{b}}$ and $c_{p,\mathrm{f}}$ are the specific heat capacity of the boiler and flue gas, respectively.

Oxygen Concentration. A common stoichiometric approach can be applied to determine the O_2 concentration in the flue gas [110, 163]:

$$\tau_{\mathrm{O_2}} \frac{d\mathrm{O_2}}{dt} = 21 \frac{(\lambda - 1)}{\lambda} - \mathrm{O_2} \tag{5.7}$$

where $\tau_{\mathrm{O_2}}$ is some time constant and λ is the stoichiometric air to fuel ratio. The latter represents the quotient between real air supply and the minimum required air supply for complete combustion and can be written as follows [110]

$$\lambda = \frac{(\dot{m}_{\mathrm{PA}} + \dot{m}_{\mathrm{SA}})}{k_{L_{\mathrm{min}}} \dot{m}_{\mathrm{c}}}. \tag{5.8}$$

where $k_{L_{\mathrm{min}}}$ is some factor related to the minimum required air supply for complete combustion.

Flue Gas Temperature. Finally, the flue gas temperature is computed from (cf. [72])

$$\tau_{\mathrm{th}} \frac{dT_{\mathrm{f}}}{dt} = \frac{Q_{\mathrm{th}}}{\dot{m}_{\mathrm{f}} c_{p,\mathrm{f}}} + T_{\mathrm{FA}} - T_{\mathrm{f}} \tag{5.9}$$

where T_{FA} denotes the temperature of fresh air, $\dot{m}_{\mathrm{f}}$ is flue gas mass flow and τ_{th} is computed from (5.6). The flue gas mass flow is given by

$$\dot{m}_{\mathrm{f}} = \dot{m}_{\mathrm{c}} + \dot{m}_{\mathrm{PA}} + \dot{m}_{\mathrm{SA}} + \dot{m}_{\mathrm{rec}} \tag{5.10}$$

where $\dot{m}_{\mathrm{rec}}$ is the mass flow rate of recycle gas.

Table 5.1: Nominal parameter values for the moving grate biomass furnace model

Variable	Value	Variable	Value
$c_{p,\mathrm{b}}$	2.5 kJ/kgK	k_{in}	5/3600 m/kg
$c_{p,\mathrm{f}}$	1.3 kJ/kgK	k_{pos}	1800 s/mkg
m_{b}	200 kg	k_{c}	3.6 s/mkg
T_{FA}	15 °C	k_{th}	10 MJ/kg
$\bar{v}_{\mathrm{g}}$	2.5/3600 m/s	$k_{L_{\min}}$	4.54
$\tau_{\mathrm{O_2}}$	60 s	τ_{pos}	120 s

5.1.4 Model Verification

To verify the proposed first-principle model real measurements from the biomass plant in Grossarl (Salzburg) are utilized. In Grossarl two biomass boilers with a heat power of 3.5 MW and 1.5 MW are installed. Additionally, for peak heat demand and as a reserve boiler a conventional oil-fired boiler with 5.7 MW can be operated. Here the data from the 3.5 MW boiler is considered for validation purposes. Unfortunately, no reliable measurements of the combustion position have been made. The attempt to reconstruct data using measurements from temperature sensors located along the grate was not successful. Further note that the input data of primary air, secondary air and recycle gas were given in percent of the total ventilator speed and it was therefore necessary to use the corresponding characteristic curves of the ventilators to get the individual mass flow rates.

The parameter values for the model are summarized in Tab. 5.1. These values were partly determined by expert knowledge and by fitting the model to the real data using the Levenberg-Marquardt optimization technique. The comparison of simulated output with real process measurements in Fig. 5.2 shows good agreement. In particular, the simulated thermal power as well as oxygen concentration indicates minimal error. However, for flue gas temperature the dynamic pattern cannot be adequately reproduced. Nevertheless, the results are acceptable and allow to use this validated grey-box model as "plant model" in the predictive control setup in Sec. 7.2.

5.2 Data-Driven Model

5.2.1 Introduction to Fuzzy Models

As human knowledge and expertise often comes in terms of verbal rules, it is attractive to integrate such linguistic information into the modeling process by using a fuzzy system [1, 196, 153, 7]. The basic element of a fuzzy system is a set of *fuzzy inference rules* which is also called the *knowledge base*. In general, each inference rule consists of two elements: the IF-part, called the antecedent of a rule, and the THEN-part, called the consequent of the rule. The structure of a single rule can be presented as follows:

$$\text{IF an antecedent proposition THEN a consequent proposition.} \tag{5.11}$$

The antecedent defines the condition, and the consequent the conclusion which will only be implemented if the condition is true. The antecedent of a fuzzy rule can include logical operators such as conjunction (and), disjunction (or) and negation (not) [196, 153]. For instance, it can take the following form:

$$\text{IF } z_1 \text{ is } A_1 \text{ and } z_2 \text{ is } (\text{not} A_2) \cdots \text{ or } z_m \text{ is } A_m \text{ THEN a consequent proposition.} \tag{5.12}$$

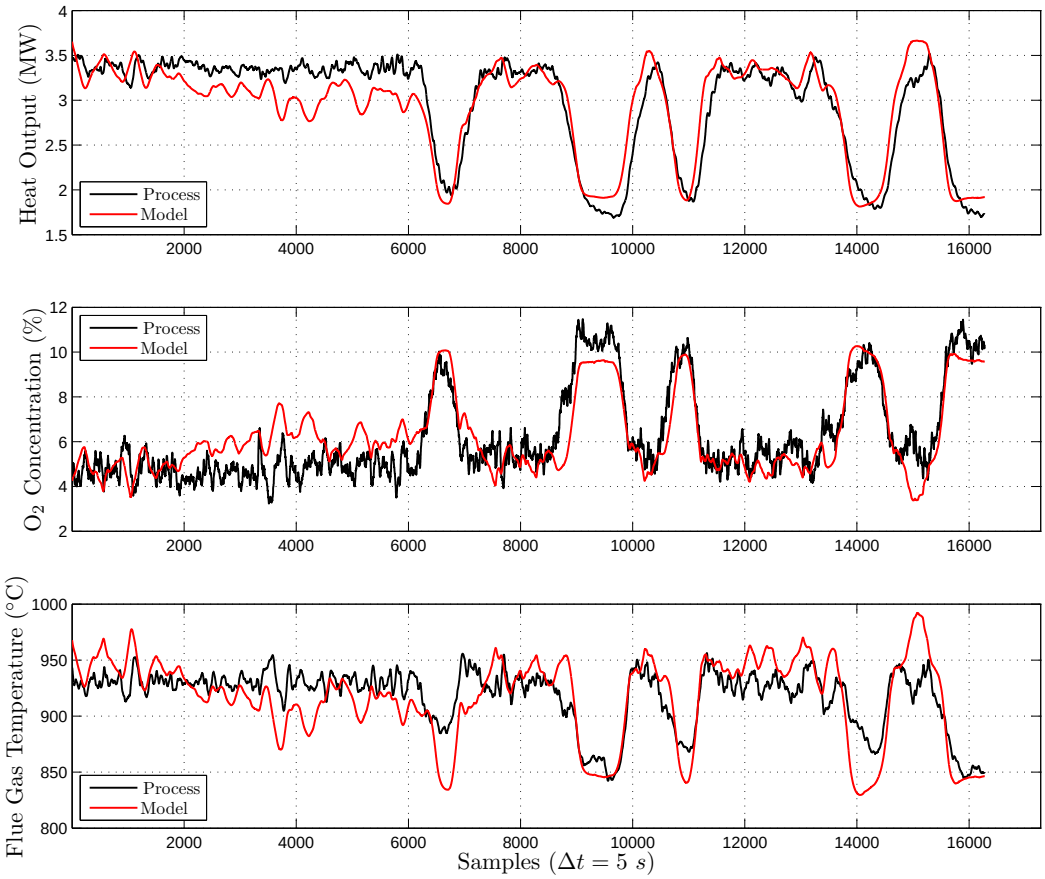


Figure 5.2: Comparison of simulated output of the first-principle plant model with process measurement data for 15.8.2012

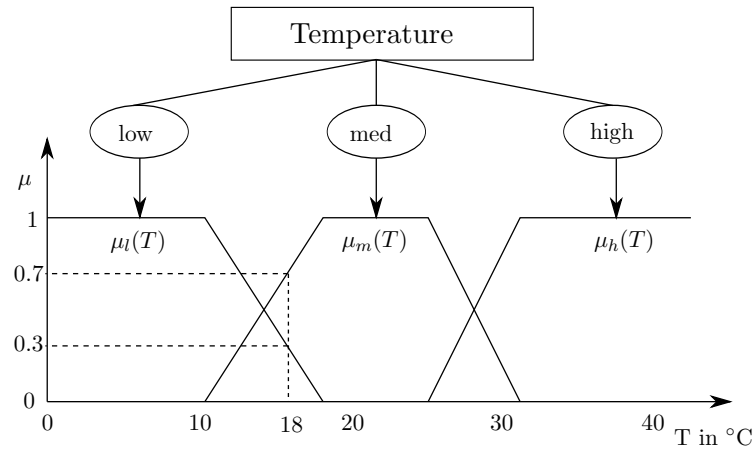


Figure 5.3: Illustration of a linguistic variable "temperature" with tree linguistic terms: low, medium, and high

Here $z_1, \ldots, z_m$ are the linguistic variables or fuzzy variables being input to the fuzzy system, and $A_1, \ldots, A_m$ are fuzzy sets described by *membership functions* (MSFs) $\mu_{A_j}(z_j) :\rightarrow [0,1]$, $j = 1, \ldots, m$. The consequent proposition can generally take one the following forms [196]

1. Crisp consequent:
$$\text{IF an antecedent proposition THEN } y = y_a \tag{5.13}$$
where y_a is a numerical value or a symbolic value,

2. Fuzzy consequent:
$$\text{IF an antecedent proposition THEN } y \text{ is } Y_k \tag{5.14}$$
where y is a fuzzy variable (linguistic) and Y_k is a fuzzy set,

3. Functional consequent:
$$\text{IF} z_1 \text{ is } A_1 \text{ and } z_2 \text{ is } A_2 \cdots \text{ and } z_m \text{ is } A_m \text{ THEN } y = f(z_1, z_2, \ldots, z_m) \tag{5.15}$$
where f is a certain function with input variables $(z_1, z_2, \ldots, z_m)$.

Fuzzy systems which have rules with functional consequents are also called Takagi-Sugeno fuzzy systems or simply TS fuzzy systems [194]. This type of fuzzy model allows effective modeling of nonlinear relations using a small number of rules and is described in greater detail in the next subsection.

For illustrative purposes consider Fig. 5.3 (cf. [7, 153, 196]). Here *temperature* is the linguistic input variable quantified by the linguistic terms *low, med* and *high*. These linguistic terms are defined by their associated membership functions $\mu_l(T)$, $\mu_m(T)$ and $\mu_h(T)$ which have trapezoidal shape in this example. Triangle or Gaussian membership function shapes are also popular [1]. These MSFs define the degree of membership for a specific temperature value to the fuzzy sets. For instance, in case of $T = 18\,°\text{C}$ the degree of memberships are: $\mu_l(18) = 0.7$, $\mu_m(18) = 0.3$ and $\mu_h(18) = 0$, respectively. This means that $18\,°\text{C}$ is considered low with weight of 0.7, medium with 0.3 and high with 0. The procedure which calculates from a crisp input the degree of membership for the fuzzy sets is called *fuzzification* [153]. Further suppose that there are more linguistic input statements, e.g., temperature and wind speed with given membership functions. The degree of membership of each input variable can be combined by appropriate operators. The result is then called the degree of rule fulfillment or firing strength [153].

5.2.2 Takagi-Sugeno (TS) Fuzzy Models

Structure of TS Fuzzy Models

Complex dynamical system can often be represented by an nonlinear ARX (NARX) model structure [1]. In general, this structure can be considered as a nonlinear relation between past inputs and outputs and the predicted output of the system:

$$\hat{y}(k+1) = f\left(y(k), \ldots, y(k-n_y+1), u(k-n_d), \ldots, u(k-n_u-n_d+1)\right), \tag{5.16}$$

where n_y and n_u are the maximum lags considered for the output and input terms, respectively, n_d is the discrete dead time, and f represents the nonlinear mapping. TS fuzzy models are proved to be suitable for approximation of such systems by interpolating between local (affine) linear, time-invariant (LTI) ARX models [150, 1]. For each rule R^j the following structure holds:

$$\begin{aligned} &\mathrm{R}^j : \text{IF } z_1 \text{ is } A_1^j \text{ and } \cdots \text{ and } z_m \text{ is } A_m^j \\ &\quad \text{THEN } y^j(k+1) = \sum_{i=1}^{n_y} a_i^j y(k-i+1) + \sum_{i=1}^{n_u} b_i^j u(k-i-n_d+1) + c^j. \end{aligned} \tag{5.17}$$

Here $\boldsymbol{z} = [z_1, \ldots, z_m]$ is the vector of input fuzzy variables and $A_1^j, \ldots, A_m^j$ are the antecedent fuzzy sets or regions for the jth rule R^j with corresponding membership functions $\mu_{A_1}^j, \ldots, \mu_{A_m}^j$. The elements of the fuzzy vector are usually a subset of the past inputs and outputs [1]:

$$\boldsymbol{z} \in \{y(k), \ldots, y(k-n_y+1), u(k-n_d), \ldots, u(k-n_u-n_d+1)\}. \tag{5.18}$$

In the special case that the elements of $\boldsymbol{z}$ are only external variables the term parameter scheduling is used which is closely related to gain scheduling [153]. Note that only single-input, single-output (SISO) dynamical systems are considered here. The extension to multiple input, single output (MISO) or multiple input, multiple output (MIMO) TS fuzzy models is straightforward (see, e.g., [1, p. 102 f]).

The antecedent propositions in TS fuzzy models are in conjunctive form. Hence, the degree of fulfillment of the jth rule can be computed using the product operator [7, 153]:

$$\mu^j(\boldsymbol{z}) = \prod_{i=1}^{m} \mu_{A_i}^j(z_i). \tag{5.19}$$

Furthermore, by denoting the normalized degree of fulfillment with

$$w^j(\boldsymbol{z}) = \frac{\mu^j(\boldsymbol{z})}{\sum_{i=1}^{r} \mu^i(\boldsymbol{z})} \tag{5.20}$$

the overall output of the fuzzy model can be written as

$$y(k+1) = \sum_{j=1}^{r} w^j(\boldsymbol{z}) y^j(k+1) \tag{5.21}$$

where r denotes the number of rules. Hence, the model domain will be divided into r fuzzy sub-domains each corresponding to a local process model. Further if, all consequents of the rules have identical structure, the TS model can be expressed as a pseudo-linear model with input-dependent parameters [7]

$$y(k+1) = \sum_{i=1}^{n_y} a_i(\boldsymbol{z}) y(k-i+1) + \sum_{i=1}^{n_u} b_i(\boldsymbol{z}) u(k-i-n_d+1) + c(\boldsymbol{z}) \tag{5.22}$$

where

$$a_i(\boldsymbol{z}) = \sum_{j=1}^{r} w^i(\boldsymbol{z}) a_i^j, \quad b_i(\boldsymbol{z}) = \sum_{i=1}^{r} w^i(\boldsymbol{z}) b_i^j \quad c(\boldsymbol{z}) = \sum_{i=1}^{r} w^i(\boldsymbol{z}) c^j. \tag{5.23}$$

It is worth mentioning that for some applications the interpolation properties of the TS model can be improved by replacing the weighted mean interpolation (5.21) by another function. For instance, using the max-function $s(z) = \max(0, z),\ z \in \mathbb{R}$, as proposed in [7, p. 36], the interpolation formula for two rules becomes

$$y(k+1) = y^1(k+1) + s\left[y^2(k+1) - y^1(k+1)\right]. \tag{5.24}$$

In [220] it was further suggested to replace the max-function by a smooth piece-wise polynomial to eliminate the drawback of non-differentiability at $z = 0$.

It is obvious that there are systematic similarities between conditional parametric models introduced in Sec. 3.2 and TS fuzzy models. In the context of conditional parametric models the local neighborhood around each fitting point is determined by Kernel functions where as for fuzzy models the term membership functions is used . In both approaches the product operator can be used to calculate the overall model output. Also both modeling frameworks can basically be reduced to the concept of parameter-varying models.

Optimization

The optimization of TS fuzzy models comprises several different aspects. Since fuzzy models are closely related to conditional parametric models some of these points were already discussed in Sec. 3.2. Generally spoken, the components which can be optimized are the following (for details see [153, pp. 93 ff.]):

- Consequent Parameters
- Antecedent Parameters
- Rule Structure

For fuzzy models of TS type the consequent proposition is a linear parametrized function. Therefore, the consequent parameters can be optimized by standard least-squares techniques. Formal descriptions of different least-squares approaches are pursued below. Antecedent parameters are related to the input membership function characteristics such as partitions and widths. They can either be chosen by expert knowledge or be optimized by nonlinear optimization techniques. However due to the high computational burden of nonlinear optimization techniques it is common to optimize only the rule consequent parameters. The optimal complexity of the fuzzy model is determined by optimization of the fuzzy rule structure. This is a combinatorial problem and nonlinear global search methods are necessary. An alternative to direct rule selection are heuristic construction algorithms such as the LOLIMOT algorithm (see Sec. 3.2.3).

Least-Squares Optimization. The identification of the consequent parameters in TS fuzzy models (5.17) can be formulated as a classical linear regression problem under the assumption that the membership functions are known. In general, two different approaches can be distinguished: *global* and *local* estimation [153, 1, 151, 7]. In the global approach one compact least-squares problem is solved in which the consequent parameters of all rules are estimated simultaneously. Whereas in the local estimation problem the parameters for each rule are estimated separately resulting in r (weighted) least-squares problems.

In both methods, first, the training data set $\{y(k), u(k), z(k)\},\ k = 1, \ldots, N$ has to be arranged into a design matrix $\boldsymbol{X}$, process output vector $\boldsymbol{y}$ and weight matrix $\boldsymbol{W}^j$ as follows

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}(1)^T \\ \boldsymbol{x}(2)^T \\ \vdots \\ \boldsymbol{x}(N)^T \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix}, \quad \boldsymbol{W}^j = \begin{bmatrix} w_1^j & 0 & \cdots & 0 \\ 0 & w_2^j & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_N^j \end{bmatrix}, \tag{5.25}$$

with the regression vector

$$\boldsymbol{x}(k) = [y(k-1), \ldots, y(k-n_y), u(k-n_d-1), \ldots, u(k-n_d-n_u), 1]^T \tag{5.26}$$

and where w_k^j denotes the normalized degree of fulfillment of the jth rule which is calculated from (5.20) with the use of $\boldsymbol{z}(k) \in \boldsymbol{x}(k)$. Next, the consequent parameter vector of the jth rule is concatenated into a single parameter vector $\boldsymbol{\theta}^j$:

$$\boldsymbol{\theta}^j = \left[a_1^j, \ldots, a_{n_y}^j, b_1^j, \ldots, b_{n_u}^j, c^j\right]. \tag{5.27}$$

Finally, the rule consequent parameters $\boldsymbol{\theta}^j,\ j = 1, \ldots, r$ for given process data $\boldsymbol{X}$,$\boldsymbol{y}$ and corresponding weight matrices $\boldsymbol{W}^j, j = 1, \ldots, r$ have to be estimated.

Global Least-Squares Method. In the global approach the consequent parameter of all r rules are estimated with one least-squares problem [7]:

$$\min_{\theta} J = \frac{1}{N}\left(\boldsymbol{y} - \tilde{\boldsymbol{X}}\boldsymbol{\theta}\right)^T \left(\boldsymbol{y} - \tilde{\boldsymbol{X}}\boldsymbol{\theta}\right) \tag{5.28}$$

where the matrix $\tilde{\boldsymbol{X}}$ is composed as follows:

$$\tilde{\boldsymbol{X}} = \left[\boldsymbol{W}^1\boldsymbol{X}, \boldsymbol{W}^2\boldsymbol{X}, \ldots, \boldsymbol{W}^r\boldsymbol{X}\right] \tag{5.29}$$

and the parameter vector $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\theta} = \left[(\boldsymbol{\theta}^1)^T, (\boldsymbol{\theta}^2)^T, \ldots, (\boldsymbol{\theta}^r)^T\right]^T. \tag{5.30}$$

When the matrix $\tilde{\boldsymbol{X}}$ has full rank, there is a unique solution to (5.28) which can be written as [108]

$$\hat{\boldsymbol{\theta}} = \left(\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{X}}\right)^{-1}\tilde{\boldsymbol{X}}^T\boldsymbol{y}. \tag{5.31}$$

Although the global least-squares yields a minimal prediction error [7], its computational load increases cubically with the number of rules [153, 1]. Thus, it is not suited for fuzzy systems with many rules.

Local Least-Squares Method. In the local estimation approach the parameters of the rules are estimated separately. This means that the interactions between the different rules, i.e., the overlaps between the membership functions (cf. Fig. 5.3) are neglected. Formally, the local approach can be written as r weighted least-squares problems [1]:

$$\min_{\theta^j} J = \frac{1}{N}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}^j\right)^T \boldsymbol{W}^j \left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}^j\right), \quad j = 1, \ldots, r, \tag{5.32}$$

where the corresponding weighted least-squares estimates are computed by [108]

$$\hat{\boldsymbol{\theta}}^j = \left(\boldsymbol{X}^T\boldsymbol{W}^j\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}^j\boldsymbol{y}. \tag{5.33}$$

Compared to global least-squares, the computational burden of the local least-squares approach grows only linearly with the number of rules [153]. However, this comes at the cost of introducing a systematic error due to the neglected interactions between the different rules.

As shown in [151] the local estimation implements an inherent regularization effect. This means it reduces the degrees of freedom of the model by decreasing the number of *effective parameters* of the model. This results on the one hand in an increased bias error and on the other hand in a reduced variance error as predicted by the famous bias-variance dilemma [93]. Hence, local estimation may improve the overall performance, if the additional bias error is compensated or even over-compensated by a reduction in the variance error. Other benefits of local learning are the possibility to locally interpret the estimated parameters. This is very important in the process industry, where one wants to gain insight into the physical process behavior locally at particular operating points. Another advantage is regarding adaptive learning which was introduced in Sec. 3.2.2. If changes in the estimates are only anticipated for specific rules, it makes no sense, from a computational point of view, to apply the global least-squares approach, i.e, estimate the consequent parameter of all rules. For a thorough discussion of global versus local learning see [151] and [153, Sec. 13.2]. Interestingly, it was also proposed in the literature to combine the global (5.28) and local optimization problem (5.32) to generate models with good tradeoff in terms of global fitting and local interpretation. For further details on this subject the reader is referred to [216] and [1].

Note that the global model output can be written as follows (see Eq. (5.21) and (5.33))

$$\hat{\boldsymbol{y}} = \sum_{j=1}^{r} \boldsymbol{W}^j \boldsymbol{X} \hat{\boldsymbol{\theta}}^j = \boldsymbol{S}\boldsymbol{y} \tag{5.34}$$

where $\boldsymbol{S}$ is defined as

$$\boldsymbol{S} = \sum_{j=1}^{r} \boldsymbol{W}^j \boldsymbol{X} \left(\boldsymbol{X}^T \boldsymbol{W}^j \boldsymbol{X} \right)^{-1} \boldsymbol{X}^T \boldsymbol{W}^j. \tag{5.35}$$

The matrix $\boldsymbol{S}$ is also called the smoothing matrix [151]. The effective number of parameters (also known as the effective degrees-of-freedom) is now calculated from [151, 93, 38]

$$n_{\text{eff}} = \text{tr}(\boldsymbol{S}) \tag{5.36}$$

where $\text{tr}(\boldsymbol{S})$ denotes the trace of the matrix $\boldsymbol{S}$, that is, the sum of its diagonal elements. The smoothing matrix can be used to calculate the leave-one-out cross-validation (LOOCV) score [209, 93]

$$LOOCV = \frac{1}{N} \sum_{k=1}^{N} \left[\frac{y(k) - \hat{y}(k)}{1 - S_{kk}} \right]^2 \tag{5.37}$$

where S_{kk} is the kth element of $\boldsymbol{S}$ or the leave-one-out generalized cross-validation score

$$LOOGCV = \frac{1}{N} \sum_{k=1}^{N} \left[\frac{y(k) - \hat{y}(k)}{1 - \text{tr}(\boldsymbol{S})/N} \right]^2 . \tag{5.38}$$

For instance, the LOOCV was used by the author to determine the optimal width of the MSFs in [79].

Output-Error Optimization Until now, it was assumed that the dynamical process is represented by a nonlinear ARX model structure and can be approximated locally by linear, time-invariant ARX models. In such a setting the parameters can be estimated with least-squares,

minimizing the one-step ahead prediction error, if the process output and relevant input variables are measured during operation. This setting is also called series-parallel model in the identification literature [153].

In contrast, if the focus lies on simulation rather than prediction, the so called parallel model configuration minimizing the output-error might be preferred. Simulation here simply means that on the basis of previous process inputs, only the model simulates future outputs. Due to the feedback of the process output a nonlinear output-error model structure (NOE) is obtained:

$$\hat{y}(k+1) = f\left(\hat{y}(k), \ldots, \hat{y}(k-n_y+1), u(k-n_d), \ldots, u(k-n_u-n_d+1)\right), \tag{5.39}$$

Similar to the NARX approach, (5.39) can be approximated by a TS fuzzy model using output-error (OE) models as consequent propositions. Output-error models are nonlinear in their parameters and consequently nonlinear optimization techniques, such as the Levenberg-Marquart algorithm, have to be utilized [153]. As with local ARX models, the training of the local OE models can be carried out globally or locally [153]. In the following only the global approach is discussed. Consider the following nonlinear least-squares problem

$$\min_{\theta} J = \frac{1}{N}\left(\boldsymbol{y} - \hat{\boldsymbol{y}}(\theta)\right)^T \left(\boldsymbol{y} - \hat{\boldsymbol{y}}(\theta)\right) \tag{5.40}$$

where $\hat{\boldsymbol{y}}(\boldsymbol{\theta}) = [\hat{y}(1,\boldsymbol{\theta}), \ldots, \hat{y}(N,\boldsymbol{\theta})]^T$ is the global output of the fuzzy model with

$$\begin{aligned}\hat{y}(k,\boldsymbol{\theta}) = f_{\text{fuzzy}}\big(&\boldsymbol{\theta}, \hat{y}(k-1,\boldsymbol{\theta}), \ldots, \hat{y}(k-n_y,\boldsymbol{\theta}),\\ &u(k-n_d+1), \ldots, u(k-n_u-n_d), w^1(\boldsymbol{z}_k), \ldots, w^r(\boldsymbol{z}_k)\big)\end{aligned} \tag{5.41}$$

and $\boldsymbol{\theta} = \left[(\boldsymbol{\theta}^1)^T, \ldots, (\boldsymbol{\theta}^r)^T\right]^T$ is the stacked vector of the consequent parameters. Note that f_{fuzzy} represents the fuzzy relationship given in (5.22). The Levenberg-Marquardt update step for this problem can now be written as follows [141, 153]

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i - \eta_i \left(\boldsymbol{\mathcal{J}}(\boldsymbol{\theta}_i)^T \boldsymbol{\mathcal{J}}(\boldsymbol{\theta}_i) + \alpha_i \mathbf{I}\right)^{-1} \boldsymbol{\mathcal{J}}(\boldsymbol{\theta}_i)^T \left(\boldsymbol{y} - \hat{\boldsymbol{y}}(\boldsymbol{\theta}_i)\right). \tag{5.42}$$

where $\boldsymbol{\mathcal{J}}(\boldsymbol{\theta}_i)$ represents the Jacobian at iteration step i, η_i is the stepsize, and α_i is the regularization parameter. Hence, the Hessian is approximated with $\boldsymbol{\mathcal{J}}^T\boldsymbol{\mathcal{J}}$ in combination with a regularization term. The regularization parameter and the step size are considered for tuning. Detailed strategies are discussed, e.g., in [153, 141]. The tricky part is now calculating the Jacobian efficiently. Starting with given initial values $\frac{\partial \hat{y}(k,\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}_i}$ $k = 1, \ldots, n_y$ the Jacobi matrix $\boldsymbol{\mathcal{J}}$ is recursively obtained at each iteration step i as follows [153, p. 566]:

$$\begin{aligned}
\frac{\partial \hat{y}(n_y+1,\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}_i} &= \frac{\partial f_{\text{fuzzy}}(\,\cdot\,)}{\partial \boldsymbol{\theta}_i} + \sum_{j=1}^{n_y} \frac{\partial f_{\text{fuzzy}}(\,\cdot\,)}{\partial \hat{y}(n_y+1-j,\boldsymbol{\theta}_i)} \frac{\partial \hat{y}(n_y+1-j,\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}_i}\\
\frac{\partial \hat{y}(n_y+2,\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}_i} &= \frac{\partial f_{\text{fuzzy}}(\,\cdot\,)}{\partial \boldsymbol{\theta}_i} + \sum_{j=1}^{n_y} \frac{\partial f_{\text{fuzzy}}(\,\cdot\,)}{\partial \hat{y}(n_y+2-j),\boldsymbol{\theta}_i)} \frac{\partial \hat{y}(n_y+2-j,\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}_i}\\
&\vdots\\
\frac{\partial \hat{y}(N,\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}_i} &= \underbrace{\frac{\partial f_{\text{fuzzy}}(\,\cdot\,)}{\partial \boldsymbol{\theta}_i}}_{\text{static term}} + \underbrace{\sum_{j=1}^{n_y} \frac{\partial f_{\text{fuzzy}}(\,\cdot\,)}{\partial \hat{y}(N-j,\boldsymbol{\theta}_i)} \frac{\partial \hat{y}(N-j,\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}_i}}_{\text{dynamic term}}.
\end{aligned} \tag{5.43}$$

Note that for building the derivatives of the global fuzzy model output (5.22) with respect to the parameters it can be assumed that the degrees of fulfillment are constant and therefore do not depend on the parameters.

5.2.3 Identification and Validation

In order to verify the proposed fuzzy model architecture real measurements from the biomass plant in Grossarl (Salzburg) are utilized. In Grossarl two biomass boilers with a heat power of 3.5 MW and 1.5 MW are installed. Here in particular data of the biomass boiler with 3.5 MW are considered for identification of the fuzzy models. The time frame of the available data is 04.08.2012 - 24.08.2012 with a resolution of 5 s. About 80% of the data were used for training and the remaining part for validation. Unfortunately, for the combustion position no reliable data were available. To generate sufficient identification data, simulated model output from the grey-box model developed in Sec. 5.1 was used. The following results were also published by the author in [74].

$$y(k+1) = \sum_{i=1}^{n_y} a_i(\boldsymbol{z})y(k-i+1) + \sum_{i=1}^{n_u} b_i(\boldsymbol{z})u(k-i+1) + c(\boldsymbol{z}). \tag{5.44}$$

Model Structures

For identification purposes it is reasonable to decompose the MIMO structure of the biomass furnace model into multiple MISO fuzzy models, i.e, to construct one fuzzy model for each output separately. This makes the modelling approach not only simpler but also more flexible, since a different model architecture can be applied to each MISO subproblem [153, 7]. More precisely, the following models were found to be appropriate:

$$\begin{aligned} C_{\text{pos}} &= f_{\text{fuzzy}}(C_{\text{pos}}, \dot{m}_{\text{in}}, \dot{m}_{\text{PA}} | Q_{\text{th}}), & n_y &= 2, n_{u_{1\ldots2}} = 2 \\ Q_{\text{th}} &= f_{\text{fuzzy}}(Q_{\text{th}}, \dot{m}_{\text{in}}, \dot{m}_{\text{PA}} | Q_{\text{th}}), & n_y &= 3, n_{u_{1\ldots2}} = 3 \\ \text{O}_2 &= f_{\text{fuzzy}}(\text{O}_2, \dot{m}_{\text{in}}, \dot{m}_{\text{PA}}, \dot{m}_{\text{SA}} | Q_{\text{th}}), & n_y &= 4, n_{u_{1\ldots3}} = 4 \\ T_{\text{f}} &= f_{\text{fuzzy}}(T_{\text{f}}, \dot{m}_{\text{in}}, \dot{m}_{\text{PA}}, \dot{m}_{\text{SA}}, \dot{m}_{\text{rec}} | Q_{\text{th}}, T_{\text{f}}), & n_y &= 5, n_{u_{1\ldots4}} = 5 \end{aligned} \tag{5.45}$$

where f_{fuzzy} again represents the fuzzy relationship (5.22) with the respective input fuzzy variables indicated after the vertical line and n_y and n_{u_i} indicate the order of the output and ith input, respectively. The input fuzzy variables were partly determined by expert knowledge and by considering the overall fitting performance. Theoretically, the LOLIMOT algorithm could also be applied directly with a set of candidate input fuzzy variables. Then only those variables would have been considered for the final model representation where axis-orthogonal splits at the internal nodes occur.

Model Fitting

For fitting of the fuzzy models a staged optimization approach as illustrated in Fig. 5.4 was used (cf. [153]). To be more precise, in a first step the LOLIMOT procedure is run with weighted least-squares estimation of the consequent parameters minimizing the one-step ahead prediction error. Then in a subsequent phase the fuzzy model is improved with respect to its simulation performance by using the output error optimization approach. Note that in the second step the least-squares parameter estimates are taken as initial parameter values for the iterative Levenberg-Marquart algorithm. This staged approach has the main advantage of being able to define the focus of the fuzzy model already in the training phase. If the fuzzy model is intended to be used for predictive purposes e.g, in a predictive control setup, the training procedure could be stopped after the first step, thus reducing the computational effort.

To determine the optimal number of fuzzy rules the convergence curve of LOLIMOT can be used as depicted in Fig. 5.5. Here the normalized root-mean-square simulation error (NRMSE) is plotted against the number of rules for the training and validation dataset. The NRMSE is

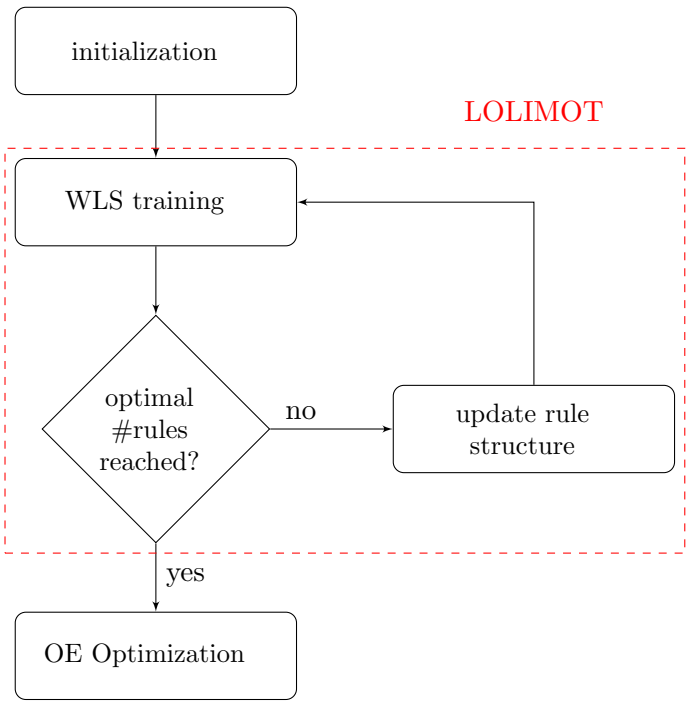


Figure 5.4: Illustration of the staged optimization approach

simply the classical root-mean-square error divided by the amplitude of the output, i.e., $\max(y) - \min(y)$. The strong decrease in the model error compared with the linear model obtained in the first few iterations reveals the highly nonlinear process characteristics. Theoretically, a minimum of the curve for the validation data set should indicate the optimal number of fuzzy rules. This implies that the model produces a good fit but does not overfit the training data. The selected model complexity for each individual fuzzy model is marked by the arrows in Fig. 5.5.

The partitioning of the fuzzy model for the combustion position, oxygen concentration of the flue gas, and the heat output is shown in Fig. 5.6. On the y-axis the degrees of membership or fulfilment are plotted which are obtained from normalizing the Gaussian membership functions. In general, two issues have a significant influence on the partitioning procedure: First, the nonlinearity of the process in the specific operating regime, and second that LOLIMOT utilizes the local sum of squared error loss function for performance comparison. Hence, splits are preferred in regions that contain more data samples. The latter is confirmed by inspecting the result for flue gas temperature shown in Fig. 5.7. In particular, for high thermal power and low flue gas temperature no training data was available. This clearly demonstrates the limitation of algorithms based on axis-orthogonal construction and further motivates the use of more advanced methods which allow for axis-oblique splits (see, e.g., [87, 88]).

Model Validation

The simulated model output for the validation data set is presented in Figures 5.8 to 5.11. All models were trained with the output error optimization approach using the Levenberg-Marquardt algorithm with 20 iterations. For step sizes $\eta = 0.08$ and for the regularization parameter $\alpha = 1$ was chosen. Aside from the process noise of the measured data (except for the combustion position where noise-free simulated output data are utilized), good agreement between measured output and simulated output can be found. In particular, for the oxygen concentration (see Fig. 5.10) dominant process noise can be observed. This is one reason for the relatively poor model performance with respect to the simulation error. Contrary, the fuzzy

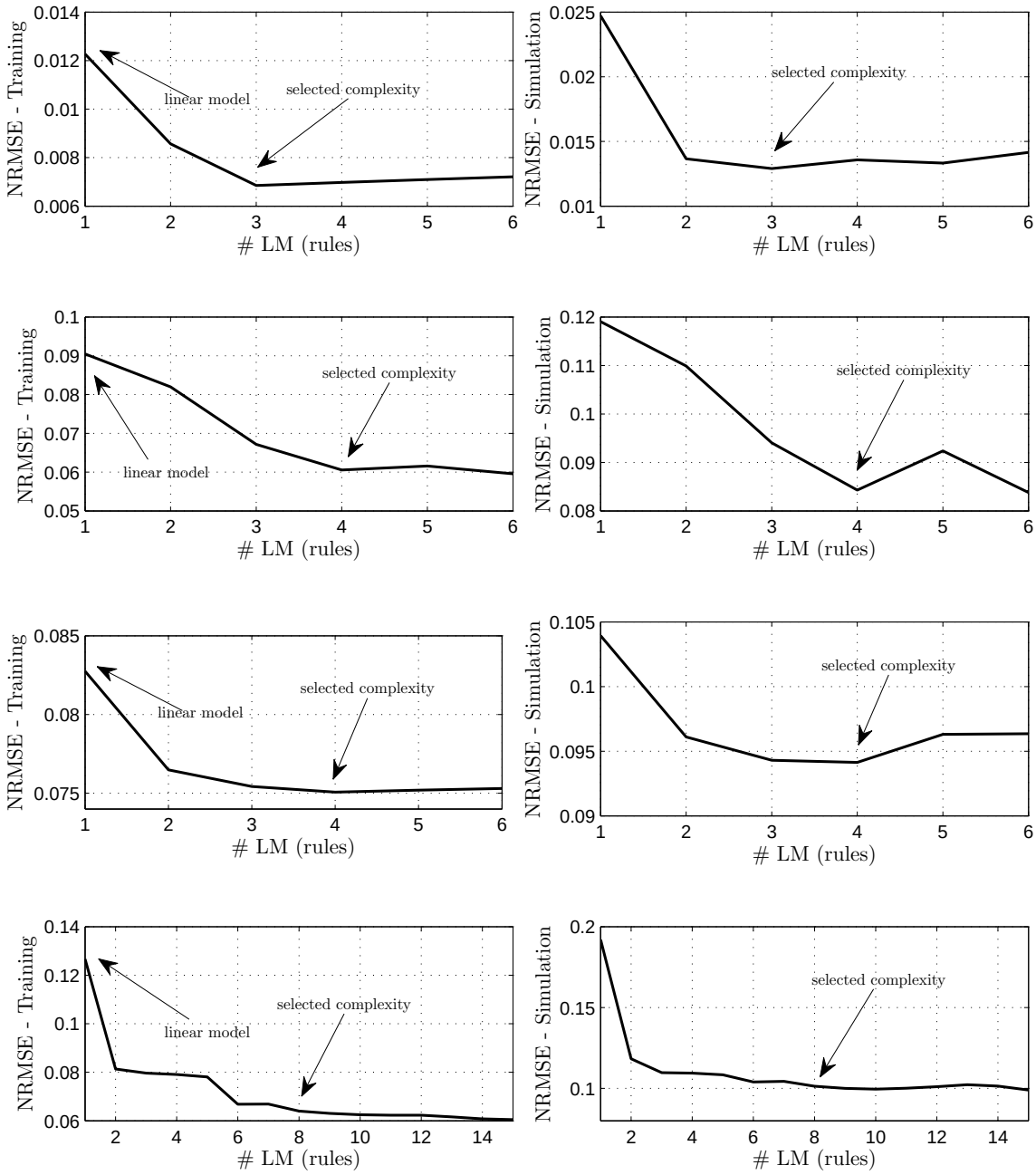


Figure 5.5: Normalized mean squared simulation error for the training dataset and the validation dataset against the number of local models (fuzzy rules); (from top to bottom) combustion position, thermal power, oxygen concentration, and flue gas temperature; the arrow indicates the selected model complexity

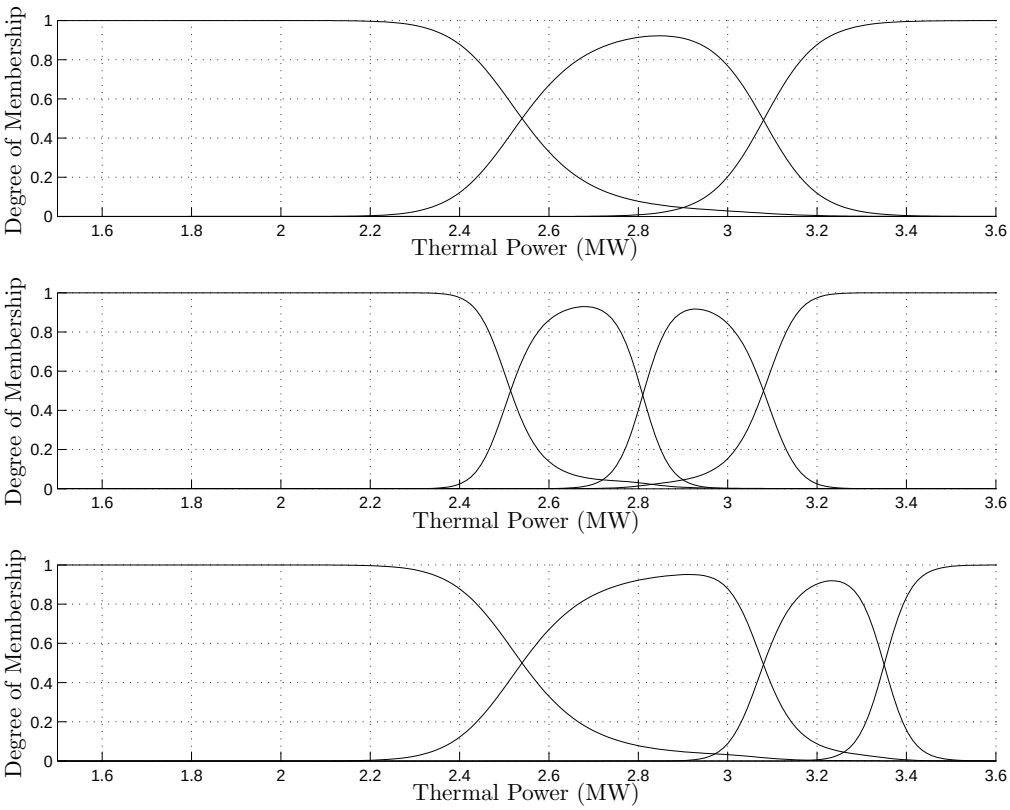


Figure 5.6: Normalized Gaussian membership functions constructed by LOLIMOT for combustion position (top), thermal power (middle), and oxygen concentration (bottom)

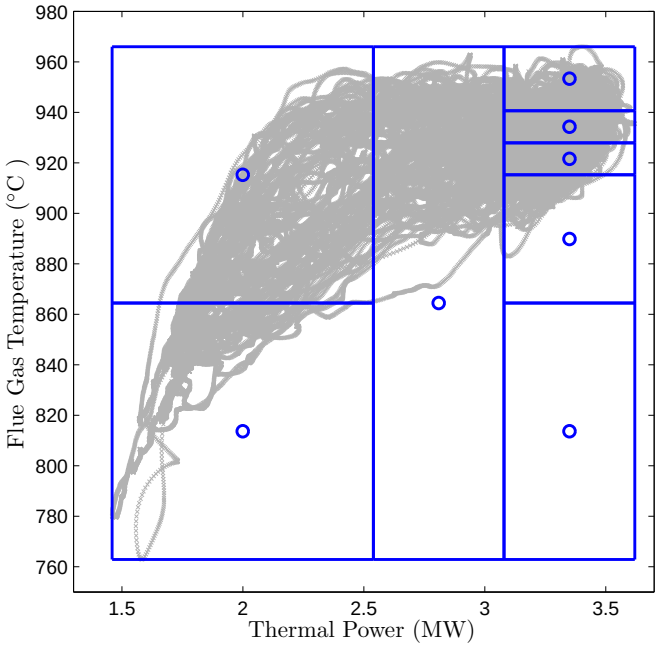


Figure 5.7: Fuzzy space decomposition constructed by LOLIMOT for flue gas temperature; the gray points are the training data samples

model for heat output shows minimal error and produces nearly a perfect fit. For the combustion position also a perfect fit is obtained which is due to the noise-free data.

Dimension Reduction via PCA

A clear relationship between the observations of flue gas temperature and thermal power can be seen in Fig. 5.12. This fact could be exploited to construct a lower dimensional fuzzy input space by using dimension reduction techniques such as principal component analysis (PCA) [93]. The basic idea of PCA is to build linear combinations of the original variables that capture most of the information, and using a subset of these transformed variables (principal components) [73]. The extension to nonlinear principal component analysis (NLPCA) generalizes the principal components from straight lines to principal curves that approximate the nonlinear relationship between a set of two variables [120]. For instance, Kernel-PCA finds principal components which are nonlinearly related to the original input variables by performing PCA in the so called feature space $\mathcal{H}$ produced by a nonlinear mapping $\Phi\colon x \to \mathcal{H}$ [69]. Note that Kernel-PCA utilizes the so called kernel matrix to approximate the nonlinear mapping. The size of the kernel matrix is square with the number of observations in the data set and is thus useful when there are much more input variables than observations [203].

To evaluate the performance with reduced fuzzy input space, identification runs using PCA and NLPCA were carried out. For the latter the following scheme is proposed:

- Fit a quadratic polynomial function through the standardized data
- Project the data points onto the fitted quadratic curve
- Calculate the arc-lengths between the projected data points and the origin

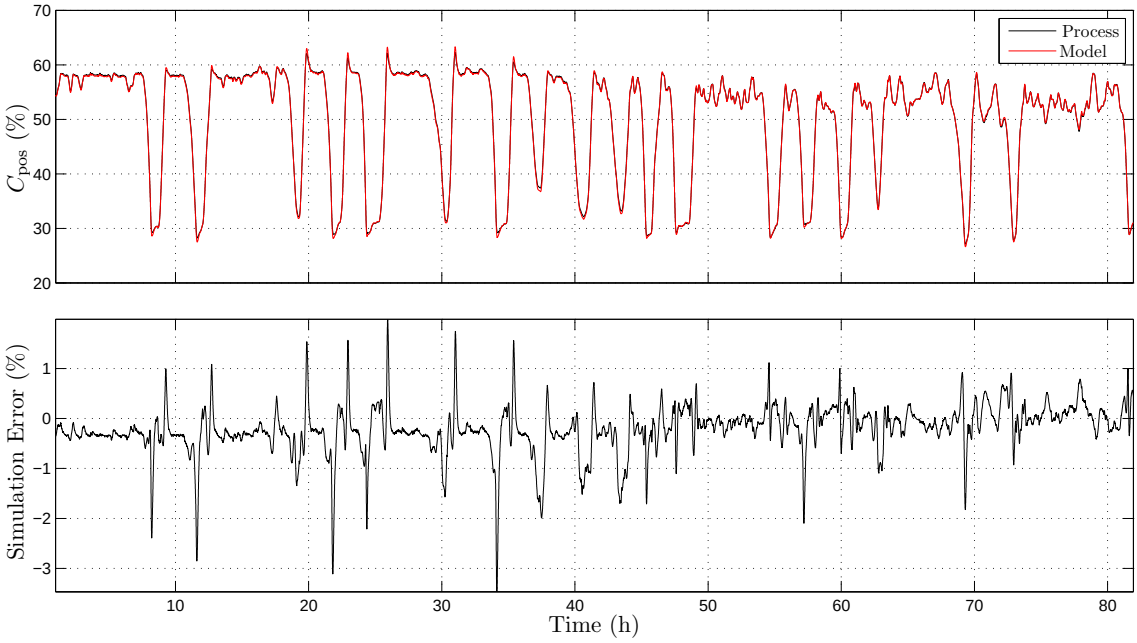


Figure 5.8: Oxygen concentration: comparison between process and model output (top); simulation error for the validation dataset (bottom)

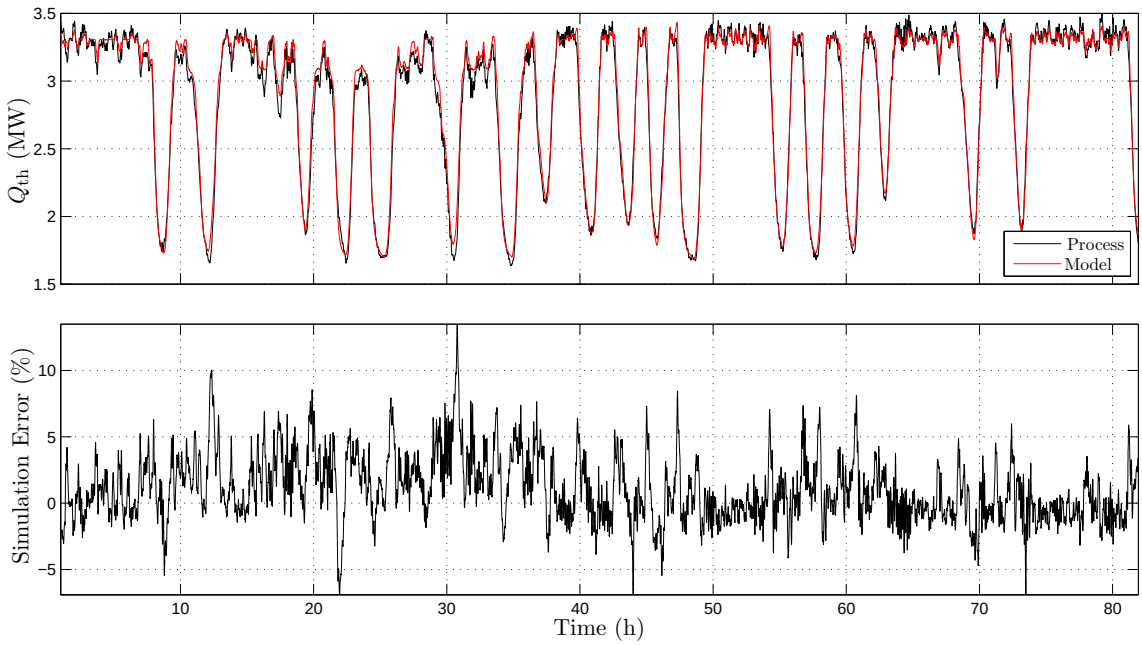


Figure 5.9: Heat output: comparison between process and model output (top); simulation error for the validation dataset (bottom)

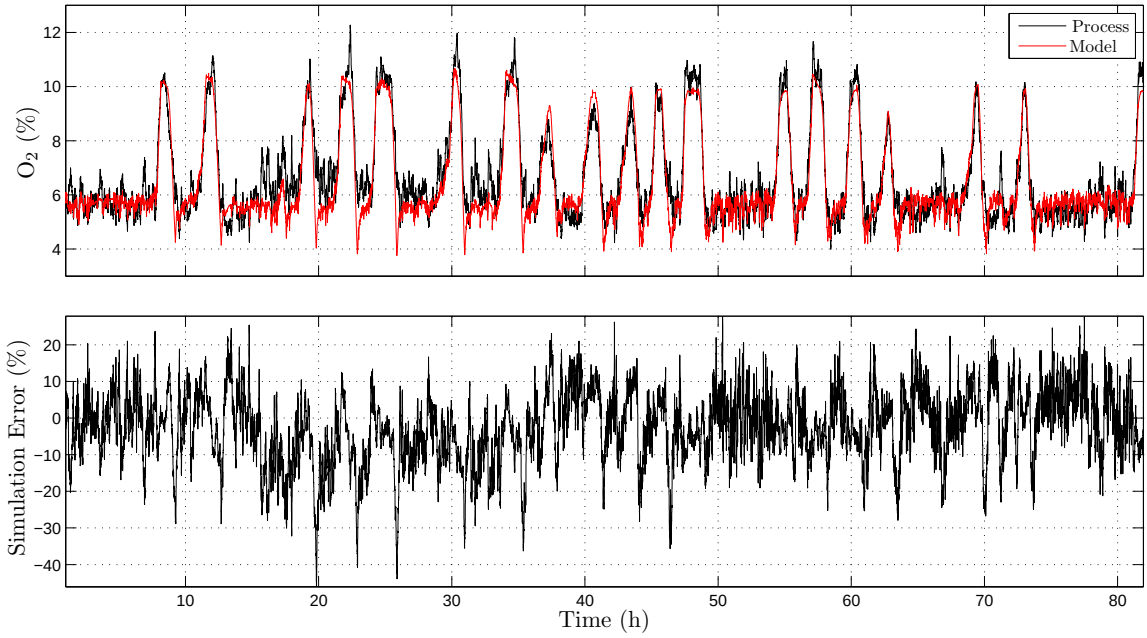


Figure 5.10: Oxygen concentration: comparison between process and model output (top); simulation error for the validation dataset (bottom)

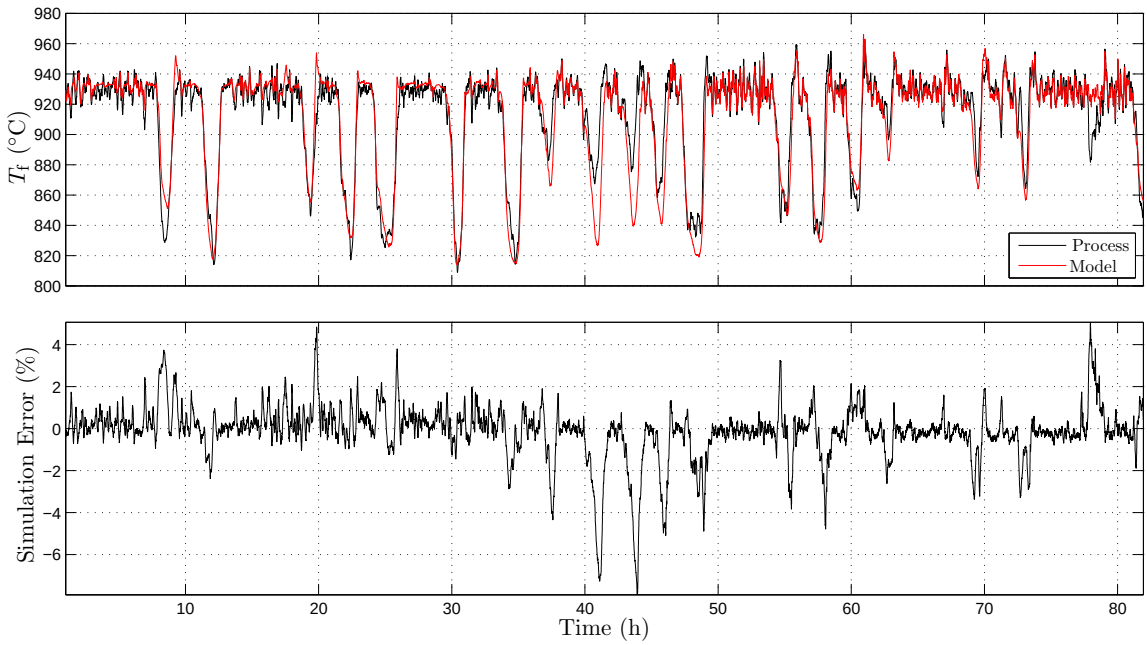


Figure 5.11: Flue gas temperature: comparison between process and model output (top); simulation error for the validation dataset (bottom)

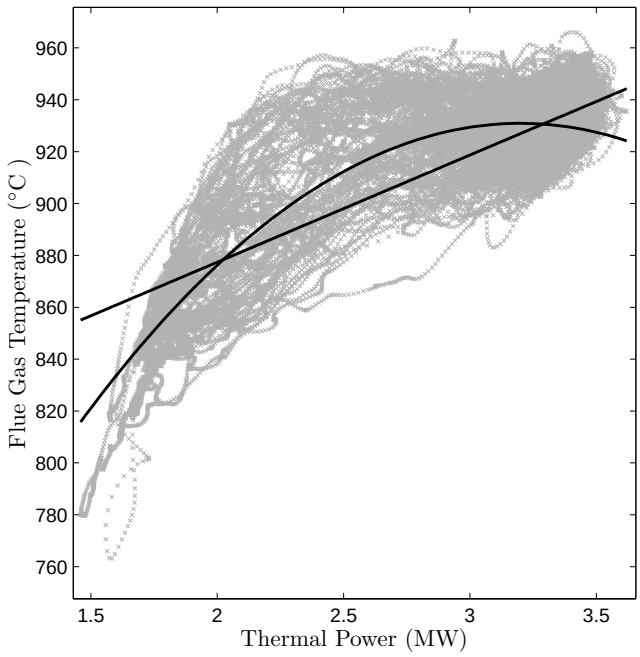


Figure 5.12: Scatterplot of flue gas temperature versus heat output plus linear and quadratic function describing the dependency

Table 5.2: Comparison of the fitting performance using PCA and NLPCA

	Classic	PCA	NLPCA
#Dim $\boldsymbol{z}$	2	1	1
#LM (rules)	8	4	4
RMSE Training	8.720	5.993	5.732
RMSE Validation	11.826	7.473	6.930

The arc-length here can be regarded similar to the score variable in PCA that represents the distance of the projected data point from the origin. (cf. [120]). The RMSE for the training and validation data set after output error optimization is summarized in Tab. 5.2. For both the PCA and NLPCA case the number of local models in the LOLIMOT algorithm was reduced to four and only the first principal component/curve is considered as fuzzy input space. The results are highly unexpected. The training and validation error using NLPCA could be decreased by about 34% and 41% respectively, compared to the 2-dimensional fuzzy input space with eight local models. This also motivates to apply NLPCA for future problems in different settings.

Chapter 6

Model Predictive Control

Model predictive control (MPC) is a general methodology for solving control problems in the time domain and is well suited for constrained and multivariate process applications [28, 207, 171, 80, 152]. Some of the popular names associated with MPC are Dynamic Matrix Control (DMC), Model Algorithmic Control (MAC), Generalized Predictive Control (GPC), etc. While these algorithms differ in certain details, the main ideas behind them are very similar [12]. All approaches are based on the following main concepts:

1. An explicit process model is used to predict the process output at future discrete time instants, over a *prediction horizon*.
2. A sequence of future control actions is computed over a *control horizon* by minimizing a given objective function.
3. Only the first control action in the sequence is applied, the horizons are moved one sample towards the future and the entire calculation is repeated at subsequent control intervals. This is called the *receding horizon* principle.

In the next section the basic concepts of predictive control are reviewed in more detail and the mathematical formulation for the MPC using state-space models is presented.

6.1 Basic Concept

Process Model. An accurate model of the process is vitally important in MPC. A good process model should fully capture the process dynamics, should be capable of allowing the predictions to be computed, to be intuitive, and to permit theoretic analysis [28]. In the framework of predictive control the process model is a necessity to predict the process output at future instants $\hat{y}(k+i|k)$ $i = 1, 2, \ldots$. Due to the general structure of MPC, various types of process models can be applied to represent the relationship between the outputs and the measurable inputs. In general, the methods for constructing the process model can be classified into two categories: *First-principles* or *white-box* models which use physical principles such as energy or mass balances to construct a model, and *black-box* models which use measured data to determine parameters (system identification) of a mathematical equation describing the input-output behaviour.

The relationship between inputs and outputs is in many cases modeled linearly, although real processes are often inherently nonlinear. Such nonlinear systems can be handled with, for instance, Fuzzy MPC [7] which is introduced in Sec. 6.3. For more details about general nonlinear MPC the reader is referred to [3].

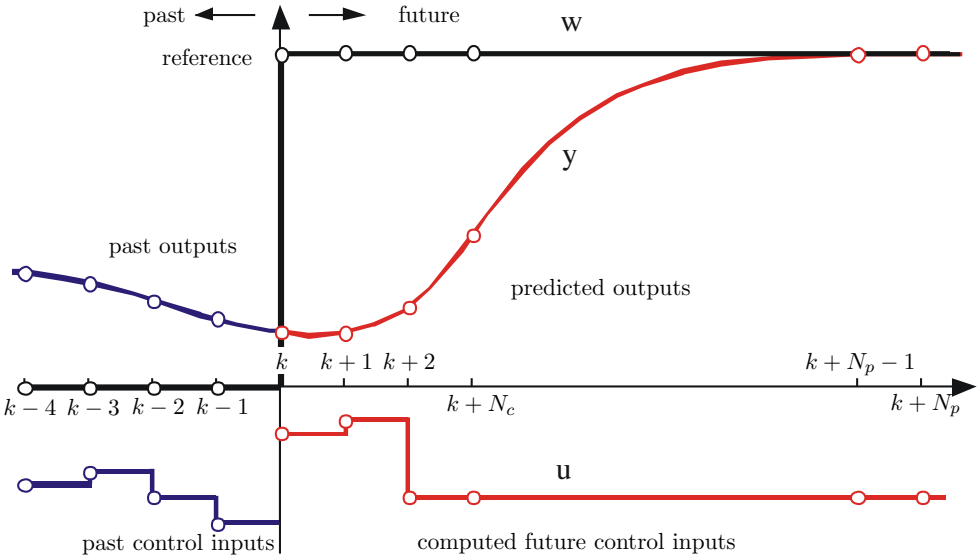


Figure 6.1: Principle of predictive control

Prediction and Control Horizons. The future process outputs are predicted over the so called *prediction horizon* N_p using the aforementioned process model. The predicted output values $\hat{y}(k+i|k),\ i = 1, \ldots, N_p$, depend on the current state of the process at time instant k and on the future input control signals $u(k+i),\ i = 0, \ldots, N_c-1$, where $N_c \leq N_p$ denotes the *control horizon*. It is further assumed that the control signal is only manipulated within the control horizon and is kept constant afterwards, i.e., $u(k+1) = u(k+N_c-1)$ for $i = N_c, \ldots, N_p-1$ [7, 207] (cp. Fig. 6.1). There is no general rule for determining the optimal prediction and control horizons. Both parameters are considered to be tuning parameters in the predictive control setup. Note that the prediction horizon has a direct influence on the condition of the optimization problem [207] and the control horizon determines the number of manipulated variables and thus affects the computing time.

Objective Function. The sequence of future control signals $u(k+i),\ i = 0, \ldots, N_c-1$ is computed by optimizing a given objective (cost) function. The aim is to bring and keep the process output as close as possible to the given reference trajectory or set-point signal r. A typical objective function has the following quadratic form [35]

$$J_k = \sum_{i=1}^{N_p} q_i[r(k+i) - \hat{y}(k+i)]^2 + \sum_{i=1}^{N_c} r_i \Delta u(k+i-1)^2. \tag{6.1}$$

The first term accounts for minimizing the variance of the process output from the reference trajectory, while the second terms penalizes excessive control actions. Note also that the control input u itself can be included in the objective function, or other filtered forms of u [7]. The parameters q_i and r_i define the weighting of the output error and the control effort with respect to each other. These weighting terms are considered to be tuning parameters for the desired closed-loop performance [207]. Now each time instant the following optimization problem has to be solved

$$\begin{gathered} \min_{\Delta u(k|N_c)} J_k \\ \text{with respect to linear constraints on } \Delta u, u \text{ and } y. \end{gathered} \tag{6.2}$$

Here, additionally, constraints on the control input, the control input increments, and the model output are imposed to guarantee that physical limits are not violated. Optimization problems

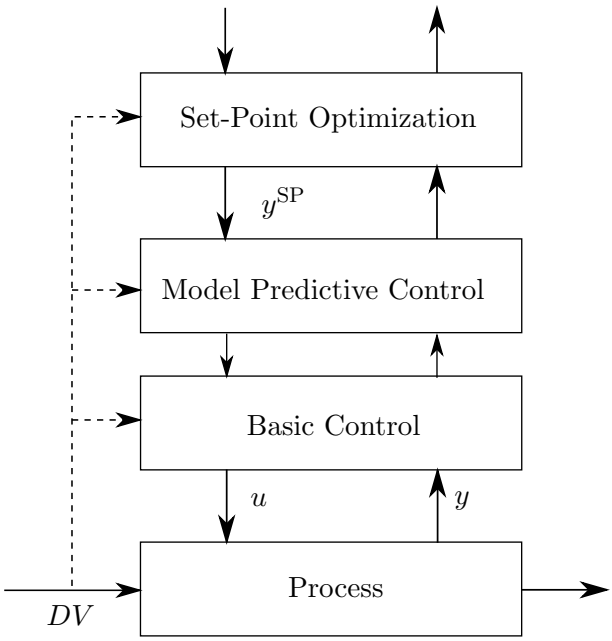


Figure 6.2: Multilayer control structure (reproduced with modifications from [197])

of this type can efficiently be solved by quadratic programming techniques [70]. In the absence of constraints, due to the quadratic nature of the cost function, an explicit analytical solution can be obtained. Generally, any other suitable cost function can be used. For instance, in the context of robust MPC, infinity norms l_∞ are used [28]. This leads to min-max problems which can be solved with methods of linear programming.

Receding Horizon Principle. Although the output of the optimization problem contains future control increments $\Delta u(k+i),\ i = 0, \ldots, N_c - 1$, only the first sample of this sequence is applied to the process, while disregarding the rest of the sequence. At the next sampling instant, the process output $y(k+1)$ is measured and the optimization and prediction can be repeated with the updated values. This procedure is called the receding horizon principle. The control action $\Delta u(k+1)$ computed at time step $k+1$ will be generally different from the one computed at time step k, since more up-to-date information about the process is available. As there will always be some mismatch between real process and process model [8], this feedback not only allows to update the current system, but also to re-identify the process model.

A common multilayer control structure in managing process plants is depicted in Fig. 6.2 [196, 197]. The MPC is here vertically placed in the so called advanced control layer trying to operate the process efficiently and safely at the desired steady state. The basic control layer is typically equipped with simple controllers for maintaining the process in a safe operation mode. The optimization layer generates optimal set-point values for the MPC by trying to minimize an overall economic profit function.

6.2 State-Space MPC

This section is dedicated to the design of a model predictive controller with the main focus lying on plant processes which can be represented by state-space models. The use of a state-space model should not pose any restriction, since other process model structures such as FIR or ARX can be put into state-space form as demonstrated in the end of this section. For more details about, e.g, DMC which use FIR process models or GPC see [28, 196].

6.2.1 Augmented Model Design

Consider the following form of the process state-space equations

$$\boldsymbol{x}(k+1) = \boldsymbol{A}\boldsymbol{x}(k) + \boldsymbol{B}\boldsymbol{u}(k) + \boldsymbol{E}\boldsymbol{z}(k) + \boldsymbol{F}\boldsymbol{w}(k) \tag{6.3}$$

$$\boldsymbol{y}(k) = \boldsymbol{C}\boldsymbol{x}(k), \tag{6.4}$$

where $\boldsymbol{u}(k)$ is the vector of input variables or manipulated variables with dimension n_u, $\boldsymbol{y}(k)$ is the process output with dimension n_y, $\boldsymbol{x}(k)$ is the state vector with dimension n_x, $\boldsymbol{z}(k)$ is the vector of measurable input disturbances with dimension n_z, $\boldsymbol{w}(k)$ denotes the unknown input disturbances and $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{E}$ and $\boldsymbol{F}$ are the state matrix, input matrix, output matrix and input disturbance matrices with appropriate dimension, respectively. Further assume that the non-measurable input disturbance $\boldsymbol{w}(k)$ follows an integrated white noise process. This means that $\boldsymbol{w}(k) - \boldsymbol{w}(k-1) = \boldsymbol{\epsilon}(k)$, where $\boldsymbol{\epsilon}(k)$ is a zero-mean, white noise sequence.

In the next step the state-space model is formulated with respect to the increments of the control signals, which are the variables to be optimized (decision variables) in the MPC problem (6.2). This is done by embedding an integrator which also allows offset free control. The following derivations and notation are largely based on that used in [207]. Taking the difference on both sides of (6.3) one gets

$$\Delta\boldsymbol{x}(k+1) = \boldsymbol{A}\Delta x(k) + \boldsymbol{B}\Delta\boldsymbol{u}(k) + \boldsymbol{E}\Delta\boldsymbol{z}(k) + \boldsymbol{F}\boldsymbol{\epsilon}(k) \tag{6.5}$$

where $\Delta\boldsymbol{x}(k+1) = \boldsymbol{x}(k+1) - \boldsymbol{x}(k)$ denote the difference of the state variables and $\Delta\boldsymbol{u}(k)$ and $\Delta\boldsymbol{z}(k)$ denote the difference of the input variables and measurable disturbance vector, respectively. Next the process output differences $\Delta y(k)$ are linked to $\Delta\boldsymbol{x}(k)$ using (6.4) as follows

$$\Delta\boldsymbol{y}(k) = \boldsymbol{C}\Delta\boldsymbol{x}(k+1) = \boldsymbol{C}\boldsymbol{A}\Delta x(k) + \boldsymbol{C}\boldsymbol{B}\Delta\boldsymbol{u}(k) + \boldsymbol{C}\boldsymbol{E}\Delta\boldsymbol{z}(k) + \boldsymbol{C}\boldsymbol{E}\boldsymbol{\epsilon}(k). \tag{6.6}$$

Putting together (6.5) with (6.6) leads to an augmented state-space model:

$$\overbrace{\begin{bmatrix}\Delta\boldsymbol{x}(k+1)\\ \boldsymbol{y}(k+1)\end{bmatrix}}^{\boldsymbol{\xi}(k+1)} = \overbrace{\begin{bmatrix}\boldsymbol{A} & \boldsymbol{0}^T\\ \boldsymbol{C}\boldsymbol{A} & \mathbf{I}\end{bmatrix}}^{\boldsymbol{\mathcal{A}}}\overbrace{\begin{bmatrix}\Delta\boldsymbol{x}(k)\\ \boldsymbol{y}(k)\end{bmatrix}}^{\boldsymbol{\xi}(k)} + \overbrace{\begin{bmatrix}\boldsymbol{B}\\ \boldsymbol{C}\boldsymbol{B}\end{bmatrix}}^{\boldsymbol{\mathcal{B}}}\Delta\boldsymbol{u}(k) + \overbrace{\begin{bmatrix}\boldsymbol{C}\\ \boldsymbol{C}\boldsymbol{E}\end{bmatrix}}^{\boldsymbol{\mathcal{E}}}\Delta\boldsymbol{z}(k) + \overbrace{\begin{bmatrix}\boldsymbol{C}\\ \boldsymbol{C}\boldsymbol{F}\end{bmatrix}}^{\boldsymbol{\mathcal{F}}}\boldsymbol{\epsilon}(k) \tag{6.7}$$

$$\boldsymbol{y}(k) = \overbrace{\begin{bmatrix}\boldsymbol{0} & \mathbf{I}\end{bmatrix}}^{\boldsymbol{\mathcal{C}}}\begin{bmatrix}\Delta\boldsymbol{x}(k)\\ \boldsymbol{y}(k)\end{bmatrix} \tag{6.8}$$

where the $\mathbf{0}$ symbol denotes zero matrices and the $\mathbf{I}$ symbol the identity matrices with appropriate dimension, respectively. The triplet $(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{C}})$ (quadruplet $(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{C}}, \boldsymbol{\mathcal{E}})$) with state vector $\boldsymbol{\xi}(k)$ is called the *augmented model*, which will be used in the design of the predictive controller.

Controllability, Observability and Minimal Realization Important concepts in the context of predictive control are controllability and observability of the state-space triplet $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ in (6.3) and (6.4). Controllability is a prerequisite for the MPC to achieve the desired closed-loop control performance and observability is a pre-requisite for the design of an observer [207]. A system is said to be state controllable if it is possible to transfer the system from any arbitrary initial state $\boldsymbol{x}(0)$ to any desired arbitrary state in a finite time period. That is, a control system is controllable if every state variable can be controlled in a finite time period by some unconstrained input signal. On the other hand, a system is said to be completely observable if every initial state $\boldsymbol{x}(0)$ can be determined from the observation of $\boldsymbol{y}(k)$ over a finite number of

sampling instances $k = 1, \ldots, N$ [162]. Mathematically, these conditions can be formulated in terms of the controllability matrix $\boldsymbol{\mathcal{C}}(\boldsymbol{B}, \boldsymbol{A})$ and observability matrix $\mathcal{O}(\boldsymbol{C}, \boldsymbol{A})$:

$$\boldsymbol{\mathcal{C}}(\boldsymbol{B}, \boldsymbol{A}) = \begin{bmatrix} \boldsymbol{B} & \boldsymbol{AB} & \boldsymbol{A}^2\boldsymbol{B} & \cdots & \boldsymbol{A}^{n_x-1}\boldsymbol{B} \end{bmatrix} \in \mathbb{R}^{n_x \times (n_x n_u)}, \tag{6.9}$$

$$\mathcal{O}(\boldsymbol{C}, \boldsymbol{A}) = \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{CA} \\ \boldsymbol{CA}^2 \\ \vdots \\ \boldsymbol{CA}^{n_x-1} \end{bmatrix} \in \mathbb{R}^{(n_y n_x) \times n_x}. \tag{6.10}$$

The system $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ is called controllable if the matrix $\boldsymbol{\mathcal{C}}(\boldsymbol{B}, \boldsymbol{A})$ has full rank (i.e., is nonsingular) and observable if the matrix $\mathcal{O}(\boldsymbol{C}, \boldsymbol{A})$ has full rank. Since the MPC design is performed on the basis of the augmented state-space model, it is important that the augmented model is both controllable and observable [207]. First, recall the definition of minimal realization (see [207, p. 24]):

Definition 6.2.1 (Minimal Realization)**.** *A realization of transfer function $\boldsymbol{G}(z)$ is given by any state-space triplet $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ such that $\boldsymbol{G}(z) = \boldsymbol{C}(z\mathbf{I} - \boldsymbol{A})^{-1}\boldsymbol{B}$. If such system $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ exists, then $\boldsymbol{G}(z)$ is said to be realizable. A realization $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ is called a minimal realization of a transfer function if no other realization of smaller dimension of the triplet exists.*

In other words, a realization $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ is minimal (i.e., the smallest number of state variables) if there exist no other triple $(\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{C}})$ with $\tilde{\boldsymbol{A}}$ having smaller dimension than $\boldsymbol{A}$ and such that $\tilde{\boldsymbol{C}}(z\mathbf{I} - \tilde{\boldsymbol{A}})^{-1}\tilde{\boldsymbol{B}} = \boldsymbol{C}(z\mathbf{I} - \boldsymbol{A})^{-1}\boldsymbol{B}$ [107]. Furthermore, as shown in [107, p. 127], a realization is minimal if the denominator $\boldsymbol{a}(z) = \det(z\mathbf{I} - \boldsymbol{A})$ and numerator $\boldsymbol{b}(z) = \boldsymbol{C}\mathrm{Adj}(z\mathbf{I} - \boldsymbol{A})\boldsymbol{B}$ are relatively prime where $\mathrm{Adj}(\,\cdot\,)$ denotes the adjugate matrix. A minimal realization has the following important property:

Theorem 6.2.1. *A minimal realization is both controllable and observable.*

Proof. See Theorem 2.2-4, 2.2-5 and 2.2-6 in [107]. □

Based on this it can now be shown (see Theorem 2.1 in [207]) that if the plant state-space model is both controllable and observable and has the transfer function $\boldsymbol{G}_m(z)$ with minimal realization, then the augmented model (6.7) is both controllable and observable if and only if the plant model $\boldsymbol{G}_m(z)$ has no zero at $z = 1$.

Classical Models in State-Space Representation. The basic idea is to put the feedback variables that have been used in the classical predictive control scheme into the state vector of the state-space model [207]. First, assume that the process model has the following MIMO-ARX representation

$$\boldsymbol{y}(k+1) = \sum_{i=1}^{n_y} \boldsymbol{F}_i \boldsymbol{y}(k-i+1) + \sum_{i=1}^{n_u} \boldsymbol{H}_i \boldsymbol{u}(k-i+1) + \boldsymbol{c}, \tag{6.11}$$

then one way to obtain a state-space representation is by choosing the state vector $\boldsymbol{x}(k)$ as [207, 150]

$$\boldsymbol{x}(k)^T = \begin{bmatrix} \boldsymbol{y}(k)^T & \ldots & \boldsymbol{y}(k-n_y+1)^T & \boldsymbol{u}(k-1)^T & \ldots & \boldsymbol{u}(k-n_u+1)^T & 1 \end{bmatrix}. \tag{6.12}$$

Hence, the state vector accomodates the previous outputs and inputs of the system as well as the one as last element which is required for offset terms. The corresponding state-space system matrices are defined as the block matrices

$$\boldsymbol{A} = \left[\begin{array}{cc|c} \boldsymbol{A}_1 & \boldsymbol{A}_2 & \\ \boldsymbol{A}_3 & \boldsymbol{A}_4 & \boldsymbol{A}_5 \end{array}\right]; \quad \boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_1 \\ \boldsymbol{B}_2 \end{bmatrix}; \quad \boldsymbol{C} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \tag{6.13}$$

$$
\begin{aligned}
\boldsymbol{A}_1 &= \begin{bmatrix} \boldsymbol{F}_1 & \boldsymbol{F}_2 & \cdots & \boldsymbol{F}_{n_y-1} & \boldsymbol{F}_{n_y} \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} & \mathbf{0} \end{bmatrix}; \quad \boldsymbol{A}_2 = \begin{bmatrix} \boldsymbol{H}_2 & \boldsymbol{H}_3 & \cdots & \boldsymbol{H}_{n_u-1} & \boldsymbol{H}_{n_u} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \end{bmatrix}; \\
\boldsymbol{A}_4 &= \begin{bmatrix} \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I} & \mathbf{0} \end{bmatrix}; \quad \boldsymbol{A}_5 = \begin{bmatrix} \boldsymbol{c} \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \\
\boldsymbol{B}_1 &= \begin{bmatrix} \boldsymbol{H}_1 & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix}^T; \quad \boldsymbol{B}_2 = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix}^T;
\end{aligned}
\tag{6.14}
$$

where $\boldsymbol{A}_3$ is a zero matrix. For offset free control the resulting state-space triplet can be augmented as shown in (6.7). Notice that in this case the offset term vanishes from (6.12) due to the representation in deviation form.

In a similar way the state-space representation for MIMO-FIR models is obtained. Here all terms related to the previous outputs of the process vanish from the state vector (6.12) which is then written as

$$
\boldsymbol{x}(k)^T = \begin{bmatrix} \boldsymbol{u}(k-1)^T & \dots & \boldsymbol{u}(k-n_u+1)^T \end{bmatrix}. \tag{6.15}
$$

Thus, the state-space matrix $\boldsymbol{A}$ reduces to a pure shift matrix with ones only on the subdiagonal and zeroes elsewhere. On the other hand, the $\boldsymbol{C}$ matrix accommodates all the FIR coefficients as well as the offset terms. In order to be able to cover the dynamic response of the process the order n_u of FIR models is typically large. Hence, also the state vector $\boldsymbol{x}(k)$ has large dimension. In such cases a more parsimonious predictive controller such as the classical DMC might be preferable.

It is important to mention that the state-space triplet $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ obtained as above is in general not a minimal realization [207]. Such models are also termed non-minimal state-space realizations (NMSS) [207]. This further means, according to (6.2.1), that the state-space model is not observable or/and not controllable. Although with this special choice of state variables as given in 6.12, the entire state vector is available through measurements. Hence, there is no need to design an observer where observability would be a prerequisite. Furthermore, state controllability can be extended to output controllability which describes the ability of an input to move the output from any initial condition to any final condition in a finite time period [162]. Output controllability is equivalent to $\boldsymbol{C}\mathcal{C}(\boldsymbol{B}, \boldsymbol{A})$ having full rank. There also exist methods to eliminate uncontrollable or unobservable states (e.g., command `minreal` in MATLAB®). For detailed discussion on this see [186].

6.2.2 Prediction of Process Output

The goal is here to predict future state variables and output variables based on the current state vector $\boldsymbol{\xi}(k)$, the future set of control moves $\Delta\boldsymbol{u}(k+i-1),\ i = 1, \dots, N_c$ and future known disturbance increments $\Delta\boldsymbol{z}(k+i-1),\ i = 1, \dots, N_p$. Note that it is assumed here that the entire state vector is available through measurements or can be estimated with appropriate observers [196, 205]. Furthermore, due to the assumption that $\boldsymbol{\epsilon}(k)$ is zero-mean, white noise sequence, the predicted value of $\boldsymbol{\epsilon}(k+i|k)$ at any future time instant k is zero. Hence, for the sake of clarity, future noise effects to the predicted variables are eliminated by considering the expectation of the respective variables. For notational simplicity, let $\hat{\boldsymbol{\xi}}(k+i),\ i = 1, \dots, N_p$ denote future expected state variables and $\hat{\boldsymbol{y}}(k+i),\ i = 1, \dots, N_p$ future expected output variables.

Mathematically, future values are computed by iterating the model (6.7) as follows [207]

$$\begin{aligned}
\hat{\boldsymbol{\xi}}(k+1|k) &= \boldsymbol{\mathcal{A}}\boldsymbol{\xi}(k) + \boldsymbol{\mathcal{B}}\Delta\boldsymbol{u}(k) + \boldsymbol{\mathcal{E}}\Delta\boldsymbol{z}(k) \\
\hat{\boldsymbol{\xi}}(k+2|k) &= \boldsymbol{\mathcal{A}}\boldsymbol{\xi}(k)(k+1|k) + \boldsymbol{\mathcal{B}}\Delta\boldsymbol{u}(k+1) \\
&= \boldsymbol{\mathcal{A}}^2\boldsymbol{\xi}(k) + \boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{B}}\Delta\boldsymbol{u}(k) + \boldsymbol{\mathcal{B}}\Delta\boldsymbol{u}(k+1) + \boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{E}}\Delta\boldsymbol{z}(k) + \boldsymbol{\mathcal{E}}\Delta\boldsymbol{z}(k+1) \\
&\vdots \\
\hat{\boldsymbol{\xi}}(k+n_p|k) &= \boldsymbol{\mathcal{A}}^{N_p}\boldsymbol{\xi}(k) + \sum_{i=1}^{N_c}\boldsymbol{\mathcal{A}}^{N_p-i}\boldsymbol{\mathcal{B}}\Delta\boldsymbol{u}(k+i-1) + \sum_{i=1}^{N_p}\boldsymbol{\mathcal{A}}^{N_p-i}\boldsymbol{\mathcal{E}}\Delta\boldsymbol{z}(k+i)
\end{aligned}$$

From the predicted state variables, the predicted output variables are easily obtained through

$$\begin{aligned}
\hat{\boldsymbol{y}}(k+1|k) &= \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}\boldsymbol{\xi}(k) + \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{B}}\Delta\boldsymbol{u}(k) + \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{E}}\Delta\boldsymbol{z}(k) \\
\hat{\boldsymbol{y}}(k+2|k) &= \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}^2\boldsymbol{\xi}(k) + \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{B}}\Delta\boldsymbol{u}(k) + \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{B}}\Delta\boldsymbol{u}(k+1) + \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{E}}\Delta\boldsymbol{z}(k) + \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{E}}\Delta\boldsymbol{z}(k+1) \\
\hat{\boldsymbol{y}}(k+3|k) &= \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}^3\boldsymbol{\xi}(k) + \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}^2\boldsymbol{\mathcal{B}}\Delta\boldsymbol{u}(k) + \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{B}}\Delta\boldsymbol{u}(k+1) + \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{B}}\Delta\boldsymbol{u}(k+2) \\
&\quad + \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}^2\boldsymbol{\mathcal{E}}\Delta\boldsymbol{z}(k) + \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{E}}\Delta\boldsymbol{z}(k+1) + \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{E}}\Delta\boldsymbol{z}(k+2) \\
&\vdots \\
\hat{\boldsymbol{y}}(k+N_p|k) &= \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}^{N_p}\boldsymbol{\xi}(k) + \boldsymbol{\mathcal{C}}\sum_{i=1}^{N_c}\boldsymbol{\mathcal{A}}^{N_p-i}\boldsymbol{\mathcal{B}}\Delta\boldsymbol{u}(k+i-1) + \boldsymbol{\mathcal{C}}\sum_{i=1}^{N_p}\boldsymbol{\mathcal{A}}^{N_p-i}\boldsymbol{\mathcal{E}}\Delta\boldsymbol{z}(k+i)
\end{aligned}$$

Now, the stacked vector of predicted outputs can be written in compact matrix form as

$$\hat{\boldsymbol{Y}} = \underbrace{\boldsymbol{\Phi}_u\Delta\boldsymbol{U}}_{\text{forced response}} + \underbrace{\boldsymbol{F}\xi(k) + \boldsymbol{\Phi}_z\Delta\boldsymbol{Z}}_{\text{free response}} \tag{6.16}$$

with

$$\begin{aligned}
\hat{\boldsymbol{Y}} &= \begin{bmatrix}\hat{\boldsymbol{y}}(k+1|k)^T & \hat{\boldsymbol{y}}(k+2|k)^T & \dots & \hat{\boldsymbol{y}}(k+N_p|k)^T\end{bmatrix}^T \\
\Delta\boldsymbol{Z} &= \begin{bmatrix}\Delta\boldsymbol{z}(k)^T & \Delta\boldsymbol{z}(k+1)^T & \dots & \Delta\boldsymbol{z}(k+N_p-1)^T\end{bmatrix}^T \\
\Delta\boldsymbol{U} &= \begin{bmatrix}\Delta\boldsymbol{u}(k)^T & \Delta\boldsymbol{u}(k+1)^T & \dots & \Delta\boldsymbol{z}(k+N_c-1)^T\end{bmatrix}^T.
\end{aligned}$$

The matrices $\boldsymbol{F}$ and $\boldsymbol{\Phi}$ are calculated as

$$\boldsymbol{F} = \begin{bmatrix}\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}} \\ \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}^2 \\ \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}^3 \\ \vdots \\ \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}^{N_p}\end{bmatrix}, \quad \boldsymbol{\Phi}_u = \begin{bmatrix}\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{B}} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{B}} & \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{B}} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}^2\boldsymbol{\mathcal{B}} & \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{B}} & \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{B}} & \cdots & \boldsymbol{0} \\ \vdots & & & \ddots & \vdots \\ \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}^{N_p-1} & \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}^{N_p-2}\boldsymbol{\mathcal{B}} & \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}^{N_p-3} & \cdots & \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}^{N_p-N_c}\end{bmatrix} \tag{6.17}$$

$$\boldsymbol{\Phi}_z = \begin{bmatrix}\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{E}} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{E}} & \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{E}} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}^2\boldsymbol{\mathcal{E}} & \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{E}} & \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{E}} & \cdots & \boldsymbol{0} \\ \vdots & & & \ddots & \vdots \\ \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{E}}^{N_p-1} & \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{E}}^{N_p-2}\boldsymbol{\mathcal{B}} & \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{E}}^{N_p-3} & \cdots & \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{E}}\end{bmatrix} \tag{6.18}$$

The forced response is related to the current and future control input changes which are decision variables of the controller optimization problem, whereas the free response depends only on the current state vector and the future disturbance increments. In other words, the free response is the output of the system if there is no future control action.

6.2.3 Constraints

The ability to efficiently handle constraints gives a strong advantage to MPC compared to standard linear controllers, such as PID (proportional-integral-derivative) [134] or LQR (linear-quadratic-regulator) control [80]. Especially in process industry, the constraint handling ability is of significant importance. This is motivated by the fact that the most profitable operation is often obtained when a process is running at or close to a constraint [143]. Often these constraints can be directly associated with costs. Hence, intelligent constraint handling not only guarantees that process limitations are not violated, but can also improve the overall performance.

In general the following categories may be introduced according to the type of constraint:

- Constraints on the control input variable incremental variation
- Constraints on the amplitude of the control input variable
- Output constraints

The rate of change constraints can be used to impose directional movement constraints on the control variables, such as limited slew rate of valves or actuators. Constraints on the amplitude of the control signal most often emerge in the form of saturation characteristics: valves with a finite range of adjustment, flow rates with maximum values due to fixed pipe diameters. Output constraints can be used, for instance, to ensure that the process output is of a certain quality or quantity. However, output constraints often cause large changes in both the control and incremental control variables when they are enforced. Therefore output constraints are often formulated as "soft" constraints by introducing slack variables (see [207, p. 48 f]). The set of constraints can formally be written as

$$\begin{aligned}
&\Delta u^{min} \leq \Delta u(k+i) \leq \Delta u^{max}, \quad i = 0, \ldots, N_c - 1 \\
&u^{min} \leq u(k+i) \leq u^{max}, \quad i = 0, \ldots, N_c - 1 \\
&y^{min} \leq y(k+i) \leq y^{max}, \quad i = 1, \ldots, N_p
\end{aligned}$$

or in compact vector notation

$$\begin{aligned}
&\Delta \boldsymbol{U}^{min} \leq \Delta \boldsymbol{U} \leq \Delta \boldsymbol{U}^{max} \\
&\boldsymbol{U}^{min} \leq \boldsymbol{U} \leq \boldsymbol{U}^{max} \\
&\boldsymbol{Y}^{min} \leq \boldsymbol{Y} \leq \boldsymbol{Y}^{max}.
\end{aligned}$$

In the next step all constraints are formulated as lower inequalities in terms of the control increments which are the decision variables in the quadratic programming problem. For $\Delta \boldsymbol{U}$ itself this relation becomes

$$\underbrace{\begin{bmatrix} -\mathbf{I} \\ \mathbf{I} \end{bmatrix}}_{\boldsymbol{M}_1} \Delta \boldsymbol{U} \leq \underbrace{\begin{bmatrix} -\Delta \boldsymbol{U}^{min} \\ \Delta \boldsymbol{U}^{max} \end{bmatrix}}_{\boldsymbol{N}_1} \tag{6.19}$$

The control variables are linked with the control increments by the following expression

$$\begin{bmatrix} \boldsymbol{u}(k) \\ \boldsymbol{u}(k+1) \\ \boldsymbol{u}(k+2) \\ \vdots \\ \boldsymbol{u}(k+N_c-1) \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{I} \\ \mathbf{I} \\ \vdots \\ \mathbf{I} \end{bmatrix} \boldsymbol{u}(k-1) + \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{I} & \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{I} & \mathbf{I} & \mathbf{I} & \cdots & \mathbf{0} \\ \vdots & & & & \\ \mathbf{I} & \mathbf{I} & \cdots & \mathbf{I} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \Delta \boldsymbol{u}(k) \\ \Delta \boldsymbol{u}(k+1) \\ \Delta \boldsymbol{u}(k+2) \\ \vdots \\ \Delta \boldsymbol{u}(k+N_c-1) \end{bmatrix} \tag{6.20}$$

Rewriting (6.20) in matrix form gives

$$\underbrace{\begin{bmatrix} -\boldsymbol{C}_2 \\ \boldsymbol{C}_2 \end{bmatrix}}_{\boldsymbol{M}_2} \Delta \boldsymbol{U} \leq \underbrace{\begin{bmatrix} -\boldsymbol{U}^{min} + \boldsymbol{C}_1 \boldsymbol{u}(k-1) \\ \boldsymbol{U}^{max} - \boldsymbol{C}_1 \boldsymbol{u}(k-1) \end{bmatrix}}_{\boldsymbol{N}_2}. \tag{6.21}$$

where $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ corresponds to the appropriate matrices in (6.20). Finally, the output is expressed in terms of $\Delta \boldsymbol{U}$ as (cp. Eq. (6.16))

$$\underbrace{\begin{bmatrix} -\boldsymbol{\Phi}_u \\ \boldsymbol{\Phi}_u \end{bmatrix}}_{\boldsymbol{M}_3} \Delta \boldsymbol{U} \leq \underbrace{\begin{bmatrix} -\boldsymbol{Y}^{min} + \boldsymbol{F}\boldsymbol{\xi}(k) + \boldsymbol{\Phi}_z \Delta \boldsymbol{Z} \\ \boldsymbol{Y}^{max} - \boldsymbol{F}\boldsymbol{\xi}(k) - \boldsymbol{\Phi}_z \Delta \boldsymbol{Z} \end{bmatrix}}_{\boldsymbol{N}_3} \tag{6.22}$$

For reason of compactness, the constraints are represented as

$$\boldsymbol{M} \Delta \boldsymbol{U} \leq \boldsymbol{\gamma}, \tag{6.23}$$

with

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{M}_1 \\ \boldsymbol{M}_2 \\ \boldsymbol{M}_3 \end{bmatrix}, \quad \boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{N}_1 \\ \boldsymbol{N}_2 \\ \boldsymbol{N}_3 \end{bmatrix}. \tag{6.24}$$

Here it was implicitly assumed that the constraints are time-invariant. Theoretically, they can also be formulated as time dependent and be updated at each time instant.

6.2.4 Optimization

Now the control objective for the MPC can be formulated by the following cost function

$$\begin{aligned} \min_{\Delta \boldsymbol{U}} J = \quad & \|\boldsymbol{R}_s - \hat{\boldsymbol{Y}}\|_{\boldsymbol{Q}}^2 + \|\Delta \boldsymbol{U}\|_{\boldsymbol{R}}^2 \\ \text{s. t.} \quad & \boldsymbol{M} \Delta \boldsymbol{U} \leq N \end{aligned} \tag{6.25}$$

where the vector $\boldsymbol{R}_s$ contains the set-point information

$$\boldsymbol{R}_s = \begin{bmatrix} r(k+1)^T & r(k+2)^T & \dots & r(k+N_p)^T \end{bmatrix}^T \tag{6.26}$$

and $\boldsymbol{Q}$ and $\boldsymbol{R}$ are positive semidefinite weighting matrices which allow for tuning. As discussed in the beginning of this chapter the typical objective in control applications is to penalize values of squared predicted output deviations from some reference signal and excessive control actions. In the unconstrained case an analytical solution to (6.25) can be obtained [207]. First, substitute (6.16) into the cost function (6.25) which gives

$$\begin{aligned} J \ &= (\boldsymbol{R}_s - \boldsymbol{F}\boldsymbol{\xi}(k) - \boldsymbol{\Phi}_z \Delta \boldsymbol{Z})^T \boldsymbol{Q} (\boldsymbol{R}_s - \boldsymbol{F}\boldsymbol{\xi}(k) - \boldsymbol{\Phi}_z \Delta \boldsymbol{Z}) \\ &\quad - 2\Delta \boldsymbol{U}^T \boldsymbol{\Phi}_u^T \boldsymbol{Q} (\boldsymbol{R}_s - \boldsymbol{F}\boldsymbol{\xi}(k) - \boldsymbol{\Phi}_z \Delta \boldsymbol{Z}) + \Delta \boldsymbol{U}^T (\boldsymbol{\Phi}_u^T \boldsymbol{Q} \boldsymbol{\Phi}_u + \boldsymbol{R}) \Delta \boldsymbol{U}. \end{aligned} \tag{6.27}$$

From the first derivative of (6.27) the necessary condition for the minimum of the cost function is obtained as

$$\frac{\partial J}{\partial \Delta \boldsymbol{U}} = -2\boldsymbol{\Phi}_u^T \boldsymbol{Q} (\boldsymbol{R_s} - \boldsymbol{F}\boldsymbol{\xi}(k) - \boldsymbol{\Phi}_z \Delta \boldsymbol{Z}) + 2(\boldsymbol{\Phi}_u^T \boldsymbol{Q} \boldsymbol{\Phi}_u + \boldsymbol{R}) \Delta \boldsymbol{U} = 0 \tag{6.28}$$

from which the optimal solution for the control signal can be found as

$$\Delta \hat{\boldsymbol{U}} = (\boldsymbol{\Phi}_u^T \boldsymbol{Q} \boldsymbol{\Phi}_u + \boldsymbol{R})^{-1} \boldsymbol{\Phi}_u^T \boldsymbol{Q} (\boldsymbol{R}_s - \boldsymbol{F}\boldsymbol{\xi}(k) - \boldsymbol{\Phi}_z \Delta \boldsymbol{Z}), \tag{6.29}$$

with the assumption that the Hessian matrix $(\boldsymbol{\Phi}_u^T \boldsymbol{Q} \boldsymbol{\Phi}_u + \boldsymbol{R})$ is nonsingular. Equation (6.29) also allows a good interpretation of the role of the weighting matrices. Since (6.29) can be considered as the solution of a weighted least-squares problem with weights $\boldsymbol{Q}$ combined with a Tikhonov regularization added through $\boldsymbol{R}$ [24].

For constrained control problems quadratic programming (QP) solvers are necessary [207, 196, 143]. In general, a convex optimization problem is called a *quadratic program* if the objective function is (convex) quadratic, and the contraint functions are affine [24, p. 152]. This can be expressed in the form

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\boldsymbol{x}^T \boldsymbol{P} \boldsymbol{x} + \boldsymbol{x}^T \boldsymbol{q} + \boldsymbol{r} \\ \text{subject to} \quad & \boldsymbol{M} \Delta \boldsymbol{U} \leq \boldsymbol{\gamma} \end{aligned} \tag{6.30}$$

where $\boldsymbol{P}$, $\boldsymbol{M}$, $\boldsymbol{q}$, $\boldsymbol{r}$ and $\boldsymbol{\gamma}$ are compatible matrices and vectors with $\boldsymbol{P}$ being symmetric and positive semidefinite. It can easily be seen that the MPC control objective can be reformulated in terms of quadratic programming problem. Several efficient solution techniques for quadratic programs exist in literature. For an overview see [61, 24, 70]. In this work, in particular, the command `quadprog` in MATLAB® is used with the active set method.

6.3 Fuzzy Models in MPC

6.3.1 Introduction

Linear MPC refers to a class of control algorithms that compute a manipulated variable by utilizing a linear process model [152]. Many systems are, however, in general inherently nonlinear [59]. This motivates the use of nonlinear model predictive control. Here a nonlinear and generally non-convex optimization problem has to be solved online at each sampling period even without any active constraint [96, 3, 16]. The computational burden needed to solve this problem online may make this approach infeasible in practical implementations. Moreover, the solution itself, even if found, may be only a local extremum. To avoid non-convex optimization, a set of local linear models can be extracted from a TS fuzzy model which are then utilized by the MPC algorithm [1, 150, 196, 7].

6.3.2 Linear Model Extraction

There are basically two approaches to extract a linear model from a given fuzzy model at a certain operating point [1, 2]. In the first approach linear models are obtained by interpolating the parameters of the local models in the TS model (see Eq. (5.22)+ Eq. (5.23) on page 71). The second approach extracts the parameters of the linear model by Taylor expansion. In this work only the first approach is considered.

The goal is to locally represent a TS fuzzy model by a linear state-space model

$$\boldsymbol{x}(k+1) = \boldsymbol{A}_k \boldsymbol{x}(k) + \boldsymbol{B}_k \boldsymbol{u}(k) \boldsymbol{E}_k \boldsymbol{z}(k) \tag{6.31}$$

$$\boldsymbol{y}(k) = \boldsymbol{C}_k \boldsymbol{x}(k) \tag{6.32}$$

in which system matrices $\boldsymbol{A}_k$, $\boldsymbol{B}_k$, $\boldsymbol{C}_k$ and $\boldsymbol{E}_k$ are considered to be non-constant. To arrive at this representation, assume that the MIMO TS fuzzy model can be regarded as a multivariable linear parameter-varying (LPV) system [1]

$$\boldsymbol{y}(k+1) = \sum_{i=1}^{n_y} \boldsymbol{F}_i(\boldsymbol{z}) \boldsymbol{y}(k-i+1) + \sum_{i=1}^{n_u} \boldsymbol{H}_i(\boldsymbol{z}) \boldsymbol{u}(k-i+1) + \boldsymbol{c}(\boldsymbol{z}) \tag{6.33}$$

in which parameter matrices frozen at a certain operating point $\boldsymbol{z}$ are calculated as

$$\begin{aligned}
\boldsymbol{F}_i(\boldsymbol{z}) &= \sum_{j=1}^{r} \boldsymbol{W}^j(\boldsymbol{z})\boldsymbol{F}_i^j, \quad i = 1, \ldots, n_y \\
\boldsymbol{H}_i(\boldsymbol{z}) &= \sum_{j=1}^{r} \boldsymbol{W}^j(\boldsymbol{z})\boldsymbol{H}_i^j, \quad i = 1, \ldots, n_u \\
\boldsymbol{c}(\boldsymbol{z}) &= \sum_{j=1}^{r} \boldsymbol{W}^j(\boldsymbol{z})\boldsymbol{c}^j.
\end{aligned} \tag{6.34}$$

where $\boldsymbol{W}^j$ is the diagonal weight matrix whose entries are the normalized degrees of fulfillment of the jth rule. In this approach the linear parameter-varying system obtained at each operating point tries to track the nonlinear process dynamics in a way similar to using a linear adaptive model to track nonlinear dynamics [148]. Moreover, for each operating point the model (6.33) can be put into state-space form as shown in Eq. (6.13) and (6.14).

Predictive Control

In order to predict the trajectory of the controlled outputs the LPV system (6.33) can be used, i.e., the set of models extracted along the future operating points $\{M(k+i)\}_{i=1}^{N_p}$ with $M(k) = \{\boldsymbol{A}_k, \boldsymbol{B}_k, \boldsymbol{C}_k, \boldsymbol{E}_k\}$ denoting the linear state-space model obtained at the kth step. As above, the linear extracted state-space models can be augmented to provide offset free control. Let $\boldsymbol{\xi}(k)$ denote the augmented state vector at time step k, then future process outputs are computed from [148]

$$\begin{aligned}
\hat{\boldsymbol{y}}(k+N_p|k) = \boldsymbol{\mathcal{C}}_{k+N_p} \prod_{j=N_p-1}^{0} \boldsymbol{\mathcal{A}}_{k+j}\boldsymbol{\xi}(k) + \boldsymbol{\mathcal{C}}_{k+N_p} \sum_{i=0}^{N_c-1} \prod_{j=N_p-1, j\geq i+1}^{i+1} \boldsymbol{\mathcal{A}}_{k+j}\boldsymbol{\mathcal{B}}_{k+i}\Delta\boldsymbol{u}(k+i) \\
+ \boldsymbol{\mathcal{C}}_{k+N_p} \sum_{i=0}^{N_p-1} \prod_{j=N_p-1, j\geq i+1}^{i+1} \boldsymbol{\mathcal{A}}_{k+j}\boldsymbol{\mathcal{E}}_{k+i}\Delta\boldsymbol{z}(k+i)
\end{aligned} \tag{6.35}$$

where $\prod_{j=N_p-1}^{0} \boldsymbol{\mathcal{A}}_{k+j}\boldsymbol{\xi}(k) = \boldsymbol{\mathcal{A}}_{k+N_p-1}\boldsymbol{\mathcal{A}}_{k+N_p-2}\cdots\boldsymbol{\mathcal{A}}_k$. Or in compact matrix form this can be written as (cp. (6.16))

$$\hat{\boldsymbol{Y}} = \boldsymbol{F}(k)\xi(k) + \boldsymbol{\phi}_u(k)\Delta\boldsymbol{U} + \boldsymbol{\phi}_z(k)\Delta\boldsymbol{Z} \tag{6.36}$$

with

$$\boldsymbol{F}(k) = \begin{bmatrix} \boldsymbol{\mathcal{C}}_{k+1}\boldsymbol{\mathcal{A}}_k \\ \boldsymbol{\mathcal{C}}_{k+2}\boldsymbol{\mathcal{A}}_{k+1}\boldsymbol{\mathcal{A}}_k \\ \boldsymbol{\mathcal{C}}_{k+3}\boldsymbol{\mathcal{A}}_{k+2}\boldsymbol{\mathcal{A}}_{k+1}\boldsymbol{\mathcal{A}}_k \\ \vdots \\ \boldsymbol{\mathcal{C}}_{k+N_p}\boldsymbol{\mathcal{A}}_{k+N_p-1}\cdots\boldsymbol{\mathcal{A}}_k \end{bmatrix}, \tag{6.37}$$

$$\boldsymbol{\Phi}_u(k) = \begin{bmatrix} \boldsymbol{\mathcal{C}}_{k+1}\boldsymbol{\mathcal{B}}_k & \mathbf{0} & \cdots & \mathbf{0} \\ \boldsymbol{\mathcal{C}}_{k+2}\boldsymbol{\mathcal{A}}_{k+1}\boldsymbol{\mathcal{B}}_k & \boldsymbol{\mathcal{C}}_{k+2}\boldsymbol{\mathcal{B}}_{k+1} & \cdots & \mathbf{0} \\ \boldsymbol{\mathcal{C}}_{k+3}\boldsymbol{\mathcal{A}}_{k+2}\boldsymbol{\mathcal{A}}_{k+1}\boldsymbol{\mathcal{B}}_k & \boldsymbol{\mathcal{C}}_{k+3}\boldsymbol{\mathcal{A}}_{k+2}\boldsymbol{\mathcal{B}}_{k+1} & \cdots & \vdots \\ \vdots & & \ddots & \\ \boldsymbol{\mathcal{C}}_{k+N_p}\prod_{j=N_p-1}^{1}\boldsymbol{\mathcal{A}}_{k+j}\boldsymbol{\mathcal{B}}_k & \cdots & & \boldsymbol{\mathcal{C}}_{k+N_p}\prod_{j=N_p-1}^{N_c}\boldsymbol{\mathcal{A}}_{k+j}\boldsymbol{\mathcal{B}}_{k+N_c-1} \end{bmatrix}, \tag{6.38}$$

$$\boldsymbol{\Phi}_z(k) = \begin{bmatrix} \boldsymbol{\mathcal{C}}_{k+1}\boldsymbol{\mathcal{E}}_k & \mathbf{0} & \cdots & \mathbf{0} \\ \boldsymbol{\mathcal{C}}_{k+2}\boldsymbol{\mathcal{A}}_{k+1}\boldsymbol{\mathcal{E}}_k & \boldsymbol{\mathcal{C}}_{k+2}\boldsymbol{\mathcal{E}}_{k+1} & \cdots & \mathbf{0} \\ \boldsymbol{\mathcal{C}}_{k+3}\boldsymbol{\mathcal{A}}_{k+2}\boldsymbol{\mathcal{A}}_{k+1}\boldsymbol{\mathcal{B}}_k & \boldsymbol{\mathcal{C}}_{k+3}\boldsymbol{\mathcal{A}}_{k+2}\boldsymbol{\mathcal{E}}_{k+1} & \cdots & \vdots \\ \vdots & & \ddots & \\ \boldsymbol{\mathcal{C}}_{k+N_p}\prod_{j=N_p-1}^{1}\boldsymbol{\mathcal{A}}_{k+j}\boldsymbol{\mathcal{E}}_k & \cdots & & \boldsymbol{\mathcal{C}}_{k+N_p}\boldsymbol{\mathcal{E}}_{k+N_p-1} \end{bmatrix} \tag{6.39}$$

These matrices have to be calculated at each time step and are then used in the quadratic program (6.25) for determing the optimal future control sequence.

Now, depending on how the calculation of $\{M(k+i)\}_{i=1}^{N_p}$ is done, two conceptual approaches exist: Non-iterative and iterative methods [150, 148, 196]. A representative of the former is the so called *Single-model method*. Here, as the name suggests, only a single linear state-space model $M(k)$ is extracted at the current operating point $\boldsymbol{z}(k) \in \{\boldsymbol{y}(k), \boldsymbol{u}(k)\}$ and used throughout the entire prediction horizon. Such approach is feasible for weakly nonlinear processes, or when operating close to certain equilibrium points during longer time periods. However, the prediction accuracy might significantly deteriorate for strongly nonlinear processes. A way to circumvent this is by applying iterative methods such as the *Multi-models method* [150]. This approach can be summarized in the following steps:

1. Use the obtained linear model $M(k)$ at the current operating point $\boldsymbol{z}(k)$ and compute the control signal $\boldsymbol{u}(k)$ for the whole control horizon
2. Simulate the TS fuzzy model over the prediction horizon
3. Freeze the TS Fuzzy model along each point in the predicted operating point trajectory $\boldsymbol{z}(k+i)$ and obtain $M(k+i)$ for $i = 1, \ldots, N_p$.
4. Use $M(k+i)$, $i = 1, \ldots, N_p$ to construct MPC matrices (6.37)-(6.39) and compute the new control sequence $\boldsymbol{u}(k)$

Steps 3 and 4 are repeated until $\boldsymbol{u}$ converges. Of course, if the operating points are independent of input and ouput terms such as in pure gain-scheduling applications, iterative methods are obsolete.

6.3.3 Internal Model Control

Disturbances acting on the process, plant-model mismatch, and measurement noise cause differences in the behavior of the process and of the model and can lead to an error between the reference trajectory and the process output [7]. Two ways to compensate for this are: Adaptive estimation as introduced in Sec. 3.2.2 and the internal model control (IMC) scheme [68, 176, 60].

In classical feedback control the current process output or an estimate from an observer is used at each time instant to update the controller state [207]. The direct update of the state of the controller, however, might not be desirable in the presence of measurement noise

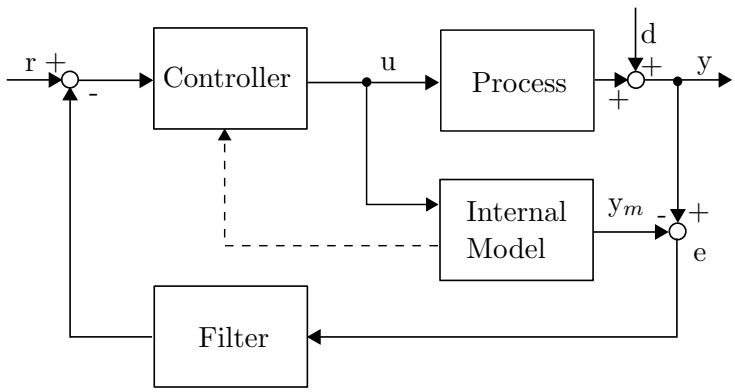


Figure 6.3: Internal model control (IMC) scheme

or significant plant-model mismatch. This could lead to oscillations or instability especially for highly nonlinear processes [7]. The IMC scheme, on the other hand, utilizes an internal model which runs in parallel with the process. The ouput of the internal model is then used to update the state of the controller instead of the measured process output. To eliminate modeling errors the difference between the outputs of the process and of the internal model is fed back. As illustrated in Fig. 6.3, the IMC scheme consists, in general, of three parts [68]: (1) An internal model (e.g., fuzzy model) to predict the process output, (2) a feedback filter to achieve robustness and (3) a controller (e.g., fuzzy MPC) to compute future values of the manipulated variables. If the predicted ouput from the internal model and the measured process output are equal, the error e is zero and the controller works in an open-loop setup. If a disturbance d is present, the error equals d and this influence is then subtracted from the original reference signal.

The two main properties of IMC are inherent stability and perfect control [68]. Inherent stability means that when the model is perfect, stability of both controller and plant is sufficient for close-loop stability. Perfect control means that if the controller is an exact inverse of the model, and the closed-loop is stable, then the control is error free. Although in practise, a perfect model does not exist and thus the feedback signal contains both the effect of unmeasured disturbances and of the modeling error. Note that for large modeling errors stability problems can even occur [7]. To stabilize the closed-loop system and to filter out measurement noise a feedback filter is introduced. A simple exponential filter was proposed in [68]. For nonlinear systems and models the design of alternate, more complex filters might be necessary in order to achieve sufficiently large robustness [7].

6.3.4 Stability

Although TS fuzzy models provide, in general, a good approximation of the process there will always be a certain plant-model mismatch present, which eventually can also destabilize the closed-loop system. Hence, tools for the design of robustly stable, constrained predictive controllers are of significant importance and therefore part of active research [146, 172]. Classical concepts for ensuring closed-loop system stability of nonlinear processes with constraints are [197, 114, 146, 172]:

- an additional set constraining terminal states
- a cost on the terminal state
- a stabilizing linear state-feedback controller, designed for a considered equilibrium point of the nonlinear process

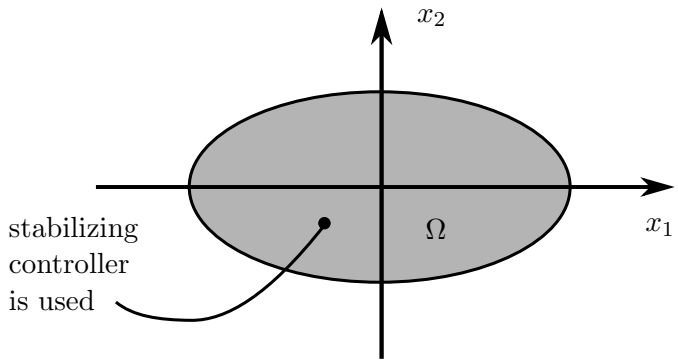


Figure 6.4: Idea of the dual-mode approach (reproduced from [145])

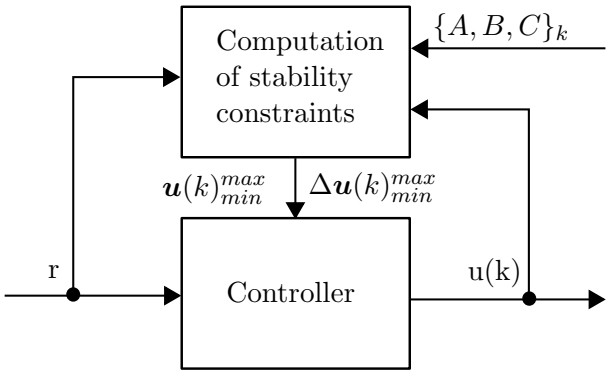


Figure 6.5: Sketch of a predictive controller with robust stability constraints (reproduced with modifications from [149])

In particular, for stabilizing fuzzy predictive controllers dual-mode strategies were proposed (e.g., [145, 14, 127]. The concept of this strategy is as follows: First, a constrained predictive contoller with a finite horizon should bring the state of the process, in a finite number of steps, to a (convex) target set or terminal region Ω (see Fig. 6.4). Then inside this set a second local stabilizing linear controller is used bringing the process to the desired equilibrium state. This also motivates the name dual-mode. The main disadvantage of this technique is that there is no guarantee that the controller drives the systems into the terminal region [115].

Robust stability techniques for fuzzy systems based on the small gain theory are proposed in [149]. Here constraints on the control input and its increments are calculated at each sampling instant which guarantees robust stability (see Fig. 6.5). The drawback of this method is that it can only be used for open loop bounded-input-bounded-output (BIBO) stable processes and that the added constraints lead to slower response.

Recently, a new concept for stability of fuzzy systems was introduced in [115]. The concept is based on two ideas: First, an added constraint on the applied control action is used to ensure the decrease of a quadratic Lyapunov function, and so guarantee Lyapunov exponential stability of the closed-loop system. Second, the feasibility of the finite-horizon optimization problem with added constraint is ensured based on an offline solution of a set of linear matrix inequalities (LMIs). For further details and a good overview about stability of unconstrained and constraind fuzzy MPC the reader is referred to [196].

Chapter 7

Model Predictive Control Implementations

7.1 Predictive Control of Supply Temperature

7.1.1 Problem Formulation

The primary goal of control scheme in district heating systems is to provide adequate thermal power to consumers while not violating any physical or technological constraints. A secondary goal is to achieve the primary goal in such way that the overall operational costs are minimized. Recall that the thermal power is proportional to the product of the mass flow rate and the temperature difference between supply and return line. Hence, district heating systems may be controlled by two different principles. Either the temperature T_s is kept constant while the flow q is varied in order to respond to the changing consumer heat demand, or the flow is fixed and the temperature varies. Due to the much faster response characteristic of the flow to changes in demand, the supply temperature is typically considered as the important variable to be manipulated [76] .

A possible approach for optimizing the control strategy is by minimizing the operational costs for time instant t: (cf. [21, 75])

$$\begin{aligned} \min_{\Delta T_s(t)} \; & J_{op}(t) = c_{elec} P_{pump}(t) + c_{fuel} Q_{loss}(t) \\ & \text{with respect to constraints on } T_s(t) \text{ and } q(t). \end{aligned} \tag{7.1}$$

Here Q_{loss} denotes the heat loss defined as the difference between produced and delievered heat power, P_{pump} is the pump power and c_{elec} and c_{fuel} are the corresponding costs per kWh. The pump power can be calculated from [18]

$$P_{pump}(t) = \frac{\rho g q(t) H(t)}{\eta_m \eta_p \eta_v} \tag{7.2}$$

where g is acceleration due to gravity, H is the pressure head and η_m, η_p and η_v are the motor, pump efficiency and the efficiency of the variable speed drive. Pressure heads and pump efficiencies for various flow conditions are provided by the characteristic curves which can be found in the manufacturer's data.

The objective function (7.1) reflects the trade-off between costs related to the supply temperature and costs related to keeping up the flow rates in the network. To understand the underlying problem of finding a solution to (7.1) one has to inspect the causalities with respect to a changes in supply temperature and volume flow rate on the operational costs. On the one hand, higher pump pressure results in higher flow rates in the network and, of course, increases

the pumping costs as can be seen in (7.2). However, on the other hand, higher flow rates mean less transport time delays and thus reduce the costs related to heat losses. Similiar arguments can be applied to the supply temperature. Hence, efficient control strategies have to take both parts of the costs and possible constraints into account. One possible strategy is introduced in this section.

7.1.2 Set-Point Strategies

Predictive control schemes require appropriate set-point or reference trajectories for the output variables. In district heating the set-point for the supply temperature is typically conservatively chosen in order to guarantee that the consumers are satisfied. Also due to technical reasons there is a minimal allowable supply temperature at heat exchangers which limits the supply temperature. Traditionally, the supply temperature is determined by so called heating curves or control curves [188, 84]. Such curves capture the general heating characteristics of consumer stations which naturally depend on the ambient temperature (see Fig. 7.1). A disadvantage of such static mappings is that operational costs such as pumping costs and heat loss costs are not explicitly taken into account. Nowadays, advanced control schemes are becoming more and more important [21]. For instance, the supply temperature could be dynamically controlled in such way that the return temperature has a certain level or that the overall operating costs are minimized. The latter is considered in this work by introducing a novel supervisory predictive on-line set-point optimizer. The idea follows the classical multilayer control approach [197, 144] in which the desired set-points are calculated in an optimization layer and then are applied to the model predictive controller (cf. Fig. 6.2 on 85).

Heating-Curves

In the context of predictive control of the supply temperature heating curves can be used for generating set-point values which define the minimal allowable supply temperature at the respective point in the network. More precisely, increasing supply temperature with decreasing ambient temperature reflects the limited capacity in the consumers space heating installations, whereas the minimum is determined by the hot tap water installations [159]. However, since the quadratic optimization problem (6.25) does not discriminate between realizations below or above the set-point, it is nessesary to add an extra safety margin to ensure that the actual supply temperatures do not fall far below the heating curve set-point values. The idea is that instead of directly imposing output constraints on the supply temperature which are given by the heating curves, the referece trajectory should by chosen in an exact way to not violate these contraints. First, assume that future values of the supply temperature at the consumers and the ambient temperature can be formulated as the sum of predictions and predictions errors

$$T_s^{\mathrm{AC}}(t+k) = \hat{T}_s^{\mathrm{AC}}(t+k|t) + \epsilon_{T_s}(t+k|t), \tag{7.3}$$

$$T_a(t+k) = \hat{T}_a(t+k|t) + \epsilon_{T_a}(t+k|t), \quad k = 1, \ldots, N_p \tag{7.4}$$

where the corresponding error terms are assumed to be mutually independent Gaussian distributed:

$$\begin{pmatrix} \epsilon_{T_s}(t+k|t) \\ \epsilon_{T_a}(t+k|t) \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \sigma^2_{T_s}(k) & 0 \\ 0 & \sigma^2_{T_a}(k) \end{bmatrix}\right). \tag{7.5}$$

The idea presented in [158] is to ensure that future supply temperature values belong to some admissible area Ω_2 for some predetermined probability p (see Fig. 7.1), i.e.,

$$\begin{aligned} \mathbb{P}\{(\hat{T}_s^{\mathrm{SP}}(t+k|t) + \epsilon_{T_s}(t+k|t), \hat{T}_a(t+k|t) + \epsilon_{T_a}(t+k|t)) \\ \in \Omega_2\} = p, \quad k = 1, \ldots, N_p \end{aligned} \tag{7.6}$$

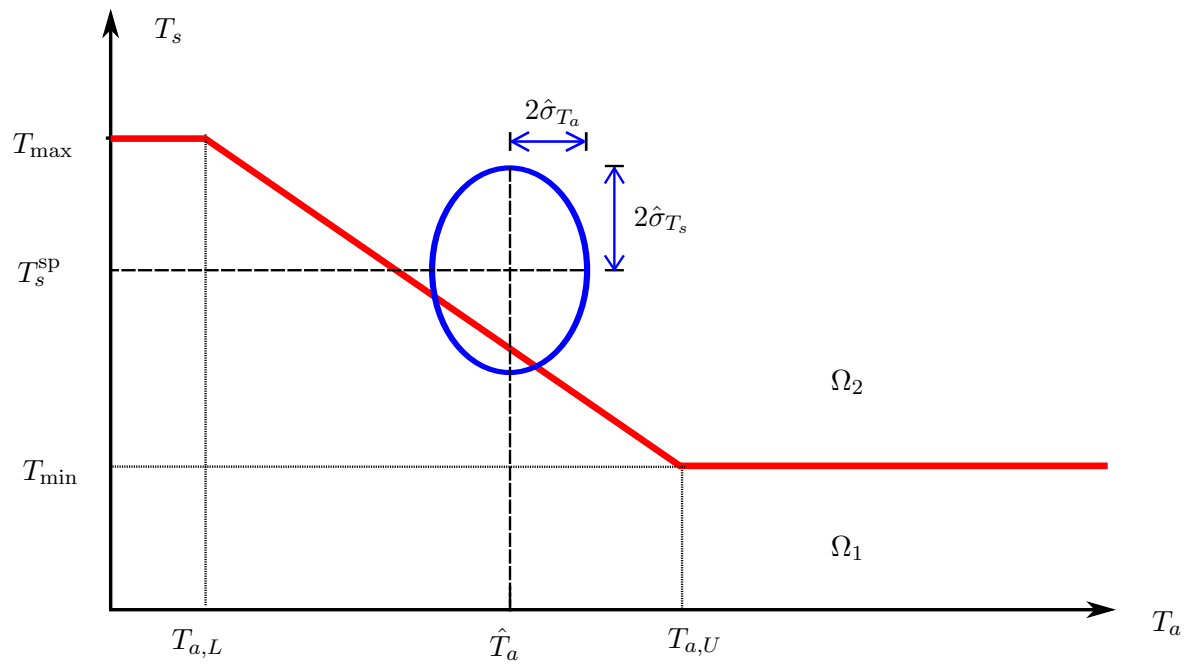


Figure 7.1: Schematic heating curve with characteristic break-points (reproduced with modifications from [159]).

Here the predicted temperature $\hat{T}_s^{\mathrm{AC}}(t+k|t)$ was substituted by the set-point value $\hat{T}_s^{\mathrm{SP}}(t+k|t)$, since it is assumed that the process is controlled according to the model used to obtain this predictor. Now Eq. (7.6) has to be solved with respect to $\hat{T}_s^{\mathrm{con}}(t+k|t),\ k=1,\ldots,N_p$ which can be achieved by a combination of Simpson integration and minimization algorithms. For details the reader is referred to [158].

Remark. *In [158] it was additionally proposed to filter the ambient temperature before constructing the relationship between heat load and ambient temperature. This was motivated due to the ability to store heat in the mass of buildings.*

Minimize Operational Costs

The control strategy based on heating curves, although being widely used, might not be optimal in terms of minimizing operational costs. As discussed previously the major sources of operational costs in district heating networks are heat loss costs and pumping costs. Heat loss costs can be reduced by lowering the supply temperature from the plant and by increasing the thermal insulation thickness in pipes [39]. This, however, results in higher pumping cost due to increased flow rates in the network assuming that the delivered heat to the consumers and the return temperature is kept constant. Due to this trade-off, choosing the optimal supply temperature is a complicated task and efficient strategies are highly desirable.

In the following one possible approach is described which tries to find optimal set-points (cp. [144, 197]) for the critial point considered by solving a (stochastic) nonlinear optimization problem at each time instant. First, the assumptions as listed below are made (cp. Fig. 7.2)

- The entire network is reduced to one supply pipe to the aggregated consumer and one return pipe from the aggregated consumer
- Return temperature T_r^{P} at the plant is known and system heat load Q_{P} can be predicted

- Heat losses in the supply pipe can be approximated using the stationary gains estimated by the fuzzy models
- Supply pipe and return pipe have the same stationary gains

Under these assumptions the entire stationary heat loss can be expressed as the sum of the heat losses of the supply and return pipe as follows ([29])

$$Q_{loss}(t) = Q_{loss,s}(t) + Q_{loss,r}(t) \tag{7.7}$$

with

$$\begin{aligned} Q_{loss,s}(t) &= c_p \rho q(t)[1 - K(q(t))]T_s^{\mathrm{P}}(t) \\ Q_{loss,r}(t) &= c_p \rho q(t)[1/K(q(t)) - 1]T_r^{\mathrm{P}}(t) \end{aligned} \tag{7.8}$$

and where $0 < K(\cdot) \leq 1$ represents the stationary gain of the cFIR/cARX model at flow rate $q(t)$ (cp. Fig. 3.19 on page 38). It is important to ensure that $K(\cdot)$ is monotonically increasing and reasonably smooth.

The optimization problem itself can be formulated as

$$\begin{aligned} \min_{\Delta \boldsymbol{T}_s^{\mathrm{P}}} J(\boldsymbol{T}_s^{\mathrm{P}}, \boldsymbol{T}_s^{\mathrm{SP}}, \boldsymbol{q}) = \quad & \mathbb{E}\left\{ \sum_{k=1}^{N_c} c_{fuel} Q_{loss}(t+k) + c_{elec} P_{pump}(t+k) \right\} \\ \text{s. t.} \quad & c_p \rho q(t+k)[T_s^{\mathrm{P}}(t+k) - T_r^{\mathrm{P}}(t+k)] = \hat{Q}_{\mathrm{P}}(t+k|t) \\ & \mathbb{P}\{q(t+k) \leq q^{\max}\} \geq p \\ & \mathbb{P}\{q(t+k) \geq q^{\min}\} \geq p \\ & T_{s,1}^{\min} \leq T_s(t+k) \leq T_{s,1}^{\max} \\ & T_{s,2}^{\min} \leq T_s^{\mathrm{SP}}(t+k) \leq T_{s,2}^{\max} \\ & \boldsymbol{T_s}^{\mathrm{SP}} = \boldsymbol{F}(q)\boldsymbol{\xi}(t) + \boldsymbol{\Phi}(q)\Delta \boldsymbol{T}_s^{\mathrm{P}} \end{aligned} \tag{7.9}$$

where $\boldsymbol{T}_s$, $\boldsymbol{T}_s^{\mathrm{SP}}$ and $\boldsymbol{q}$ are vectors of plant supply temperature, set-point supply temperature at critical nodes, and volume flow rates, respectively. The matrices $\boldsymbol{F}$ and $\boldsymbol{Q}$ are related to the forced and free response and are given in Eq. (6.37). The first constraint in (7.9) links the volume flow rate in the network to the predicted heat-load, supply temperature, and return temperature at the plant. The other constraints are related to input and ouput restrictions, and upper and lower bounds on the volume flow rate. The latter has to be formulated in a probabilistic manner, due the uncertainty in the heat load predictions. In particular, the minimum flow rate must be ensured in order to avoid rapid evaporation phenomena, possible malfunctions and pipe restrictions [202].

The return temperature at the plant is assumed to be known at time instance t and is further assumed to be constant over the prediction horizon, since in general it varies only smoothly (cp. Fig. 4.13 on 62).

To reduce the computational complexity of (7.9) two steps were taken. First, it was decided to disregard the ouput and input constraints, which could alternatively be implemented in a more efficient way in the MPC control problem, and second, the (nonlinear) constraints on the volume flow rate is reformulated by using penalty function methods [70]. Note that a deterministic counterpart to the flow contraints can be written as follows [20, 158]

$$\begin{aligned} q(t+k) &\leq q^{\max} - \frac{\sigma(t+k|t)^2}{c_p \rho[T_s^{\mathrm{P}}(t+k) - T_r^{\mathrm{P}}(t+k)]}\Phi(p), \\ q(t+k) &\geq q^{\min} + \underbrace{\frac{\sigma(t+k|t)^2}{c_p \rho[T_s^{\mathrm{P}}(t+k) - T_r^{\mathrm{P}}(t+k)]}\Phi(p)}_{\text{saefty term}} \quad k = 1, \ldots, N_c \end{aligned} \tag{7.10}$$

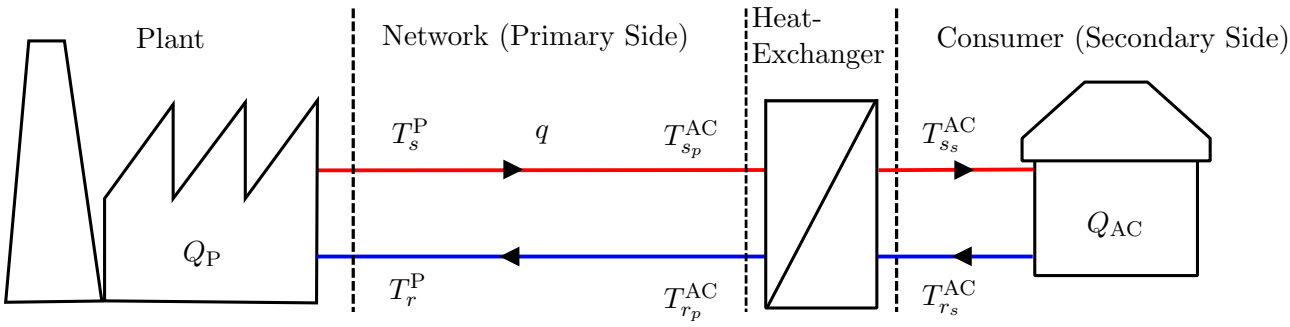


Figure 7.2: Simplified district heating network with one consumer

where $\Phi(p)$ is the p-quantile of the standard normal distribution. Thus the stochastic nonlinear optimization problem with constraints was reduced to a deterministic unconstrained problem. A minimal allowable flow rate of $5 \cdot 10^{-3}\,\mathrm{m}^3/\mathrm{s}$ and a maximal allowable flow rate of $25 \cdot 10^{-3}\,\mathrm{m}^3/\mathrm{s}$ are used in the simulations.

7.1.3 Controller Setup

The setup of the fuzzy model predictive controller used in this work is depicted in Fig. 7.3. It was decided to use the internal model control scheme as well as constrained WRLS to handle plant-model mismatch and disturbances. The control sequence can be summarized in the following steps:

1. If process data is available the estimates of the coefficient functions are updated via constrained WRLS.
2. The fuzzy model (cARX/cFIR) is linearized around the current operating point, then the internal state-space model and the MPC matrices are updated.
3. The reference trajectory for the supply temperature at the critical node is constructed using heating curves or set-point optimization.
4. In case of multi-model mode Algorithm 1 on page 102 is used to compute a sequence of future supply temperature increments otherwise (i.e., SM - mode), the classical MPC problem is solved. In both cases the first element of the optimal sequence is then applied to the plant.

Note that in the internal model control scheme the state vector of the (fuzzy) MPC is updated using the output of the internal model only, and the internal model in turn has its own internal state vector. To account for disturbances and plant-model mismatch the difference between the outputs of the internal model and the process model is subtracted from the generated reference trajectory. The fuzzy model itself is given by the conditional ARX and conditional FIR model developed in Sec. 3.2. The trigonometric offset was packed into the input disturbance matrix $\boldsymbol{E}$.

In addition to the standard recursive least squares procedure, constraints on the stationary gain were implemented. Due to heat losses in the distribution network the stationary gain is naturally not allowed to exceed one (cp. Fig. 3.19). The finite-difference models construced in Sec. 3.1 were taken as plant process model. Here again the data was up-sampled for the plant model using zero-order-hold, since the MPC operates in 30 min time frames and the finite-difference model was time discretized with 1 min. Furthermore, the return temperatures at the consumer stations are given by the estimated nonlinear, static mappings as illustrated in Fig. 4.13 on page 62.

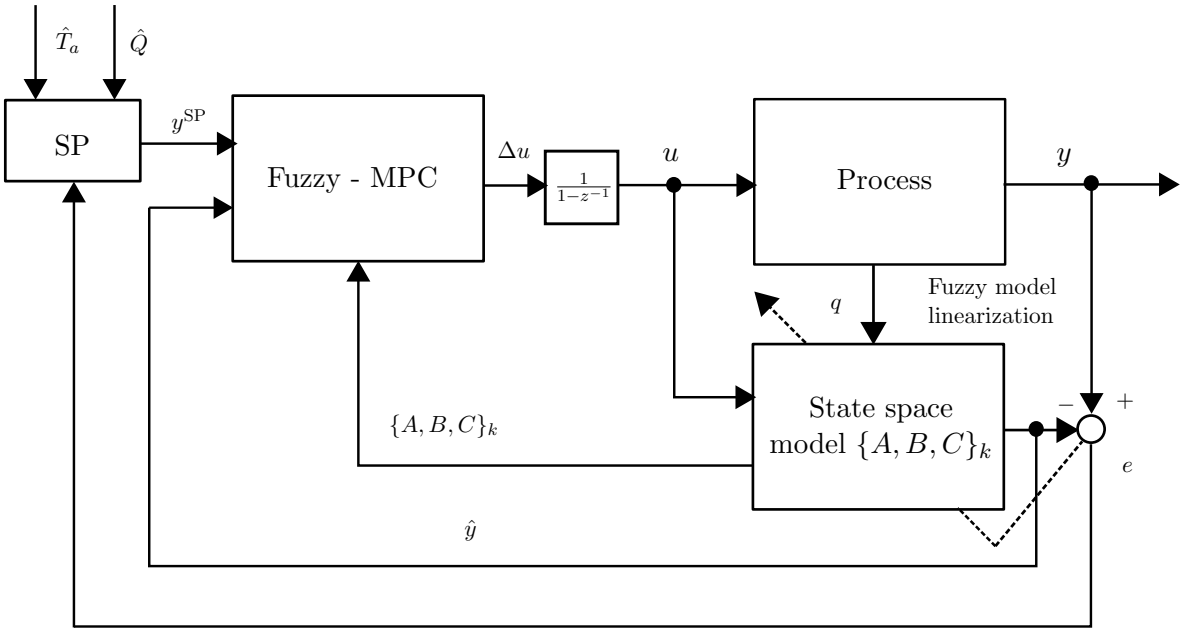


Figure 7.3: Block diagram of the Fuzzy MPC (SP: set point generation/optimization)

Algorithm 1 incorporates the forecast of the heat load into the FMPC control scheme. The idea is to iteratively update the volume flow rate based on future heat load values and optimized supply temperature values. Since the return temperature only varies smoothly, it was assumed that it can be held constant over the prediction horizon. Further it was assumed that $T_s(t+k|t) = T_s(t+N_c|t)$ for $k = N_c + 1 \ldots N_p - 1$. Experiments have shown that the algorithm converges quite quickly in practice, so $\max_{it} = 5$ was a reasonable choice.

Due to safety reasons, two constraints on the plant supply temperature are imposed:

- The rate of change for the supply temperature is restricted to 2°C per time instant (30 min).
- Minimum and maximum supply temperature values are set to 75°C and 105°C, respectively.

Algorithm 1 Fuzzy MPC multi-model iterative scheme.

1: Initialize $T_s^{(1)}(t+k|t)$ and $T_r^{(1)}(t+k|t),\ k = 0 \ldots N_p - 1$
2: Predict heat load $\hat{Q}(t+k|t),\ k = 1 \ldots N_p - 1$
3: **repeat**
4: Calculate current and future volume flow rates
$\hat{q}^{(i)}(t+k|t) = \frac{\hat{Q}(t+k|t)}{c_p \rho (T_s^{(i)}(t+k|t) - T_r^{(i)}(t+k|t))}\ ,\ k = 0 \ldots N_p - 1$
5: Use FMPC to determine future supply temperature
$T_s^{(i+1)}(t+k|t) = f_{\mathrm{FMPC}}(\hat{q}^{(i)}),\ k = 1 \ldots N_c$
6: **until** $||T_s^{(i+1)} - T_s^{(i)}|| <$ tol or #Iteration $\geq \max_{it}$

Although for the simulations considered, the constraints were hardly ever active. In fact, if the reference tajectory is chosen appropriately and the tuning of the weighting matrices is done well, the role of constraints is less important during normal operations. For the prediction horizon $N_p = 16$ (8 h) and control horizon $N = 6$ (3 h) were found to be adequate. With this choice

of prediction horizon the crucial process dynamics could be captured. The weight for penalizing the control effort is chosen constant as $R = 10$. In [75] it was found that higher penalties on the control increments for lower flow rates result in a significant performance increase. Hence, different weights for different flow conditions may be considered. For the output weights the concept of active critical node is introduced. Here positive weights are imposed (e.g., $Q = 1$) for the active critical node and all other outputs are weighted with zero weights (i.e., $Q = 0$). This means that practically only the temperature at the active critical node is controlled.

A detailed stability analysis is not done in this work. No stability problems were encountered during the simulations. For open-loop BIBO stable processes robust stability can always be guaranteed by choosing appropriate time-varying constraints on the input and input increments [149].

7.1.4 Simulation Results

Set-Point Tracking and Disturbance Rejection. First, to motivate the use of fuzzy methods, a convential MPC is compared with the fuzzy MPC for both the cARX and cFIR as underlying process model. For the design of the MPC the operating point $q = 0.008\, m^3/s$ was chosen. As performance measures for the tracking ability and the disturbance rejection the integrated absolute error (IAE):

$$IAE = \sum_{t=1}^{N} |y(t) - y^{\mathrm{SP}}(t)| \tag{7.11}$$

and the integrated control effort (ICE):

$$ICE = \sum_{t=1}^{N} |\Delta u(t)| \tag{7.12}$$

were taken. The reference signal is represented by the step signal as illustrated in Fig. 7.4. The disturbances are modeled as simultaneous heat demand jumps of all aggregated consumers. More precisely, first all consumers start with constant low heat demand and then at certain time steps the heat demand is increased by about 50%. This test procedure is repeated for all three critical aggregated consumers where each time one was considered to be the active critical point. The performance results are summarized in Table 7.1. It can be seen that the FMPC clearly outperforms the conventional MPC for both the cARX and cFIR model. The maximum performance increase is achieved for AC5 with a 18.2% reduction in terms of ICA (7.1% ICE) for the cARX process model and 32.7% IAE reduction (43.1% ICE) for the cARX process model. This can be explained mainly by the poor performance of the conventional MPC for very low flow rates as depicted in Fig. 7.4. Thus the MPC has problems in handling large transport time delays, whereas the FMPC shows good performance over the entire operating envelope.

In addition, the stepwise volume flow profile plus corresponding membership functions are shown in Fig. 7.5. Obviously, if consumer supply temperature is kept constant, the required heat demand can only be accomplished by higher flow rates and thus less heat loss in the network. Reduced heat losses also allow the supply temperature at the plant to be decreased as seen in Fig. 7.4.

Heating Curve Set-Points. To show the effects of utilizing heating curves for set-point genereration ten days of real measurements of heat-load and ambient temperature between 29.01.2009 and 07.02.2009 (Period I) and between 30.06.2009 and 09.Jul.2009 (Period II) are taken (see Fig. 7.6). In general, it can be assumed that in warmer months the heat losses are disproportionately high due to lower heat demand and consequently lower flow rates at constant

Table 7.1: Performance index comparison of set point tracking under MPC and fuzzy MPC

		cARX-(F)MPC		**cFIR-(F)MPC**	
	Error	MPC	FMPC	MPC	FMPC
AC3	IAE	164.6161	150.0188	166.61	163.02
	ICE	38.5639	34.4808	38.90	32.67
AC5	IAE	278.5856	226.71	346.94	233.47
	ICE	48.1366	44.74	71.89	40.94
AC7	IAE	111.5350	95.6119	106.30	96.38
	ICE	25.0531	23.0911	25.14	22.14

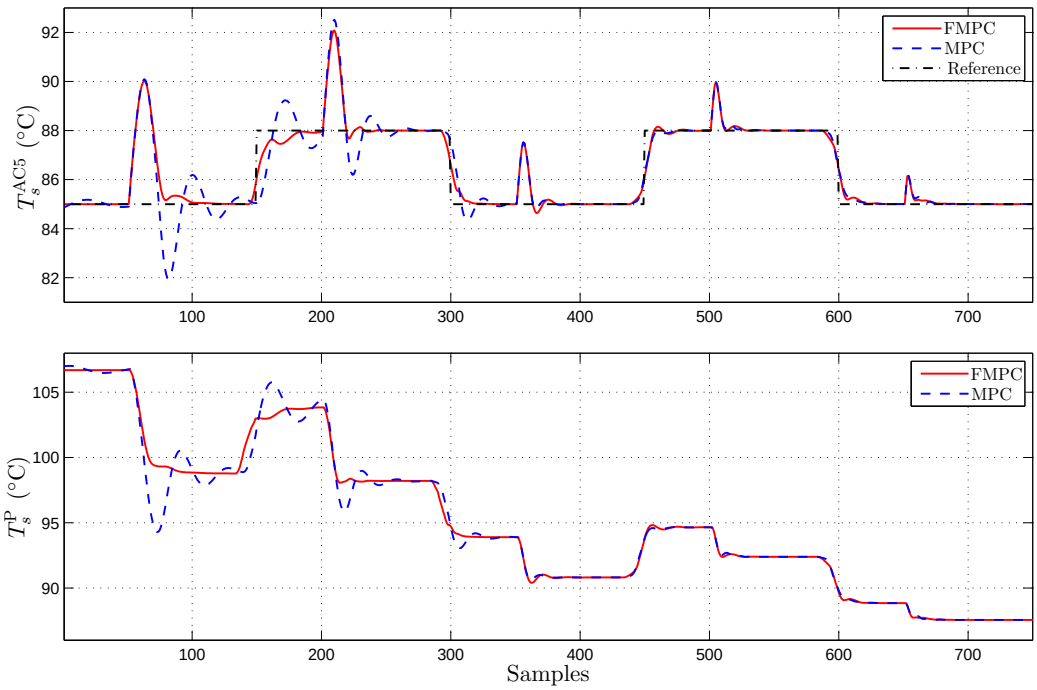


Figure 7.4: Performance comparison of set point tracking and step disturbance rejection under cFIR-MPC and cFIR-FMPC with AC5 being active

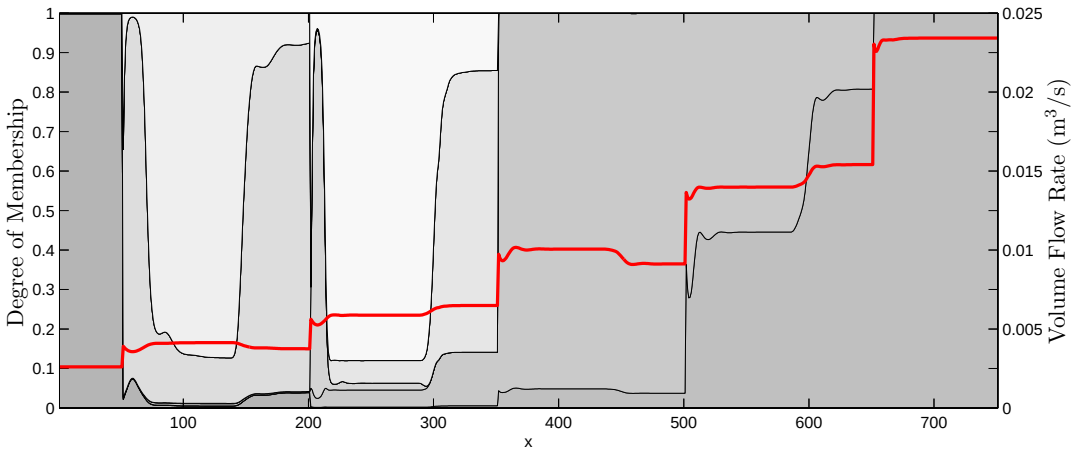


Figure 7.5: Stepwise volume flow profile plus corresponding memberhip function values of the local models for the cFIR-FMPC with AC5 being active; the thick full line represents the volume flow rate; the gray shaded areas reflect the individual local models

Table 7.2: Control Curve Specification in °C

	$T_{\max}$	$T_{\min}$	$T_{a,L}$	$T_{a,U}$
AC3 (CP1)	83.6169	55.9583	0.4324	15.4962
AC5 (CP2)	83.0846	61.9933	-1.5288	11.2344
AC7 (CP3)	85.3809	83.4613	4.7425	9.3541

supply temperatures. For constructing the individual heating curves for the critical points the steps discussed in [159] are being followed:

1. The ambient temperature observations are divided into a number of sub-intervalls
2. After grouping the obserations of supply temperature and ambient temperature into these above defined sub-intervalls, the quantile of the supply temperture distribution corresponding to $1-p$ is established for each intervall
3. Finally, the heating curves are fitted to the observed quantiles

Note that the number of observations in each sub-intervall should be large enough to provide reliable estimates. The heating curves as shown in Fig. 7.1 can be written in functional form as

$$T_s^{\mathrm{SP}} = \beta_0 + \beta_1 \left[(T_a - \zeta_1)_+ - (T_a - \zeta_2)_+ \right]. \tag{7.13}$$

where the parameters β_0, β_1 and the knots ζ_1, ζ_2 were estimated by nonlinear least-squares techniques (`lsqnonlin` in MATLAB®). The estimation results for the three cricital nodes AC3, AC5 and AC7 are presented in Table 7.2. Here the 10%-quantiles were chosen, i.e., 10% of the observations lie below the fitted heating curves.

The fitted heating curve for AC5 is depicted in Fig. 7.7 for illustration purposes. One can clearly identify the characteristic knots in the scatterplot which were also found by the nonlinear least-squares estimation approach.

Simulation results for Period I are shown in Fig. 7.8 and for Period II in Fig. 7.9. In both cases the FMPC with cARX process model is considered. Further it was decided to increase the

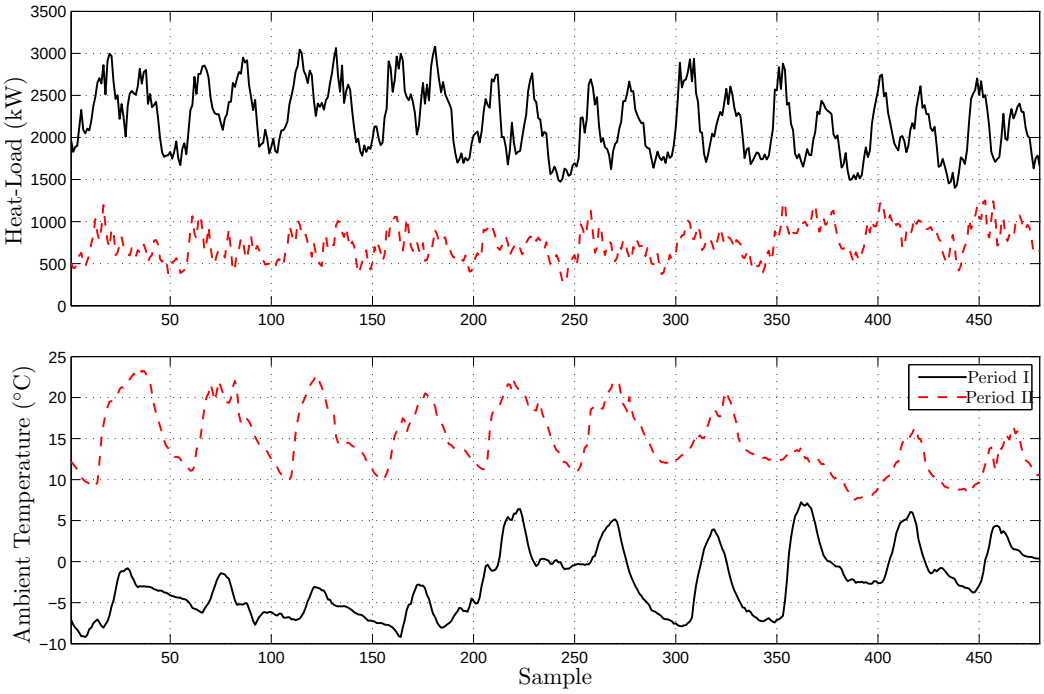


Figure 7.6: System heat load and ambient temperature profile for the two time periods considered; sampling time is 30 min

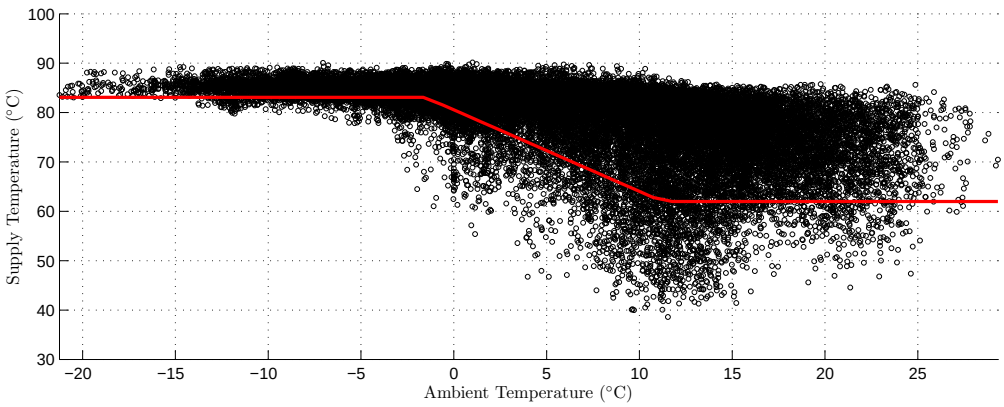


Figure 7.7: Fitted heating curve for AC5

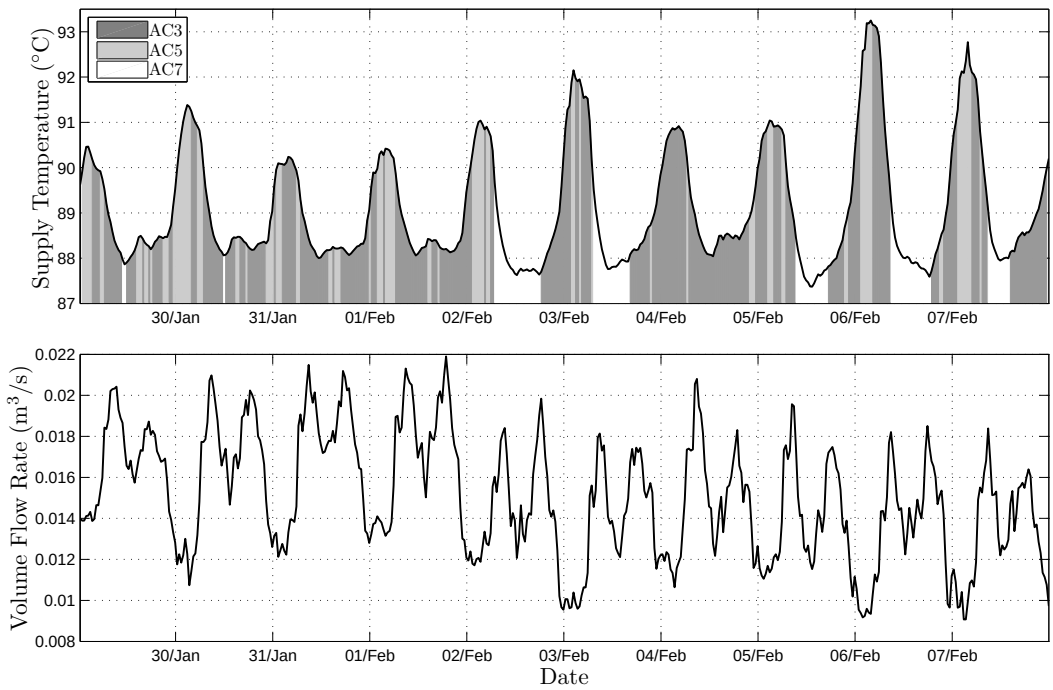


Figure 7.8: Simulation results for Period I using heating curves. The shaded areas correspond to the currently active critical point

weight of penalizing the control effort to $Q = 20$ for Period II where flow rates in the network are low.

The active critical point is determinded as follows: Predict future output at each critical point using Eq. (6.36) on page 93. Then build the differences between future reference trajectory and predicted output. The active critical node is the one where the biggest differences can be observed. In both plots the typical diurnal variation in the volume flow rate can be seen. The peaks in the supply temperature during night are due to the static mappings between supply- and ambient temperature of the heating curves. From an operational point of view this might not be optimal, since high temperatures are imposed on the network when heat-load is low.

Set-Point Optimization (SPO). In the following the performance of the primary supply temperature based control strategy using heating curves versus set-point optimization is evaluated. The cost function (7.1) for various supply temperature values is shown in Fig. 7.10. The figure on the left side assumes system heat-load of 2 MW and on the right a system heat-load of 3 MW is assumed. Hence, for higher heat demand the minimum of the cost function moves in the direction of higher supply temperatures. Note that for the electricity price 0.15 €/kWh was used and for biomass fuel 0.04 €/ kWh. In Tannheim three pumps of type *CR45-7 Grundfos* [81] (30kW) are installed in parallel. The speed of each pump is controlled through the use of a frequency converter. Pressure heads and flow rates for different pump operating points can be derived from affinity laws [212].

The simulation results plotted in Fig. 7.11 clearly show that the controlled supply temperature at AC3 with set-points obtained from SPO follows the diurnal variation of the heat-load. This means that peaks are produced in the morning hours and late afternoon and low supply temperature is imposed on the network during night. This stands in contrast to the results obtained when using static heating curves for set-point construction. From an operational point

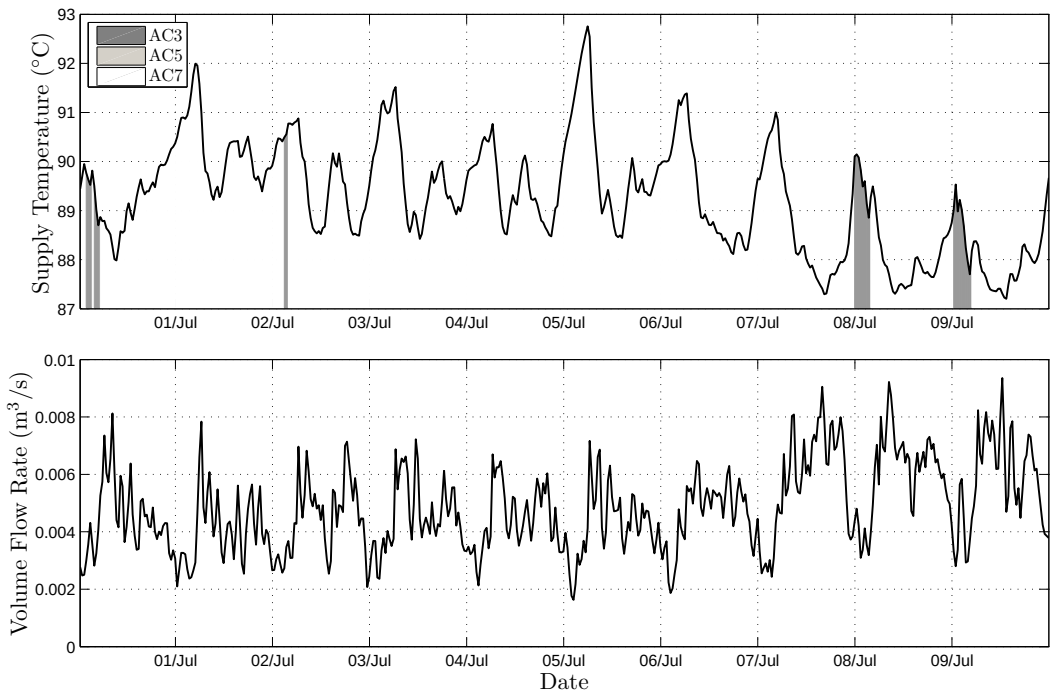


Figure 7.9: Simulation results for Period I using heating curves. The shaded areas correspond to the currently active critical point

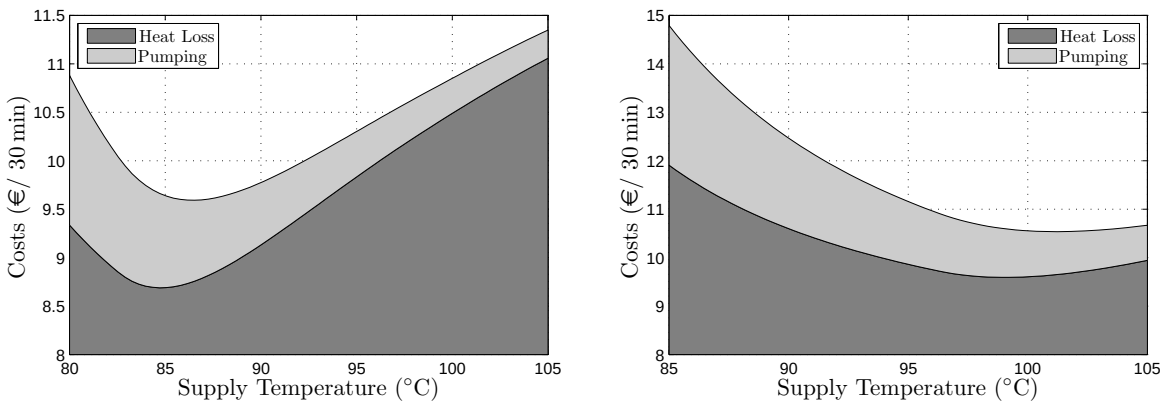


Figure 7.10: Heat loss costs and pumping costs for $Q_{\mathrm{P}} = 2\,\mathrm{MW}$ (left) and $Q_{\mathrm{P}} = 3\,\mathrm{MW}$ (right) using stationary gain curve from AC3; return temperature at plant is fixed to 55 °C

Table 7.3: Daily operating costs in € for set-points generated by heating curve and by set-point optimization, respectively

Costs	Heating Curve	SPO
Heat Loss	449.58	433.62
Pumping	32.72	32.10
Overall	482.30	465.72

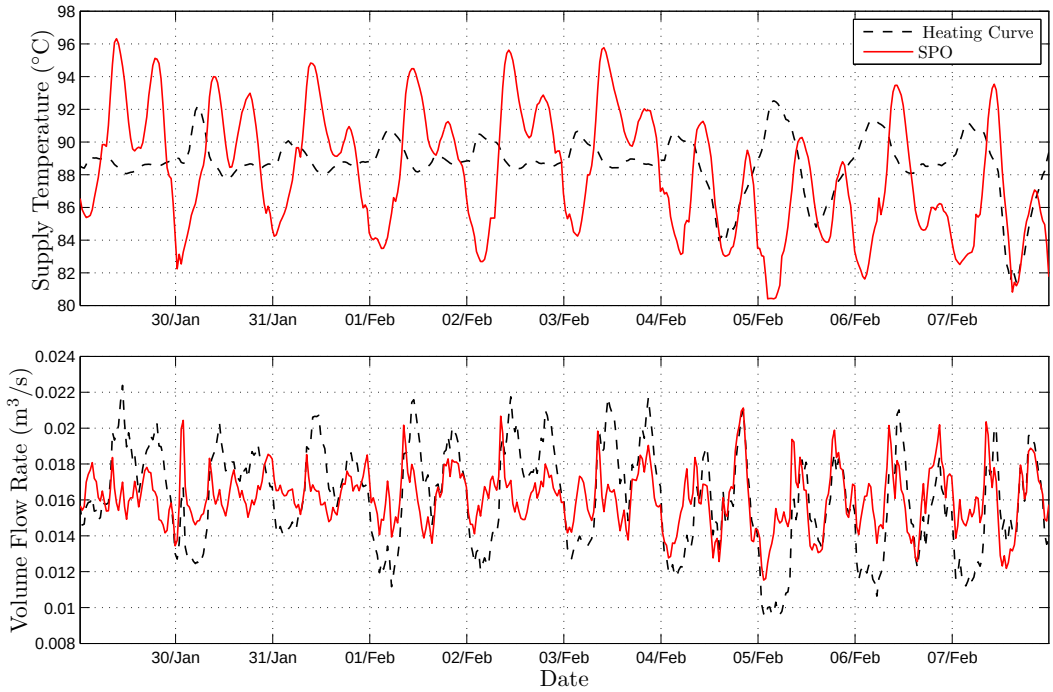


Figure 7.11: Simulation results for AC3 with set-points generated by heating curve and by set-point optimization, respectively

of view the heating curves are not optimal since the overall costs can be reduced by 3.44% when using the SPO strategy.

7.2 Predictive Control of Moving Grate Biomass Furnace

7.2.1 Problem Formulation

Although present-day grate-fired biomass boilers for heat and power production are usually well designed and sufficiently equipped with monitoring systems, the use of modern control strategies is still not very widespread [163]. Existing automation systems are widely based on individual SISO-control loops with rudimentary decoupling features and hence require substantial operator interaction [164]. The optimization in operation using advanced control concepts can greatly increase efficiency, lower emissions (e.g., the pollutants as a result of incomplete combustion), and mitigate other related problems [218, 119].

Model predictive control (MPC) [28] and its variants such as nonlinear model predictive control (NMPC) [3] or fuzzy model predictive control (FMPC) [7] are advanced control schemes especially suited for (nonlinear) multivariate process applications with constraints [185].

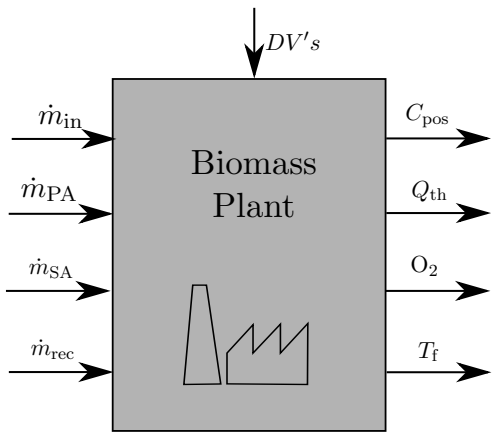


Figure 7.12: Overview of input- and output variables of the biomass plant model

In this section a fuzzy controller for the biomass plant is designed which is based on the fuzzy model constructed and identified in Sec. 5.2.

7.2.2 Process Description

A detailed process description of the considered moving grate biomass furnace is presented in Sec. 5.1. Figure 7.12 again summarizes the input and output variables of the MIMO fuzzy model. Hence, the relevant manipulated variables of the fuzzy MPC are (cf. Sec. 5.1)

1. Biomass input feed ($\dot{m}_{\text{in}}$)
2. Mass flow rate of the primary air supply ($\dot{m}_{\text{PA}}$)
3. Mass flow rate of the secondary air supply ($\dot{m}_{\text{PA}}$)
4. Mass flow rate of recycle gas, ($\dot{m}_{\text{rec}}$)

and the important output variables are

1. Combustion position (C_{pos})
2. Heat output (thermal power) (Q_{th})
3. Oxygen concentration (O_2)
4. Flue gas temperature (T_{f})

The heat output and flue gas temperature are in general considered to be the main output variables. However, for safe and secure operation it is also crucial to position the main fuel combustion approximately at the center of the grate [164]. Thus, there is need to actively control the combustion by the distribution of the primary air supply. Similar arguments hold for the oxygen concentration. Since incomplete combustion can lead to high emissions of unburnt pollutants such as Carbon monoxide [160], control of the oxygen concentration can effectively increase the efficiency of the combustion plant [119]. Moreover, disturbance variables affecting the combustion process are, for instance, the moisture content of biomass fuels or the varying grate speed. The latter is also influenced by the heat load needed to satisfy the varying consumer heat demand.

The process dynamics of the biomass plant fuzzy model for a specific operating point is depicted in Fig. 7.13. Recall that the partitioning variables of the MISO fuzzy models introduced

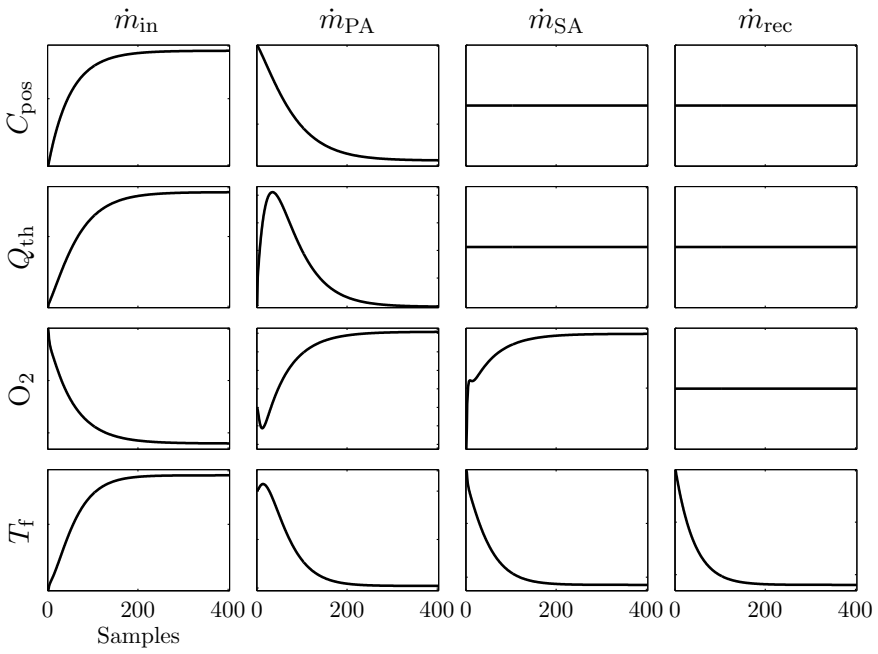


Figure 7.13: Step responses of the MIMO fuzzy model for operating point ($Q_{\mathrm{th}} = 2.5\,\mathrm{MW}$, $T_{\mathrm{f}} = 900\,^{\circ}\mathrm{C}$) with a sampling time of 5 s

in Sec. 5.2 are the heat output and the flue gas temperature. As one can see, the overall heat output is mainly influenced by the level of the input feed. The primary air supply has no net effect on the heat output because the amount of fuel burning remains constant. The combustion position depends positively on the input feed due to the fixed relationship between input feed and grate velocity and it depends negatively on the level of the primary supply. The latter can be explained by the fact that increasing the oxygen concentration results in a better combustion. Thus, the combustion position of the biomass bed on the grate is moved towards the fuel feed. Furthermore, primary and secondary air have a positive static gain on oxygen because more oxygen is provided for the same amount of fuel to be combusted. On the other hand, increased input feed has a negative effect on the oxygen concentration because more fuel will be combusted with the same amount of air, the resulting oxygen concentration in the flue gas will drop (cf. [119, p. 162]. For the flue gas temperature the exact opposite behavior can be observed. The input feed has a positive effect on the flue gas temperature because the amount of fuel burning increases. On the contrary, primary air, secondary air, and recycle gas have a negative static gain on the flue gas temperature due to the cooling of the flue gas.

7.2.3 Controller Specification

For control of the biomass furnace a setup similar to the control of the supply temperature in the network (see Sec. 7.1) is used. More precisely, the internal model control scheme (IMC) as illustrated in Fig. 7.14 is applied. Here, in each sample time, first the MIMO fuzzy model of the moving grate biomass furnace is linearized around the current operating point (heat output and flue gas temperature) and then the internal model and fuzzy MPC matrices plus states are updated. The first-principles model developed in Sec. 5.1 represents the real process. For the feedback filter a simple exponential filter with a filter parameter of 0.95 for all outputs was chosen. Other values for the relevant tuning parameters are:

- prediction horizon $N_p = 160$

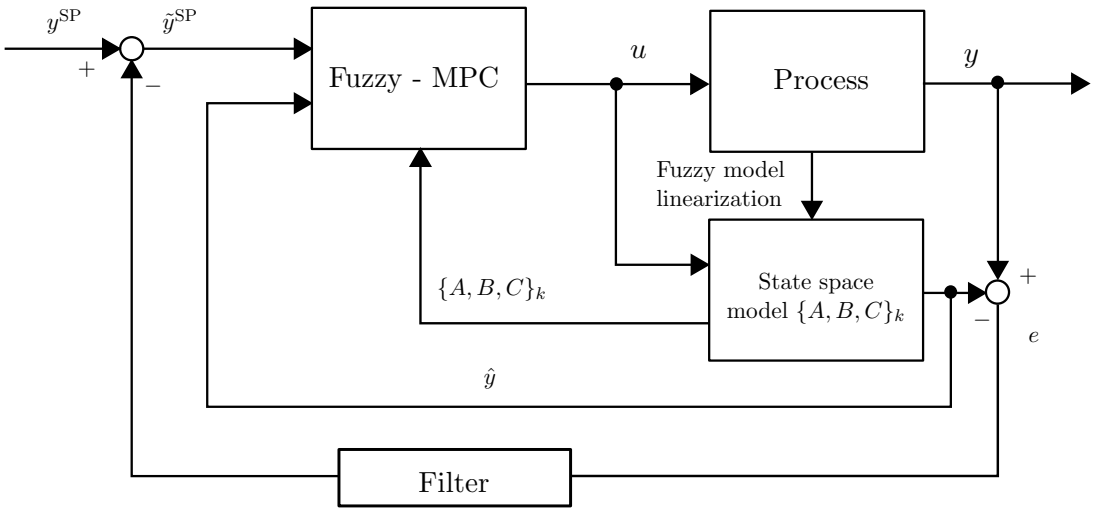


Figure 7.14: Block diagram of the fuzzy model predictive control scheme

- control horizon $N_c = 20$
- weights reference error $\boldsymbol{Q} = \mathrm{diag}\begin{bmatrix}8 \cdot 10^{-4} & 7 \cdot 10^{-1} & 3 \cdot 10^{-2} & 3 \cdot 10^{-5}\end{bmatrix}$
- weights control increments $\boldsymbol{R} = \mathrm{diag}\begin{bmatrix}200 & 1 & 1 & 0.1\end{bmatrix}$
- constraints control increments: $\Delta\dot{m}_{\mathrm{in}}$: 0 - 0.5 kg/s; $\Delta\dot{m}_{\mathrm{PA}}$, $\Delta\dot{m}_{\mathrm{SA}}$, and $\Delta\dot{m}_{\mathrm{rec}}$: 0% - 100%

Note that the input variables primary supply air, secondary supply air, and recycle gas are given in percent of total fan speed. Furthermore, the prediction horizon was chosen to be equal to about 90% of the maximal rise time of the system (cp. Fig. 7.13). Moreover, for calculating the forced and free response the non-iterative *Single-model method* was used. This means that the state-space triplet obtained from the current linearized MIMO fuzzy model is used throughout the entire prediction horizon. This was found necessary due to the long prediction horizon. Here iterative methods would significantly increase the computational burden.

7.2.4 Simulation Results

Fuzzy MPC

To study the closed-loop behavior of the fuzzy controller a set-point change in the heat output and the flue gas temperature was considered. Both the combustion position, and the oxygen concentration were kept constant at the center of the grate (i.e., 50%), and at 6%, respectively. The simulation results for the output variables are depicted in Fig. 7.15 and for the manipulated variables in Fig. 7.16. The fuzzy controller was compared with a conventional MPC with state-space triplet obtained at the operating point: $Q_{\mathrm{th}} = 2.5\,\mathrm{MW}$, $T_{\mathrm{f}} = 900\,^{\circ}\mathrm{C}$. Since from a visual point of view only minor differences between fuzzy MPC and conventional MPC are observable, the integrated absolute errors (IAE) are summarized in Tab. 7.4. The fuzzy controller performs slightly better except for flue gas temperature where practically no difference exists. One reason for that is the coarse partitioning of the fuzzy domain and hence low number of local models which leads to a similar response characteristics for a wide operating range. It is further worth analyzing the simulation results for the manipulated variables presented in Fig. 7.16. For instance, the input feed, being proportional to the heat output, decreases. On the other hand, the MPC uses the primary supply air to keep the combustion position at 50% and the secondary supply air to keep the oxygen concentration at the 6% level. The amount of recycle gas controls

Table 7.4: Integrated control error of set-point tracking under MPC and fuzzy MPC

	FMPC	MPC
Combustion position	808.24	830.27
Heat output	50.11	56.33
Oxygen concentration	143.46	146.49
Flue gas temperature	$4.986 \cdot 10^3$	$4.967 \cdot 10^3$

the flue gas temperature and has to be increased at the last quarter of the simulation horizon in order to trigger the flue gas temperature drop.

Reduced Fuzzy MPC

The simulations above are time consuming due to the long control horizon and the large number of manipulated variables. Each time instant the MPC optimization problem with $\dim(u) \times N_p = 30 \times 4$ decision variables has to be solved. Hence, more efficient algorithms and techniques which reduce the computational burden and ensure performance in real-time applications are highly desirable (see, e.g., [201, 208]).

One concept, the predictive functional controller (PFC), was first introduced by Richalet [173, 174]. The PFC is characterized by two distinctive features: The future control signal is assumed to be composed of a priori known functions, and the idea of coincidence points to evaluate the cost function along the horizon [28]. In mathematical terms this means that future control increments are described as linear combinations of basis functions $\omega_m,\ m = 1, \ldots, n_\Omega$

$$\Delta u(k+i) = \sum_{m=1}^{n_\Omega} = \mu_m(k)\omega_m(i) \quad i = 0, \ldots, N_c - 1 \tag{7.14}$$

and the predicted output is only considered at certain instants (coincidence points) $h_j, j = 1, \ldots, n_H$. Taking account of this, the original MPC cost function (6.25) on page 91 is modified by substituting $\Delta \boldsymbol{U} = \boldsymbol{\Omega\mu}$ to

$$\begin{aligned} \min_{\boldsymbol{\mu}} J = \quad & \left(\boldsymbol{R}_s^\dagger - \hat{\boldsymbol{Y}}^\dagger\right)^T \boldsymbol{Q}^\dagger \left(\boldsymbol{R}_s^\dagger - \hat{\boldsymbol{Y}}^\dagger\right) + \boldsymbol{\mu}^T \boldsymbol{\Omega}^T \boldsymbol{R} \boldsymbol{\Omega\mu} \\ \text{s. t.} \quad & \boldsymbol{M\Omega\mu} \leq \boldsymbol{N} \end{aligned} \tag{7.15}$$

with

$$\begin{aligned} \hat{\boldsymbol{Y}}^\dagger &= \left[\hat{\boldsymbol{y}}(k+h_1|k)^T \quad \hat{\boldsymbol{y}}(k+h_2|k)^T \quad \ldots \quad \hat{\boldsymbol{y}}(k+h_{n_H}|k)^T\right]^T, \\ \boldsymbol{\Omega} &= \left[\boldsymbol{\omega}_1 \quad \boldsymbol{\omega}_2 \quad \ldots \quad \boldsymbol{\omega}_{n_\Omega}\right] \end{aligned}$$

and corresponding set-point information $\boldsymbol{R}_s^\dagger$ and weighting matrix $\boldsymbol{Q}^\dagger$.

Popular basis functions are, for instance, polynomial functions [174, 28] or Laguerre functions [207]. The latter are obtained in an iterative way and are calibrated with the pole a of the Laguerre network. For details the reader is referred to Chapter 3 in [207].

To illustrate the role of the pole a, third-order Laguerre functions for $a = 0.6$ and $a = 0.8$, respectively, are depicted in Fig. 7.17. As one can see with $a = 0.6$ the functions decay to zero in less than 25 samples. In contrast to the case with $a = 0.8$, the Laguerre functions decay to zero at a much slower speed.

To measure the performance of the reduced FMPC compared to the classical FMPC a set-point change in each output variable is considered (see Fig. 7.18). For the reduced FMPC

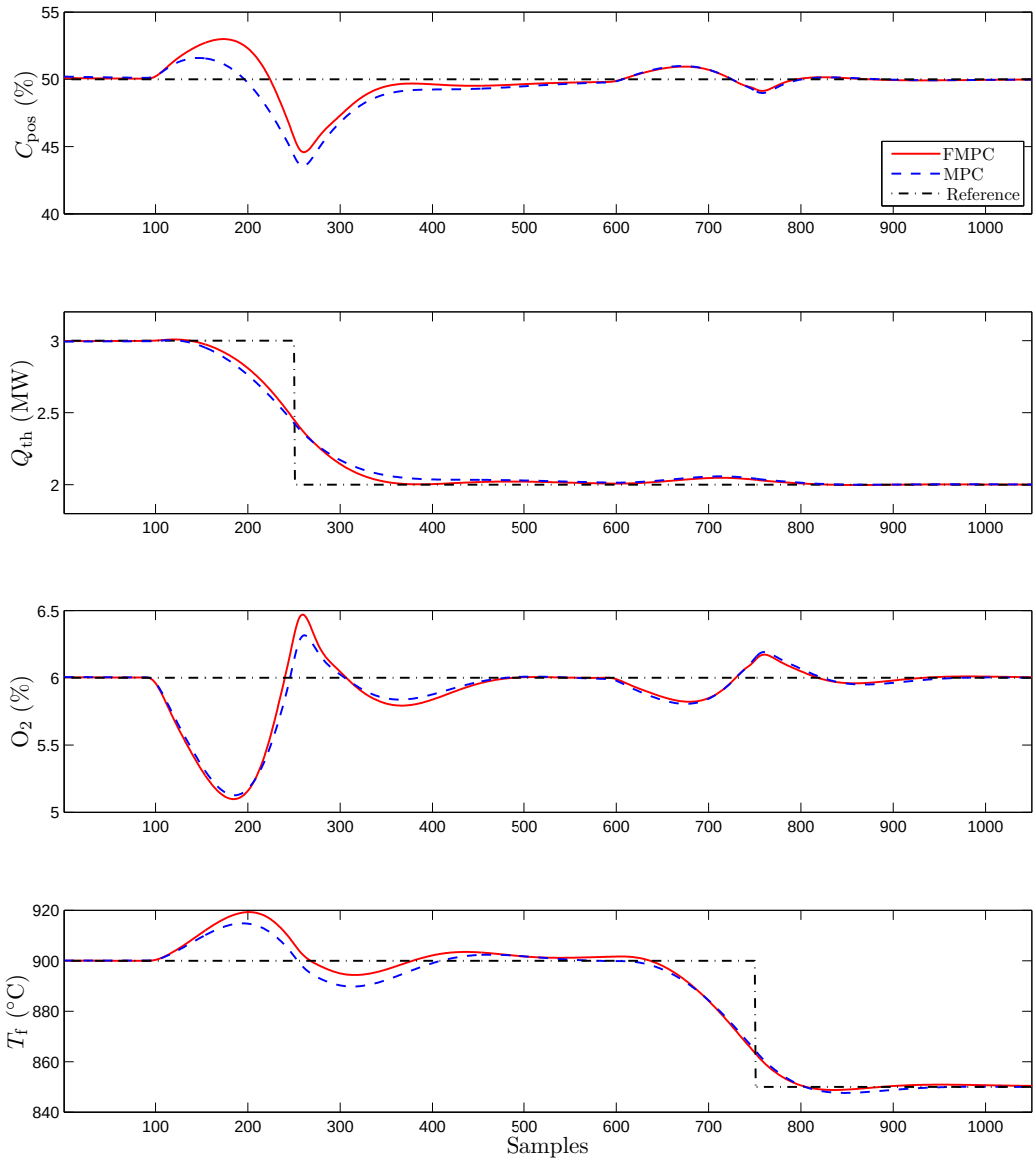


Figure 7.15: Performance comparison of set-point change in heat output and flue gas temperature

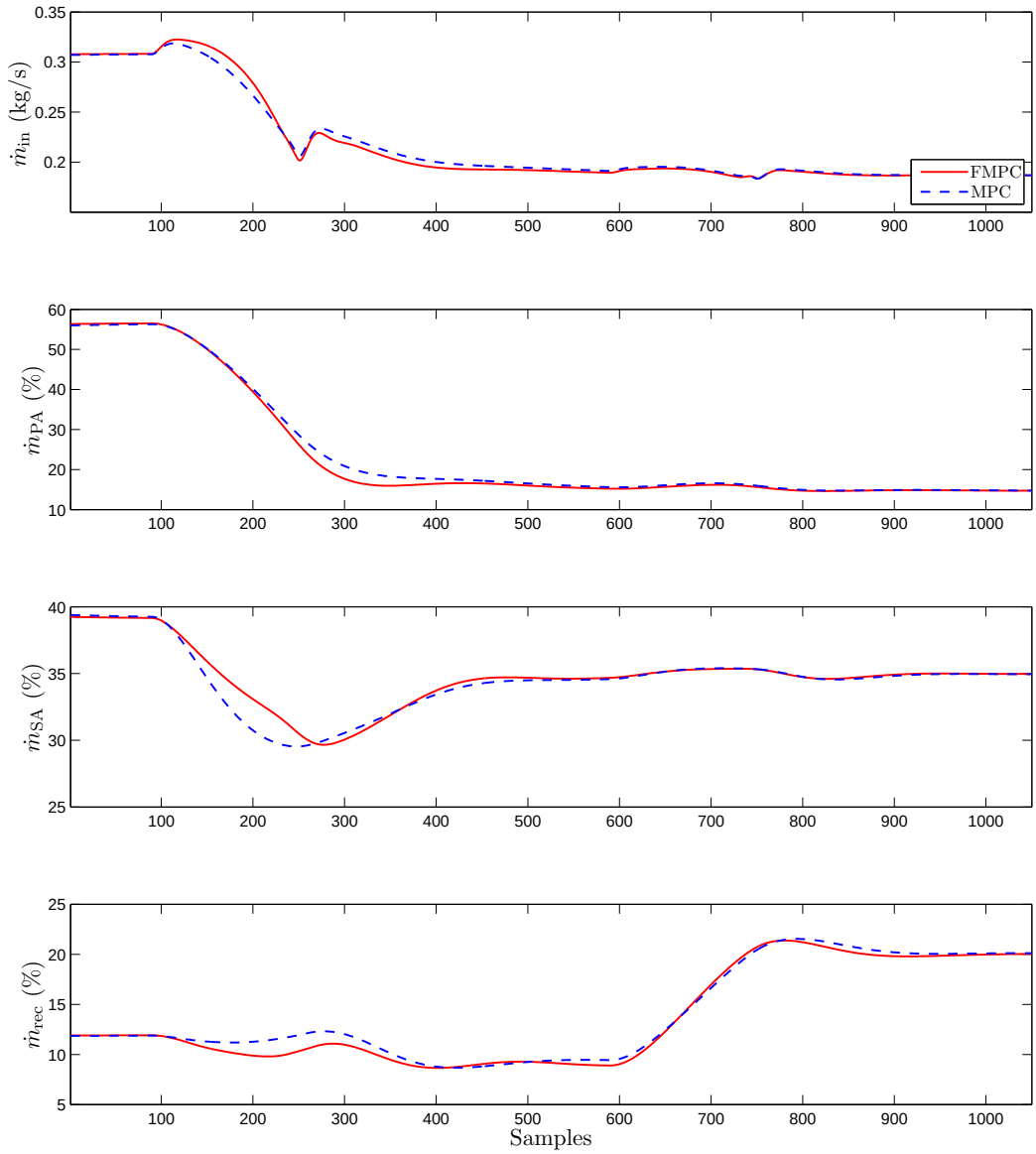


Figure 7.16: Manipulated variables of set-point change in heat output and flue gas temperature

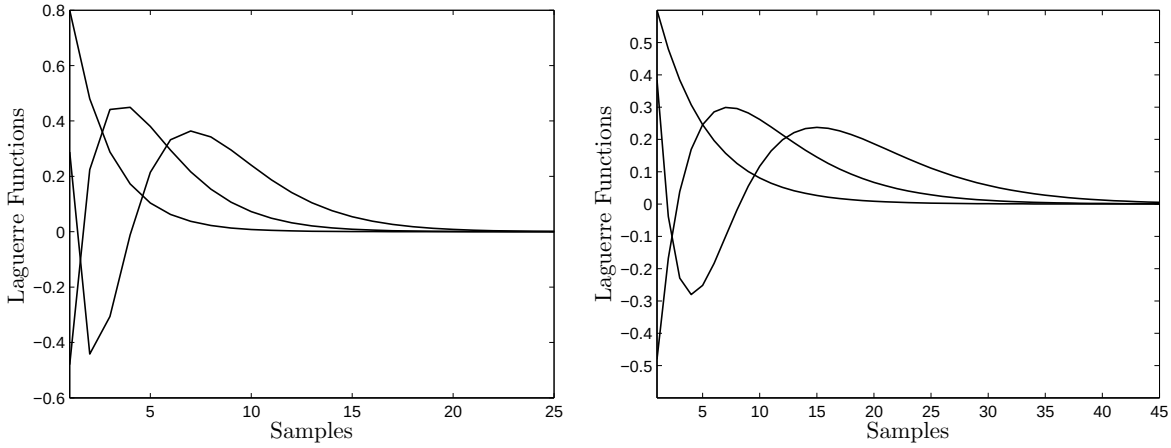


Figure 7.17: Third-order Laguerre functions with $a = 0.6$ (left) and $a = 0.8$ (right)

Table 7.5: Performance of set-point tracking of reduced FMPC and classical FMPC

#Decision Variables	80 ($= N_c \times 4$)	24	12	12
#Coincidence Points	160 ($= N_p$)	160	160	16
Avg. Exec. Time	85.79 s	58.30 s	51.19 s	42.96 s
	IAE	IAE	IAE	IAE
Combustion Position	$1.41 \cdot 10^3$	$1.44 \cdot 10^3$	$1.25 \cdot 10^3$	$1.34 \cdot 10^3$
Heat Output	59.2	59.8	56.5	56.7
Oxygen Concentration	206.8	209.5	216.7	249.9
Flue Gas Temperature	$7.69 \cdot 10^3$	$7.73 \cdot 10^3$	$7.96 \cdot 10^3$	$8.33 \cdot 10^3$

third-order Laguerre functions with $a = 0.6$ were used as basis functions for each input variable. Here the decay to zero of the functions is at about 20 samples. This corresponds to the chosen control horizon in the previous simulations.

Results such as average execution time (averaged over three runs) and integrated absolute error for each output are given in Table 7.5. For the reduced MPC with 12 ($= 3 \times \dim(u)$) decision variables and 16 coincidence points the execution time can be reduced by 50% compared to the classical FMPC, and this without negatively affecting the performance of the controller. Note that for the last setup the coincidence points were equally distributed over the prediction horizon. The closed-loop results depicted in Fig. 7.18 also show that the differences between the classical FMPC and the "most reduced" FMPC are negligibly small.

First Implementation Results

The proposed fuzzy control strategy for the moving grate biomass furnace was implemented in the biomass plant in Grossarl (Salzburg) between 14.8.2012 and 17.8.2012. On the first day it was tried to extract sufficient real measurement data from the process in order to be able to estimate the fuzzy models. However, due to technical problems the historical database was not accessible during the entire implementation period. After a brief discussion, it was decided to use the first-principles model developed in Sec. 5.1 for generating a training data set which actively excites the process over the entire operating envelope. This training data were then used to estimate the fuzzy models as described in Sec. 5.2. The workflow is summarized in the

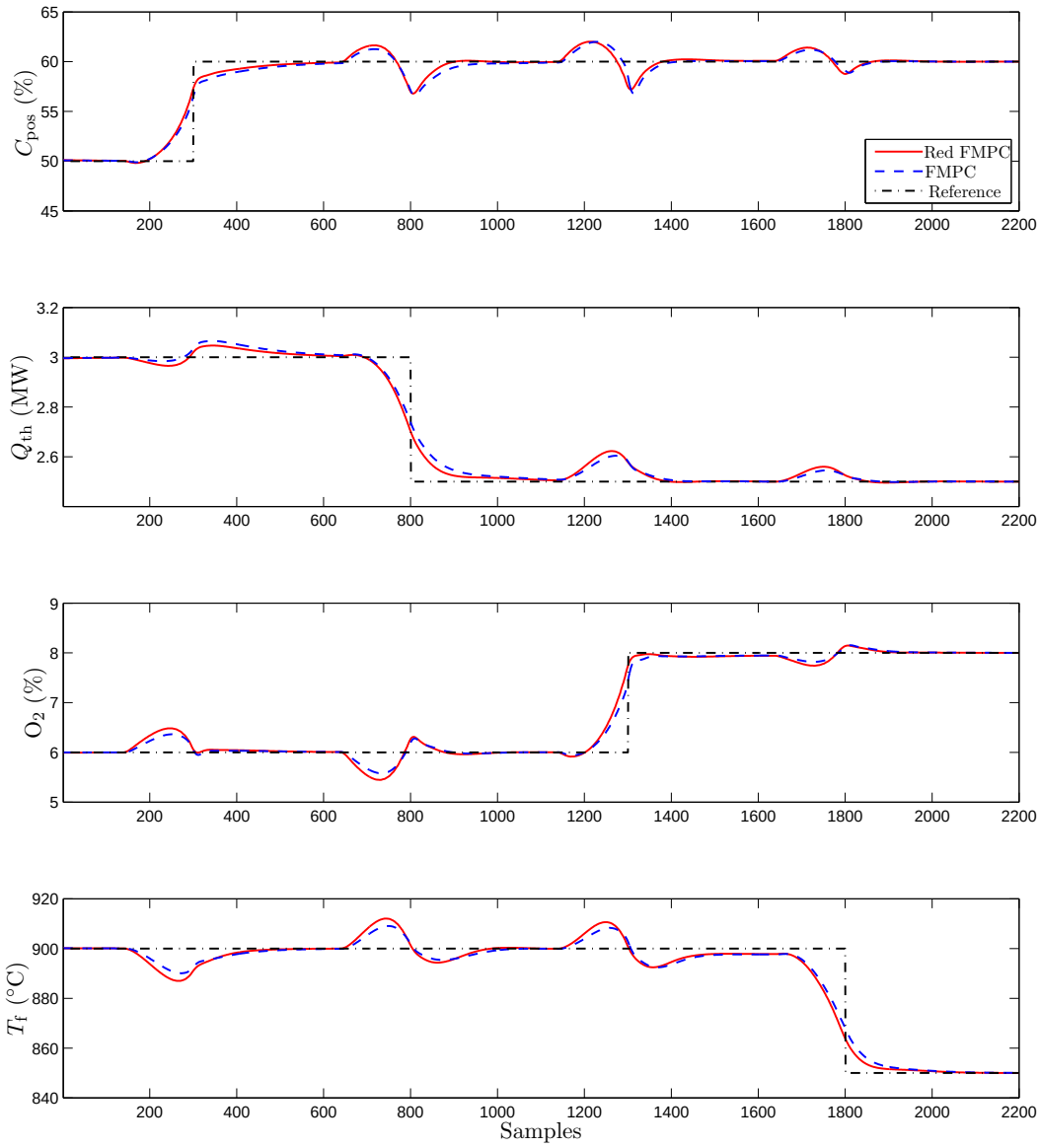


Figure 7.18: Comparison of set-point tracking of reduced FMPC with classical FMPC

following steps:

1. Calibrate the first-principle plant model using a small set of process measurement data
2. Generate representative input-output training data using the calibrated first-principle plant model
3. Construct and estimate the fuzzy models based on these training data
4. Incorporate the estimated fuzzy model into the fuzzy predictive control scheme

Since all modeling and control design was done in MATLAB® R2010a on a standalone machine and the FMPC was running on this machine, it was necessary to set up a communication between MATLAB® and the real process This was realized with the help of an OPC (Open Process Control) Server. Using the commands from the OPC Toolbox in MATLAB® it was possible to directly communicate and interact with the plant. At each time step the following steps are carried out:

1. Read current process output (combustion position, thermal power, oxygen concentration, flue gas temperature) via OPC protocol
2. Update FMPC state plus matrices and solve the quadratic programming problem
3. Send the results (i.e., control variables: input feed, primary air, secondary air, and recycle gas) to the real process via OPC protocol
4. Synchronize MATLAB® FMPC simulation with process interface sampling time (5 s)

The last step was realized with a simple `pause` command in MATLAB®. Unfortunately, no reliable measurements for the combustion position were available. Therefore it was necessary to run the first-principle model in parallel to the process and use the calculated combustion position from the first-principle model to update the controller states.

It took nearly two days to set up the communication and the interface correctly. For controller tuning and testing only one-half day remained. Nevertheless, the results of the first closed-loop run are shown in Fig. 7.19. Unexpected results were obtained with regard to the primary air supply. Primary air supply was moved towards the lower constraint of 70%. Despite this strong decrease of about 30%, there was practically no effect on the combustion position. After analyzing these results, it was found out that the internal model of the combustion position was not initialized correctly. To be more specific, the fuel mass on the grate (cf. Eq. (5.3) on page 66) was initially assumed to be 300 kg, which was too high. Hence, the simulation did not start with a stationary value of the fuel mass. Consequently, the fuel mass on the grate decreased slowly and had a significant influence on the combustion position. In response, the FMPC unsuccessfully tried to control the combustion position by means of the primary air supply. This fact, in combination with large weights on the combustion position reference error, could be the reason for these test results. Unfortunately, there was no more time for further simulation runs and the implementation had to be stopped at this point.

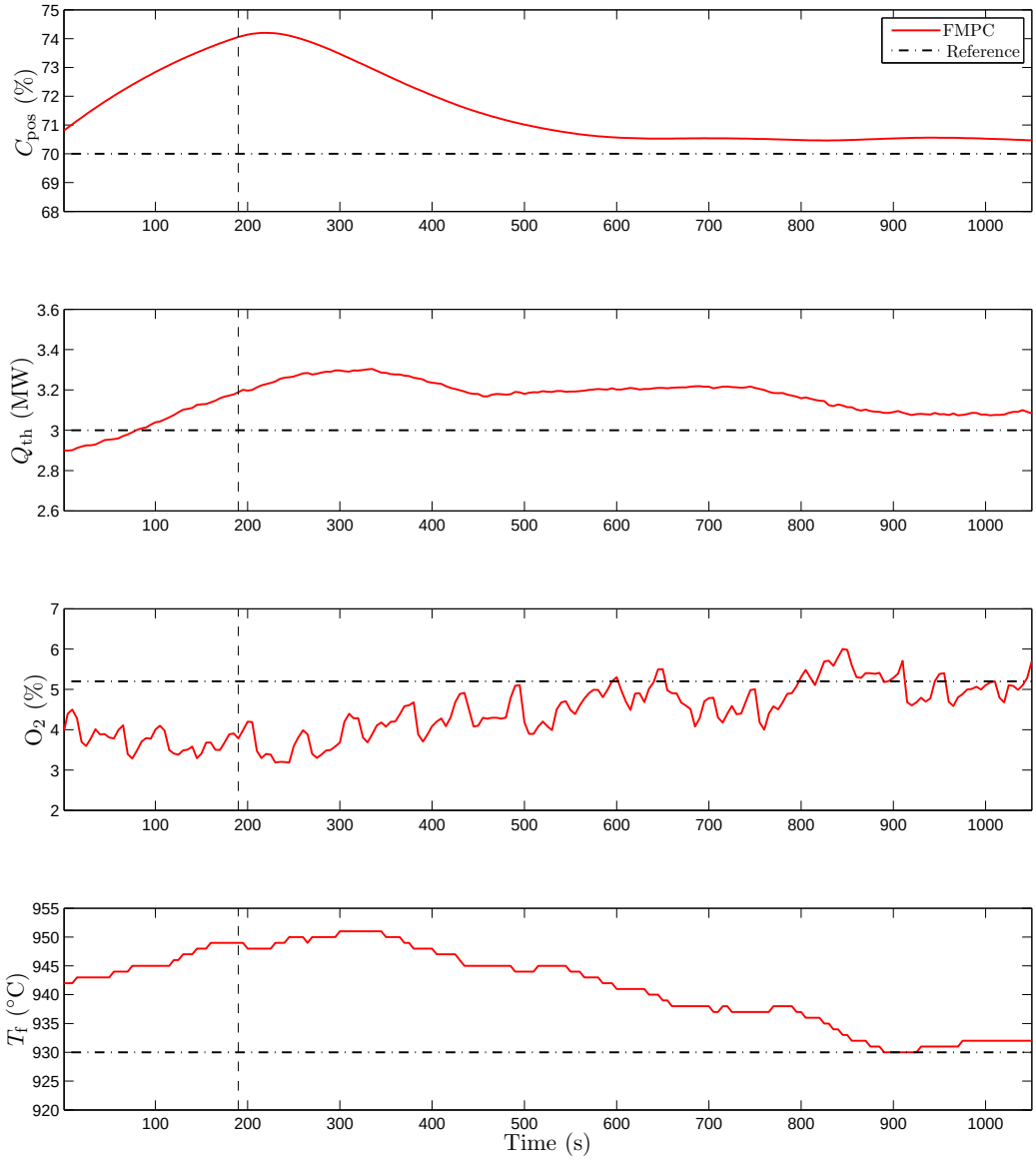


Figure 7.19: Results of output variables for the first closed-loop run; FMPC is active after the vertical dashed line

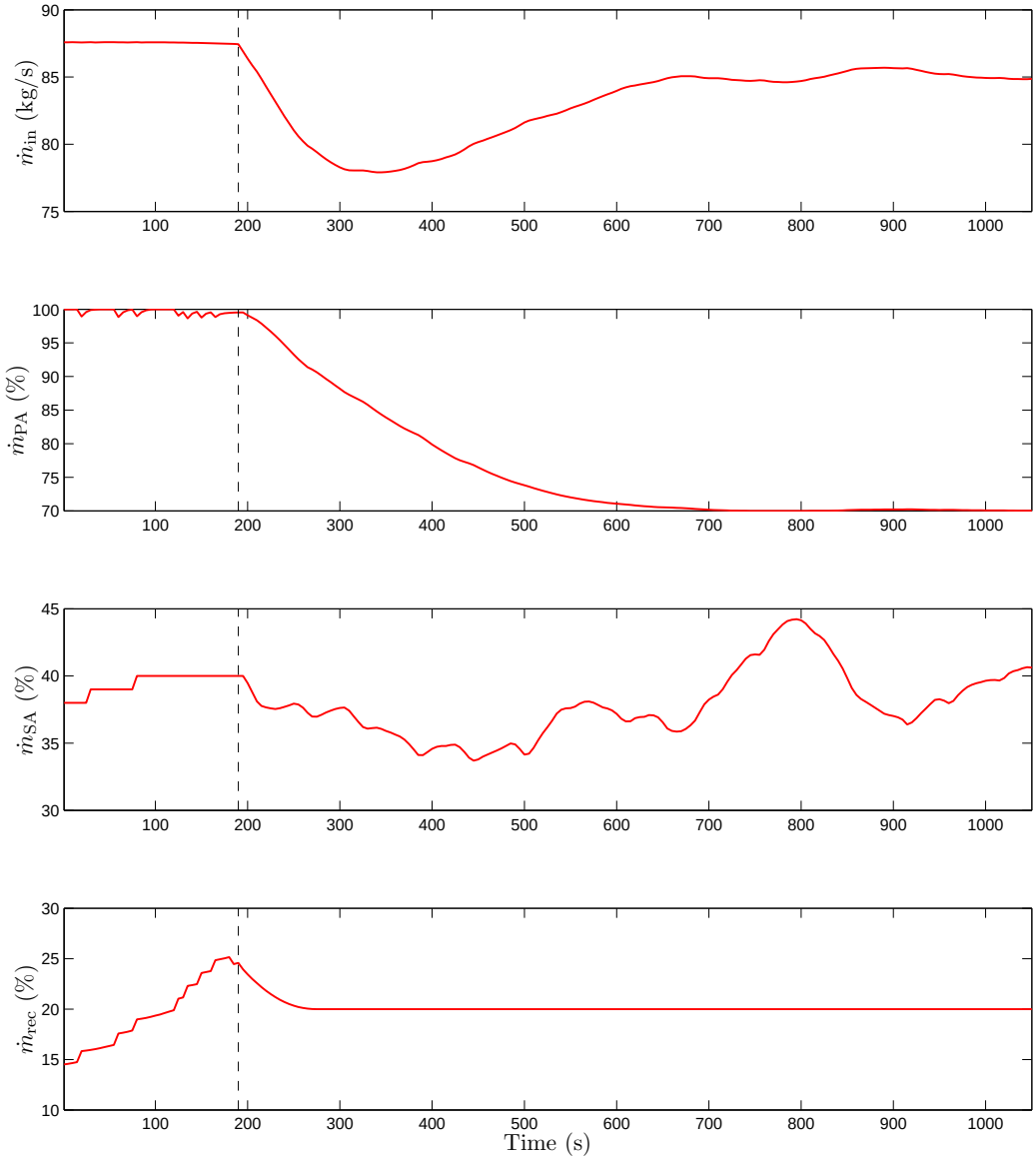


Figure 7.20: Results of manipulated variables for the first closed-loop run; FMPC is active after the vertical dashed line

Chapter 8

Conclusion

8.1 Summary of the Main Results

In this work, a general framework for modeling and control of biomass-fired district heating networks has been proposed. The main components of a district heating system are the distribution network, the biomass combustion plant and the consumer stations. For the distribution network and the biomass combustion plant, physical models as well as data-driven models have been presented. For modeling the consumer behavior statistical methods were applied. Advanced predictive control concepts have been developed for the supply temperature at critical points in the network and for a moving grate biomass furnace.

In detail, it was shown that the hydraulic flow situation in distribution networks can efficiently be determined by means of a graph-theoretical approach in combination with a Newton-based iterative algorithm. The effectiveness of this method was demonstrated on a simple diagraph with two internal loops. For modeling the thermal transients in the district heating networks a modification of the well-known QUICKEST finite-difference scheme was proposed. To overcome the problem of incorrect solutions at pipe junctions or when diameters change it was suggested to use a non-conservative formulation of the original QUICKEST approach. Moreover, it was recommended to use velocity weighted right wall values in case of pipe junctions to get a more robust scheme. The performance of the finite-difference was demonstrated by simulating the DHN in Tannheim (Austria) and comparing the simulation results with real measurements. To reduce the computational effort a physically simplified, aggregated version of the real network was used.

In addition to the physical models for the distribution network, a data-driven approach using conditional parametric models was pursued. For the three considered critical points in the DHN of Tannheim, conditional FIR models as well as a conditional ARX models were estimated. Validation results have shown that this modeling framework is capable of handling the inherent nonlinearity in the response characteristics of DHN caused by varying flow rates. Furthermore, a detailed analysis of multi-steps-ahead prediction errors was carried out which indicates that the quality of the forecasts strongly depend on the accuracy of future volume flow rate estimates.

It was demonstrated that the varying heat demand of consumers can be well described by SARIMAX and structural models in state-space form. It was further shown that exogenous influences such as the ambient temperature which attributes approximately 80% to the total heat-load can be accounted for by piece-wise linear functions. A predictive performance comparison pointed out that reliable weather forecasts are essential for accurate load predictions. For an overall simulation of the distribution network, the primary return temperature has to be known or computed. For that purpose, a nonlinear mapping describing the return temperature as a function of ambient temperature and the social load was introduced. Estimation results obtained by a sophisticated nonparametric framework illustrate that classical heat exchanger

theory is not always valid for district heating applications.

For modeling the biomass plant a simple grey-box model was introduced which is largely based on mass and energy balances. The model was designed in such way to be easily calibrated with a small set of tuning parameters using real measurements. The validation with real data from a biomass plant in Grossarl (Austria) showed acceptable performance. Beside the grey-box model, Takagi-Sugeno fuzzy models were developed for a moving grate biomass furnace. The main variables responsible for the nonlinearity of the process were identified to be the heat output and flue gas temperature. For locally approximating the process linear and nonlinear estimation techniques were introduced. The comparison with real measurements obtained from the biomass plant in Grossarl (Salzburg) showed good agreement.

The concept of fuzzy model predictive control was introduced for control of the supply temperature at critical nodes in the network. For this purpose a novel approach was proposed which iteratively determines a future fuzzy trajectory based on heat load forecasts. The comparison with a conventional MPC clearly demonstrates the performance improvement of this approach, especially in regions of low volume flow rates and thus large transport time delays. Additionally, set-point generation tools based on heating curves as well as a set-point optimizer minimizing operational costs were studied. With the latter the overall costs can be reduced by 3.44% in contrast to the results with set-points constructed by heating curves.

In the final section a fuzzy model predictive control scheme for a moving grate biomass furnace was developed. The effectiveness of this approach is demonstrated by simulating set-point changes in heat output and flue gas temperature. Furthermore, possible reduction in the computational burden of fuzzy controller was discussed using basis functions for future control increments and the idea of coincidence points to evaluate the cost function along the prediction horizon. Simple case studies have shown that this reduces the execution time significantly without affecting the overall performance. In the final part of this work first implementation results were summarized and discussed. Due to technical issues and incorrect model initialization, the implementation was only partly successful.

8.2 Outlook

The contributions presented in this work unveil interesting and challenging future research opportunities: For instance, to increase the applicability of the physical modeling framework, a general approach for aggregating complex district heating networks could be defined. Such an approach may even reach the point of software tools which are able to read, for example CAD drawings of the distribution network and automatically construct aggregated networks with corresponding incidence matrix and pipe parameters. In the context of modeling and predicting heat-load the classical Box-Jenkins methodology has reached its limits. To take into account the seasonal variations, extensions such as periodic SARIMA models [97], varying coefficient models [155] or nonparametric estimation techniques like MARS [63] might be considered.

Optimal control of the biomass plant and the distribution network was addressed separately in this work. To obtain a global optimum it is necessary to link both, the production side and the distribution side. Due to highly different response times multi-sampling (Fuzzy) MPC have to be designed for that purpose. The consideration of more than one biomass plant, co-generation plants or heat storages in DHN poses another challenging task. Such processes can easily reach a complexity which exceeds the capacity of current algorithmic and computational resources. Hence, to ensure performance in real-time applications efficient algorithms have to be designed.

The ultimate goal of such advanced control concepts is to be implemented into existing district heating systems. This is not only challenging from a technical point of view, as experienced by the author, but also considerable time has to be invested to convince plant operators of of its additional benefit compared to conventional control schemes.

List of Figures

2.1 Schematic drawing of a district heating network, with supply temperatures T_s, return temperatures T_r, pressure differentials Δp and heat Q. 7
2.2 District heating network in Tannheim (Austria) 8
2.3 Heat exchanger scheme, with the mass flow rate $\dot{m}_i$, supply temperatures T_r^i, return temperatures T_r^i, heat flow Q and characteristic transfer capability kA $(i = 1, 2)$ 9

3.1 Pipe differential elements showing the hydraulic (left) and thermal (right) physical properties; H_{12} is the head difference between the pipe ends and T_g is the ground temperature 12
3.2 Example diagraph 13
3.3 Schematic representation of the uniform grid structure in space and time (Quickest stencil is shown) 17
3.4 Representation of a pipe junction in district heating networks assuming non-uniform grid 17
3.5 Schematic representation of the finite-difference grid for the supply line with imposed boundary conditions: Dirichlet boundary condition plus mirror node upstream (UB); Zero curvature condition downstream (DB) 19
3.6 Schematic representation of the finite-difference grid for the return line with imposed boundary conditions: Dirichlet boundary condition plus mirror node upstream (UB); Zero curvature condition downstream (DB) 20
3.7 Two-dimensional plot of the amplification factor modulus, $G(\theta, c)$, for the Quickest scheme 21
3.8 Original and aggregated district heating network in Tannheim 23
3.9 Aggregated volume flow rates for AC5 and AC7 between 19.-26. Dec. 2008 . . . 25
3.10 Performance comparison of the finite-difference model of the supply temperature at AC5 (top) and at AC7 (bottom) for one week of real data ($\sigma_\epsilon^{\mathrm{AC5}} = 0.624\,^\circ\mathrm{C}$, $\sigma_\epsilon^{\mathrm{AC7}} = 0.65\,^\circ\mathrm{C}$) 26
3.11 Performance comparison of the finite-difference model of the return temperature at at the biomass plant for one week of real data ($\sigma_\epsilon = 0.39\,^\circ\mathrm{C}$) 26
3.12 Different benchmark temperature profiles 26
3.13 First four iterations of LOLIMOT algorithm for a two-dimensional input space $(m = 2)$ 32
3.14 Interpretation of Fig. 3.13 as a decision tree 33
3.15 Original and aggregated district heating network in Tannheim 34
3.16 Histogram plus estimated density of the volume flow rate 35
3.17 Root mean squared simulation error for the training dataset and the validation dataset against the number of local models. Results are shown for the cARX model and CP3 36

3.18 Membership functions generated by LOLIMOT of CP1 (AC3), CP2 (AC5) and CP3 (AC7) . . . 37
3.19 Stationary gain against volume flow rate of CP1 (AC3), CP2 (AC5) and CP3 (AC7) for cARX (left) and cFIR models (right) . . . 38
3.20 Estimated social component for the three critical points . . . 39
3.21 Impulse response function of the cARX model for critical point 3 (AC7) . . . 40

4.1 Multi-year record of system heat-load of DHN in Tannheim . . . 45
4.2 Weekly season of system heat-load of DHN in Tannheim for winter (top) and summer (bottom) . . . 46
4.3 Daily heat-load profile for summer (left) and winter (right) . . . 46
4.4 System heat load against ambient temperature plus piecewise linear function describing the dependency . . . 52
4.5 ACF (left) and PACF (right) of the residuals $\{\hat{V}_t\}$ . . . 52
4.6 ACF (left) and PACF (right) of the seasonal differenced time series $\{\nabla_{48}\hat{V}_t\}$ with 95% confidence interval . . . 53
4.7 ACF (left) and PACF (right) of the double differenced time series $\{\nabla\nabla_{48}\hat{V}_t\}$ with 95% confidence interval . . . 54
4.8 Diagnostic plots of standardized prediction error for SARIMA model: (top left) histogram plus density of normal distribution; (top right) QQ-Plot; (bottom) correlogram . . . 55
4.9 Diagnostic plots of standardized prediction error for state-space model: (top left) histogram plus density of normal distribution; (top right) QQ-Plot; (bottom) correlogram . . . 57
4.10 16-step-ahead forecasts of system heat-load during winter, including their 90% confidence interval; the full black line are the heat-load observations . . . 59
4.11 16-step-ahead forecasts of system heat-load during summer, including their 90% confidence interval; the full black line are the heat-load observations . . . 59
4.12 Supply temperature against ambient temperature for AC4 (top) and AC7 (bottom) plus piecewise linear fit . . . 61
4.13 Regression surface for the return temperature estimated by MARS for AC4 . . . 62
4.14 Regression surface for the return temperature estimated by MARS for AC7 . . . 63

5.1 Schematic drawing of the moving grate biomass furnace. The fuel is fed to the left side of the grate and is transported to the ash conveyor on the far right side (reproduced with modifications from [164]) . . . 65
5.2 Comparison of simulated output of the first-principle plant model with process measurement data for 15.8.2012 . . . 68
5.3 Illustration of a linguistic variable "temperature" with tree linguistic terms: low, medium, and high . . . 69
5.4 Illustration of the staged optimization approach . . . 76
5.5 Normalized mean squared simulation error for the training dataset and the validation dataset against the number of local models (fuzzy rules); (from top to bottom) combustion position, thermal power, oxygen concentration, and flue gas temperature; the arrow indicates the selected model complexity . . . 77
5.6 Normalized Gaussian membership functions constructed by LOLIMOT for combustion position (top), thermal power (middle), and oxygen concentration (bottom) 78
5.7 Fuzzy space decomposition constructed by LOLIMOT for flue gas temperature; the gray points are the training data samples . . . 79

5.8 Oxygen concentration: comparison between process and model output (top); simulation error for the validation dataset (bottom) 80
5.9 Heat output: comparison between process and model output (top); simulation error for the validation dataset (bottom) 80
5.10 Oxygen concentration: comparison between process and model output (top); simulation error for the validation dataset (bottom) 81
5.11 Flue gas temperature: comparison between process and model output (top); simulation error for the validation dataset (bottom) 81
5.12 Scatterplot of flue gas temperature versus heat output plus linear and quadratic function describing the dependency 82

6.1 Principle of predictive control 84
6.2 Multilayer control structure (reproduced with modifications from [197]) 85
6.3 Internal model control (IMC) scheme 95
6.4 Idea of the dual-mode approach (reproduced from [145]) 96
6.5 Sketch of a predictive controller with robust stability constraints (reproduced with modifications from [149]) 96

7.1 Schematic heating curve with characteristic break-points (reproduced with modifications from [159]). 99
7.2 Simplified district heating network with one consumer 101
7.3 Block diagram of the Fuzzy MPC (SP: set point generation/optimization) 102
7.4 Performance comparison of set point tracking and step disturbance rejection under cFIR-MPC and cFIR-FMPC with AC5 being active 104
7.5 Stepwise volume flow profile plus corresponding memberhip function values of the local models for the cFIR-FMPC with AC5 being active; the thick full line represents the volume flow rate; the gray shaded areas reflect the individual local models 105
7.6 System heat load and ambient temperature profile for the two time periods considered; sampling time is 30 min 106
7.7 Fitted heating curve for AC5 106
7.8 Simulation results for Period I using heating curves. The shaded areas correspond to the currently active critical point 107
7.9 Simulation results for Period I using heating curves. The shaded areas correspond to the currently active critical point 108
7.10 Heat loss costs and pumping costs for $Q_{\mathrm{P}} = 2\,\mathrm{MW}$ (left) and $Q_{\mathrm{P}} = 3\,\mathrm{MW}$ (right) using stationary gain curve from AC3; return temperature at plant is fixed to 55 °C 108
7.11 Simulation results for AC3 with set-points generated by heating curve and by set-point optimization, respectively 109
7.12 Overview of input- and output variables of the biomass plant model 110
7.13 Step responses of the MIMO fuzzy model for operating point ($Q_{\mathrm{th}} = 2.5\,\mathrm{MW}$, $T_{\mathrm{f}} = 900\,°\mathrm{C}$) with a sampling time of 5 s 111
7.14 Block diagram of the fuzzy model predictive control scheme 112
7.15 Performance comparison of set-point change in heat output and flue gas temperature 114
7.16 Manipulated variables of set-point change in heat output and flue gas temperature 115
7.17 Third-order Laguerre functions with $a = 0.6$ (left) and $a = 0.8$ (right) 116
7.18 Comparison of set-point tracking of reduced FMPC with classical FMPC 117
7.19 Results of output variables for the first closed-loop run; FMPC is active after the vertical dashed line 119

7.20 Results of manipulated variables for the first closed-loop run; FMPC is active after the vertical dashed line 120

List of Tables

3.1 Overview of Finite-Difference Schemes 18
3.2 Physical properties of the pipes of the aggregated district heating network; Length, diameter are in m and heat loss in W/m^2K 24
3.3 Summary of the supply temperature statistics of the identification data set . . . 35
3.4 Prediction results of cFIR models with observed flow rate measurements as input 42
3.5 Prediction results of cARX models with observed flow rate measurements as input 42
3.6 Prediction results of cFIR models with predicted flow rate measurements as input 43
3.7 Prediction results of cARX models with predicted flow rate measurements as input 43
3.8 Prediction results of cARX model for CP3 using observed/predicted/constant future flow rates as input 43

4.1 Summary of the statistics of the training data set 51
4.2 Maximum Likelihood estimation results for the state-space heat load model . . . 56
4.3 Prediction performance for the state-space temperature model; prediction accuracy measures are in °C 58
4.4 Prediction performance comparisons of heat-load with observed ambient temperature; prediction accuracy measures are in kW 58
4.5 Prediction performance comparisons of heat-load with predicted ambient temperature; prediction accuracy measures are in kW 58

5.1 Nominal parameter values for the moving grate biomass furnace model 67
5.2 Comparison of the fitting performance using PCA and NLPCA 82

7.1 Performance index comparison of set point tracking under MPC and fuzzy MPC 104
7.2 Control Curve Specification in °C 105
7.3 Daily operating costs in € for set-points generated by heating curve and by set-point optimization, respectively 109
7.4 Integrated control error of set-point tracking under MPC and fuzzy MPC 113
7.5 Performance of set-point tracking of reduced FMPC and classical FMPC 116

Bibliography

[1] J. Abonyi. *Fuzzy model identification for control.* Birkhäuser Boston, 2003.

[2] J. Abonyi and B. Feil. *Cluster Analysis for Data Mining and System Identification.* Birkhäuser Basel, 2007.

[3] F. Allgöwer and A. Zheng. *Nonlinear model predictive control*, volume 26. Birkhäuser, 2000.

[4] B.D.O. Anderson and J.B. Moore. *Optimal filtering*, volume 11. Prentice-hall Englewood Cliffs, NJ, 1979.

[5] R. Aringhieri and F. Malucelli. Optimal operations management and network planning of a district heating system with a combined heat and power plant. *Annals of Operations Research*, 120(1):173–199, 2003.

[6] D.H. Axworthy. *Water distribution network modelling: From steady state to waterhammer.* PhD thesis, Department of Civil Engineering, University of Toronto, 1997.

[7] R. Babuéska. *Fuzzy modeling for control.* Springer, 1998.

[8] A. Badwe, S. Shah, S. Patwardhan, and R. Patwardhan. Model-plant mismatch detection in mpc applications using partial correlation analysis. In *IFAC World Congress*, pages 14926–14933, 2008.

[9] H.D. Baehr and K. Stephan. *Heat and Mass Transfer.* Springer-Verlag Berlin, 2006.

[10] J. Balate, B. Chramcov, P. Navratil, and M. Princ. Strategy evolution of control of extensive district heating systems. In *International Conference on Power Engineering, Energy and Electrical Drives*, pages 678–683. IEEE, 2007.

[11] R. Baldick. The generalized unit commitment problem. *IEEE Transactions on Power Systems*, 10(1):465–475, 1995.

[12] D. Bao-Cang. *Modern predictive control.* Taylor & Francis Group, 2010.

[13] R. Bauer, M. Gölles, T. Brunner, N. Dourdoumas, and I. Obernberger. Modelling of grate combustion in a medium scale biomass furnace for control purposes. *Biomass and Bioenergy*, 34(4):417–427, 2010.

[14] K. Belarbi and F. Megri. A stable model-based fuzzy predictive control based on fuzzy dynamic programming. *IEEE Transactions on Fuzzy Systems*, 15(4):746–754, 2007.

[15] A. Benonysson, B. Bøhm, and H.F. Ravn. Operational optimization in a district heating system. *Energy conversion and management*, 36(5):297–314, 1995.

[16] B.W. Bequette. Non-linear model predictive control: A personal retrospective. *The Canadian Journal of Chemical Engineering*, 85(4):408–415, 2007.

[17] S. Bercu and F. Proïa. A sarimax coupled modelling applied to individual load curves intraday forecasting. *Arxiv preprint arXiv:1207.0360*, 2012.

[18] M.A. Bernier and B. Bourret. Pumping energy and variable frequency drives. *ASHRAE Journal*, 41(12):37–40, 1999.

[19] R.B. Bird, W.E. Stewart, and E.N. Lightfoot. *Transport phenomena*. Wiley, 2006.

[20] J.R. Birge and F. Louveaux. *Introduction to stochastic programming*. Springer Verlag, 1997.

[21] B. Bøhm, S. Ha, W. Kim, B. Kim, T. Koljonen, H.V. Larsen, M. Lucht, Y. Park, K. Sipilä, M. Wigbels, et al. Simple models for operational optimisation. Technical report, IEA district heating and cooling, 2002. Annex VI: Report 2002:S1.

[22] B. Bøhm and H.V. Larsen. *Simple models of district heating systems for load and demand side management and operational optimisation*. Department of Mechanical Engineering, Technical University of Denmark, 2004.

[23] G.E.P. Box, G.M. Jenkins, and G.C. Reinsel. *Time series analysis: forecasting and control*. Wiley Series in Probability and Statistics, 4th edition, 2008.

[24] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[25] L. Breiman, J. Friedman, C.J. Stone, and R.A. Olshen. *Classification and regression trees*. Chapman & Hall/CRC, 1984.

[26] P.J. Brockwell and R.A. Davis. *Introduction to time series and forecasting*. Springer Verlag, 2002.

[27] P.J. Brockwell and R.A. Davis. *Time series: theory and methods*. Springer Verlag, 2009.

[28] E.F. Camacho and C. Bordons. *Model predictive control*. Springer Berlin, 2004.

[29] Y.A. Cengel. *Heat transfer: a practical approach*. McGraw-Hill Boston, 2003.

[30] M.H. Chaudhry. Applied hydraulic transients. *Van Nostrand Reinhold Co,New York*, 1987.

[31] G. Chicco, R. Napoli, and F. Piglione. Application of clustering algorithms and self organising maps to classify electricity customers. In *Proceedings of Power Tech Conference, Bologna*, volume 1, pages 7–pp. IEEE, 2003.

[32] G. Chicco, R. Napoli, and F. Piglione. Comparisons among clustering techniques for electricity customer classification. *IEEE Transactions on Power Systems*, 21(2):933–940, 2006.

[33] B. Chramcov. Forecast of heat demand according the box-jenkins methodology for specific locality. *Latest Trends on Systems*, 1:252–256, 2010.

[34] B. Chramcov, P. Dostál, and J. Baláte. Forecast model of heat demand. In *The 29th Annual International Symposium on Forecasting, Hong Kong*, pages 1–9, 2009.

[35] D.W. Clarke, C. Mohtadi, and PS Tuffs. Generalized predictive control-part I. the basic algorithm. *Automatica*, 23(2):137–148, 1987.

[36] W.S. Cleveland. Coplots, nonparametric regression, and conditionally parametric fits. *Lecture Notes-Monograph Series*, pages 21–36, 1994.

[37] W.S. Cleveland and S.J. Devlin. Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83(403):596–610, 1988.

[38] W.S. Cleveland, E. Grosse, and W.M. Shyu. Local regression models. *Statistical models in S*, pages 309–376, 1992.

[39] K. Comaklı, B. Yüksel, and Ö. Comaklı. Evaluation of energy and exergy losses in district heating network. *Applied thermal engineering*, 24(7):1009–1017, 2004.

[40] J.J.F. Commandeur and S.J. Koopman. *An introduction to state space time series analysis.* Oxford University Press, USA, 2007.

[41] J. Conti and P. Holtberg. International energy outlook 2011. *US Energy Information Administration*, 2011.

[42] R. Courant, K. Friedrichs, and H. Lewy. On the partial difference equations of mathematical physics. *IBM Journal of Research and Development*, 11(2):215–234, 1967.

[43] H. Cross. Analysis of flow in networks of conduits or conductors. Technical report, College of Engineering, University of Illinois, 1936.

[44] S. Cross. Biomass 2020: Opportunities, challenges and solutions. Technical report, Union of the Electricity Industry–EURELECTRIC, 2011. http://www.eurelectric.org/media/26720/resap_biomass_2020_81111_prefinal-2011-113-0004-01-e.pdf (retrieved on 23.01.2013).

[45] J.D. Cryer and K.S. Chan. *Time series analysis: with applications in R.* Springer Verlag, 2008.

[46] L. Dannecker, M. Boem, U. Fischer, F. Rasenthal, G. Hackenbroich, and W. Lehner. A survey of forecast models for energy demand and supply. Technical report, TU Dresden, Database Technology Group / SAP Research CEC Dresden, 2010.

[47] P. De Jong. The diffuse Kalman filter. *The Annals of Statistics*, 19(2):1073–1083, 1991.

[48] Reinhard Diestel. *Graph Theory.* Springer Verlag, electronic edition, 2000.

[49] V. Dordonnat, S.J. Koopman, and M. Ooms. Dynamic factors in state-space models for hourly electricity load signal decomposition and forecasting. In *Power & Energy Society General Meeting*, pages 1–8. IEEE, 2009.

[50] V. Dordonnat, S.J. Koopman, and M. Ooms. Dynamic factors in periodic time-varying regressions with an application to hourly electricity load modelling. *Computational Statistics & Data Analysis*, 2011.

[51] V. Dordonnat, S.J. Koopman, M. Ooms, A. Dessertaine, and J. Collet. An hourly periodic state space model for modelling french national electricity load. *International Journal of Forecasting*, 24(4):566–587, 2008.

[52] E. Dotzauer. Simple model for prediction of loads in district-heating systems. *Applied Energy*, 73(3-4):277–284, 2002.

[53] J. Durbin and S.J. Koopman. *Time series analysis by state space methods.* Oxford University Press Inc., New York, 2001.

[54] R. Ehrig, M. Wörgetter, C. Kristöfel, N. Ludwiczek, C. Pointner, and C. Strasser. Biomass mobilisation for industrial-scale bioenergy plants. Technical report, Bioenergy 2020+, 2011. http://www.bioenergy2020.eu/files/publications/pdf/I1-2598.pdf (retrieved on 23.01.2013).

[55] M. Espinoza, C. Joye, R. Belmans, and B. DeMoor. Short-term load forecasting, profile identification, and customer segmentation: a methodology based on periodic time series. *IEEE Transactions on Power Systems*, 20(3):1622–1630, 2005.

[56] R.E. Ewing and H. Wang. A summary of numerical methods for time-dependent advection-dominated partial differential equations. *Journal of Computational and Applied Mathematics*, 128(1):423–445, 2001.

[57] J. Fan and I. Gijbels. *Local polynomial modelling and its applications*, volume 66. Chapman & Hall/CRC, 1996.

[58] J. Fan and W. Zhang. Statistical methods with varying coefficient models. *Statistics and its Interface*, 1(1):179, 2008.

[59] R. Findeisen and F. Allgöwer. An introduction to nonlinear model predictive control. In *21st Benelux Meeting on Systems and Control*, volume 11, 2002.

[60] A. Fink, O. Nelles, and R. Isermann. Nonlinear internal model control for miso systems based on local linear neuro-fuzzy models. In *Proceedings of the 15th IFAC World Congress*, 2002.

[61] R. Fletcher. *Practical methods of optimization.* Wiley-Interscience, New York, 2nd edition, 1987.

[62] T.R. Fortescue, L.S. Kershenbaum, and B.E. Ydstie. Implementation of self-tuning regulators with variable forgetting factors. *Automatica*, 17(6):831–835, 1981.

[63] J.H. Friedman. Multivariate adaptive regression splines. *The annals of statistics*, pages 1–67, 1991.

[64] I. Gabrielaitiene. Numerical simulation of a district heating system with emphases on transient temperature behaviour. In *Proceedings of the 8th International Conference on Enviromental Engineering*, 2011.

[65] I. Gabrielaitiene, B. Bøhm, and B. Sunden. Modelling temperature dynamics of a district heating system in naestved, denmark–a case study. *Energy conversion and management*, 48(1):78–86, 2007.

[66] I. Gabrielaitiene, B. Sunden, and R. Kacianauskas. Analysis of fluid flow and heat transfer in district heating pipelines using the finite element method. *Heat Transfer VII. Southampton: WIT Press*, pages 13–22, 2002.

[67] A. Gajdos and S. Mandelkern. Comparative study of numerical schemes used for one-dimensional transport modelling. *2nd Int. PhD Symposium in Civil Engineering, Budapest*, 1998.

[68] C.E. Garcia and M. Morari. Internal model control. a unifying review and some new results. *Industrial & Engineering Chemistry Process Design and Development*, 21(2):308–323, 1982.

[69] A. Ghodsi. Dimensionality reduction a short tutorial. Technical report, Department of Statistics and Actuarial Science, Univ. of Waterloo, Ontario, Canada, 2006.

[70] P.E. Gill, W. Murray, and M.H. Wright. *Practical optimization*, volume 1. Academic press, 1981.

[71] S.K. Godunov. A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics. *Matematicheskii Sbornik*, 89(3):271–306, 1959.

[72] M. Gölles, R. Bauer, T. Brunner, N. Dourdoumas, and I. Obernberger. Model based control of a biomass grate furnace. In *European Conference on Industrial Furnaces and Boilers*, 2011.

[73] S. Grosswindhager. Using penalized logistic regression models for predicting the effects of advertising material. Master's thesis, Vienna University of Technology, 2009.

[74] S. Grosswindhager, L. Haffner, A. Voigt, and M. Kozek. Fuzzy modelling of moving grate biomass furnace for simulation and control purposes. *Mathematical and Computer Modelling of Dynamical Systems*, in review procedure.

[75] S. Grosswindhager, M. Kozek, L. Haffner, and A. Voigt. Fuzzy predictive control of district heating network. *International Journal of Modelling, Identification and Control*, accepted.

[76] S. Grosswindhager, A. Voigt, and M. Kozek. Efficient physical modelling of district heating networks. In *Modelling and Simulation*. ACTA Press, 2011.

[77] S. Grosswindhager, A. Voigt, and M. Kozek. Linear finite-difference schemes for energy transport in district heating networks. In *Proceedings of the 2nd International Conference on Computer Modelling and Simulation-CSSim2011*, pages 35–42, 2011.

[78] S. Grosswindhager, A. Voigt, and M. Kozek. Online short-term forecast of system heat load in district heating networks. In *Proceedings of the 31nd Annual International Symposium on Forecasting*, 2011.

[79] S. Grosswindhager, A. Voigt, and M. Kozek. Predictive control of district heating network using fuzzy DMC. In *Proceedings of International Conference on Modelling, Identification & Control (ICMIC)*, pages 241–246. IEEE, 2012.

[80] Stefan Grosswindhager, Klemens Schulmeister, and Martin Kozek. Model predictive control for lateral position of endless metal belts. *Solid State Phenomena*, 198:525–532, 2013.

[81] Grundfos, net.grundfos.com/Appl/WebCAPS/Grundfosliterature-1257.pdf. *GRUNDFOS Datenheft*, 2012. retrieved on 10.05.2012.

[82] M. Grzenda and B. Macukow. Heat consumption prediction with multiple hybrid models. *Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living*, pages 1213–1221, 2009.

[83] S. Gunnarsson. Combining tracking and regularization in recursive least squares identification. In *35th IEEE conference on Decision and Control*, volume 3, pages 2551–2552. IEEE, 1996.

[84] J. Gustafsson, J. Delsing, and J. van Deventer. Improved district heating substation efficiency with a new control strategy. *Applied energy*, 87(6):1996–2004, 2010.

[85] A. Hakulinen, J. Lampinen, and J. Lavanti. Efficiency of district heating water pumping in Finland. In *The 12th International Symposium on District Heating and Cooling*, 2010.

[86] H.W. Hamacher and K. Klamroth. *Lineare Optimierung und Netzwerkoptimierung*. Vieweg Verlag, 2006.

[87] C. Hametner and S. Jakubek. Nonlinear identification with local model networks using gtls techniques and equality constraints. *IEEE Transactions on Neural Networks*, 22(9):1406 –1418, 2011.

[88] C. Hametner and M. Nebel. Operating regime based dynamic engine modelling. *Control Engineering Practice*, 2012.

[89] W. Härdle. *Applied Nonparametric Regression*. Citeseer, 1994.

[90] A. Harvey and S.J. Koopman. Structural time series models. *Encyclopedia of Biostatistics*, 1993.

[91] L.D.D. Harvey. *A handbook on low-energy buildings and district-energy systems: fundamentals, techniques and examples*. Earthscan, 2006.

[92] T. Hastie and R. Tibshirani. Varying-coefficient models. *Journal of the Royal Statistical Society, Series B*, 55(4):757–796, 1993.

[93] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference and prediction*. Springer Series in Statistics, 2nd edition, 2009.

[94] O. Hecker, O. Nelles, and O. Moseler. Nonlinear system identification and predictive control of a heat exchanger based on local linear fuzzy models. In *Proceedings of the American Control Conference*, volume 5, pages 3294–3298. IEEE, 1997.

[95] AJ Heller. Heat-load modelling for large systems. *Applied Energy*, 72(1):371–387, 2002.

[96] M.A. Henson. Nonlinear model predictive control: current status and future directions. *Computers and Chemical Engineering*, 23(2):187–202, 1998.

[97] I. Hindrayanto, S.J. Koopman, and M. Ooms. Exact maximum likelihood estimation for non-stationary periodic time series models. *Computational Statistics & Data Analysis*, 54(11):2641–2654, 2010.

[98] R.A. Horn and C.R. Johnson. *Matrix analysis*. Cambridge University Press, 1990.

[99] S.R. Idelsohn, J.C. Heinrich, and E. Oñate. Petrov–galerkin methods for the transient advective–diffusive equation with sharp gradients. *International Journal for Numerical Methods in Engineering*, 39(9):1455–1473, 1996.

[100] C.M. Jarque and A.K. Bera. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3):255–259, 1980.

[101] Gints Jekabsons. *ARESLab Adaptive Regression Splines toolbox for Matlab/Octave*, 2011. available at http://www.cs.rtu.lv/jekabsons/.

[102] A. Joensen, H. Madsen, H.A. Nielsen, and T.S. Nielsen. Tracking time-varying parameters with local regression. *Automatica*, 36(8):1199–1204, 2000.

[103] T.A. Johansen. Robust identification of Takagi-Sugeno-Kang fuzzy models using regularization. In *Proceedings of the 5th IEEE International Conference on Fuzzy Systems*, volume 1, pages 180–193. IEEE, 1996.

[104] C. Johansson. *Towards Intelligent District Heating.* PhD thesis, Blekinge Institute of Technology, 2010.

[105] R.H. Jones. Maximum likelihood fitting of ARIMA models to time series with missing observations. *Technometrics*, 22(3):389–395, 1980.

[106] GR Jónsson. A model for predicting the yearly load in district heating systems. *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy*, 216(3):277–281, 2002.

[107] T. Kailath. *Linear systems*, volume 1. Prentice-Hall Englewood Cliffs, NJ, 1980.

[108] T. Kailath, A.H. Sayed, and B. Hassibi. *Linear estimation*, volume 1. Prentice Hall, 2000.

[109] R.E. Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

[110] M. Kaltschmitt, H. Hartmann, and H. Hofbauer. *Energie aus Biomasse.* Springer-Verlag, 2009. In German.

[111] Y.H. Kareem and A.R. Majeed. Monthly Peak-load Demand Forecasting for Sulaimany Governorate Using SARIMA. In *Transmission & Distribution Conference and Exposition, Latin America*, pages 1–5. IEEE, 2006.

[112] K. Kato, M. Sakawa, K. Ishimaru, S. Ushiro, and T. Shibano. Heat load prediction through recurrent neural network in district heating and cooling systems. In *Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on*, pages 1401–1406. IEEE, 2008.

[113] P.T. Keenan. Thermal simulation of pipeline flow. *SIAM journal on numerical analysis*, 32(4):1225–1262, 1995.

[114] S.S. Keerthi and E.G. Gilbert. Optimal infinite-horizon feedback laws for a general class of constrained discrete-time systems: Stability and moving-horizon approximations. *Journal of optimization theory and applications*, 57(2):265–293, 1988.

[115] M. Khairy, A.L. Elshafei, and H.M. Emara. LMI based design of constrained fuzzy predictive control. *Fuzzy Sets and Systems*, 161(6):893–918, 2010.

[116] T. Klason and X.S. Bai. Combustion process in a biomass grate fired industry furnace: a CFD study. *International Journal of Progress in Computational Fluid Dynamics*, 6(4):278–286, 2006.

[117] L. Kölbl. Prognose der Heizlast in Fernwärmenetzwerken. Master's thesis, Vienna University of Technology, 2012.

[118] S.J. Koopman. Exact Initial Kalman Filtering and Smoothing for Nonstationary Time Series Models. *Journal of the American Statistical Association*, 92(440), 1997.

[119] J. Koppejan and S. van Loo. *The handbook of biomass combustion and co-firing.* Routledge, 2012.

[120] U. Kruger, J. Zhang, and L. Xie. Developments and applications of nonlinear principal component analysis–a review. *Principal manifolds for data visualization and dimension reduction*, pages 1–43, 2008.

[121] M. Kubat. Decision trees can initialize radial-basis function networks. *IEEE Transactions on Neural Networks*, 9(5):813–821, 1998.

[122] D. Kuzmin. A guide to numerical methods for transport equations, 2010.

[123] D. Kwiatkowski, P.C.B. Phillips, P. Schmidt, and Y. Shin. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54(1-3):159–178, 1992.

[124] M. Langthaler, E. Plunger, A. Walzer, F. Raab, M. Prosenbauer, W. Löffler, and Haneder H. Biomasse-ressourcenpotenzial in Österreich. *Brainbows Informationsmanagement GmbH*, 2007. In german.

[125] H.V. Larsen, B. Bøhm, and M. Wigbels. A comparison of aggregated models for simulation and operational optimisation of district heating networks. *Energy conversion and management*, 45(7):1119–1139, 2004.

[126] H.V. Larsen, H. Pálsson, B. Bøhm, and H.F. Ravn. Aggregated dynamic simulation model of district heating networks. *Energy conversion and management*, 43(8):995–1019, 2002.

[127] M. Lawrynczuk and W. Tadej. A computationally efficient stable dual-mode type nonlinear predictive control algorithm. *Control and Cybernetics*, 37(1):99, 2008.

[128] H.S. Lee and P.L. Siklos. Unit roots and seasonal unit roots in macroeconomic time series: Canadian evidence. *Economics letters*, 35(3):273–277, 1991.

[129] B.P. Leonard. A stable and accurate convective modelling procedure based on quadratic upstream interpolation. *Computer methods in applied mechanics and engineering*, 19(1):59–98, 1979.

[130] B.P. Leonard. The ultimate conservative difference scheme applied to unsteady one-dimensional advection. *Computer methods in applied mechanics and engineering*, 88(1):17–74, 1991.

[131] BP Leonard. Stability of explicit advection schemes. The balance point location rule. *International journal for numerical methods in fluids*, 38(5):471–514, 2002.

[132] R.J. LeVeque. *Numerical methods for conservation laws.* Birkhäuser, 1992.

[133] R.J. LeVeque. *Finite difference methods for ordinary and partial differential equations.* Society for Industrial and Applied Mathematics Philadelphia, 2007.

[134] W.S. Levine. *The control handbook.* CRC, 1996.

[135] P.A.W. Lewis and B.K. Ray. Modeling long-range dependence, nonlinearity, and periodic phenomena in sea surface temperatures using TSMARS. *Journal of the American Statistical Association*, 92(439):881–893, 1997.

[136] P.A.W. Lewis and J.G. Stevens. Nonlinear modeling of time series using multivariate adaptive regression splines (MARS). *Journal of the American Statistical Association*, 86(416):864–877, 1991.

[137] S. Li and J.M. Hyman. Adaptive mesh refinement for finite difference weno schemes. Technical report, Technical Report LA-UR-03-8927, Los Alamos National Lab, 2003.

[138] W. Li, Y. Yang, Z. Yang, and C. Zhang. Fuzzy system identification based on support vector regression and genetic algorithm. *International Journal of Modelling, Identification and Control*, 12(1):50–55, 2011.

[139] J. Lin and J. Shen. Non-linear modelling of drum-boiler-turbine unit using an evolving Takagi-Sugeno fuzzy model. *International Journal of Modelling, Identification and Control*, 12(1):56–65, 2011.

[140] G.M. Ljung and G.E.P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.

[141] L. Ljung. *System identification: theory for the user*. Prentice Hall, 1999.

[142] A. Loewen. *Entwicklung eines Verfahrens zur Aggregation komplexer Fernwärmenetze*. Fraunhofer-IRB-Verlag, 2001. In German.

[143] J.M. Maciejowski. *Predictive control: with constraints*. Pearson education, 2002.

[144] P. Marusak. Efficient predictive control and set–point optimization based on a single fuzzy model. *Adaptive and Natural Computing Algorithms*, pages 215–224, 2011.

[145] P.M. Marusak and P. Tatjewski. Effective dual-mode fuzzy DMC algorithms with online quadratic optimization and guaranteed stability. *International Journal of Applied Mathematics and Computer Science*, 19(1):127–142, 2009.

[146] D.Q. Mayne, J.B. Rawlings, C.V. Rao, and P.O.M. Scokaert. Constrained model predictive control: Stability and optimality. *Automatica*, 36(6):789–814, 2000.

[147] S.F. Meulen. Load management in district heating systems. *Energy and Buildings*, 12(3):179–189, 1988.

[148] S. Mollov, R. Babuska, J. Abonyi, and H.B. Verbruggen. Effective optimization for fuzzy model predictive control. *IEEE Transactions on Fuzzy Systems*, 12(5):661–675, 2004.

[149] S. Mollov, T. van den Boom, F. Cuesta, A. Ollero, and R. Babuska. Robust stability constraints for fuzzy model predictive control. *IEEE Transactions on Fuzzy Systems*, 10(1):50–64, 2002.

[150] S. Mollov, P. Van Der Veen, R. Babuska, J. Abonyi, J.A. Roubos, and H.B. Verbruggen. Extraction of local linear models from Takagi-Sugeno fuzzy model with application to model-based predictive control. In *Proceedings of the 7th European Conference on Intelligent Techniques and Soft Computing (EUFIT '99)*, 1999.

[151] R. Murray-Smith and T.A. Johansen. Local learning in local model networks. In *4th International Conference on Artificial Neural Networks*, pages 40–46. IET, 1995.

[152] K.R. Muske and J.B. Rawlings. Model predictive control with linear models. *AIChE Journal*, 39(2):262–287, 1993.

[153] O. Nelles. *Nonlinear system identification: from classical approaches to neural networks and fuzzy models*. Springer Verlag, Berlin, 2001.

[154] L.N. Neumann, F. Cook, A.W. Western, and K. Verburg. A one dimensional solute transport model for hydrological response units. In *Proceedings of the 18th World IMACS / MODSIM Congress*, 2009.

[155] H.A. Nielsen and H. Madsen. Modelling the heat consumption in district heating systems using a grey-box approach. *Energy and buildings*, 38(1):63–71, 2006.

[156] H.A. Nielsen, T.S. Nielsen, A.K. Joensen, H. Madsen, and J. Holst. Tracking time-varying-coefficient functions. *International Journal of Adaptive Control and Signal Processing*, 14(8):813–828, 2000.

[157] H.B. Nielsen. Methods for analyzing pipe networks. *Journal of Hydraulic Engineering*, 115, 1989.

[158] T.S. Nielsen. *Online prediction and control in nonlinear stochastic systems*. PhD thesis, Department of Mathematical Modelling Technical University of Denmark, 2002.

[159] T.S. Nielsen and H. Madsen. Control of supply temperature in district heating systems. In *Proceedings of the 8th International Symposium on District heating and Cooling*, 2002.

[160] T. Nussbaumer. Combustion and co-combustion of biomass: fundamentals, technologies, and primary measures for emission reduction. *Energy & Fuels*, 17(6):1510–1521, 2003.

[161] I. Obernberger. Decentralized biomass combustion: state of the art and future development1. *Biomass and Bioenergy*, 14(1):33–56, 1998.

[162] K. Ogata. *Discrete-time control systems*, volume 8. Prentice-Hall Englewood Cliffs, NJ, 1995.

[163] N. Paces and M. Kozek. Modeling of a grate-firing biomass furnace for real-time application. In *International Symposium on Models and Modeling Methodologies in Science and Engineering (MMMse 2011)*, volume 6, 2011.

[164] N. Paces, A. Voigt, S. Jakubek, A. Schirrer, and M. Kozek. Combined control of combustion load and combustion position in a moving grate biomass furnace. In *19th Mediterranean Conference on Control & Automation (MED)*, pages 1447–1452. IEEE, 2011.

[165] H. Pálsson. *Methods for planning and operating decentralized combined heat and power plants*. PhD thesis, Technical University of Denmark, 2000.

[166] H. Pálsson, H.V. Larsen, B. Bøhm, H.F. Ravn, and J. Zhou. Equivalent models of district heating systems: for on-line minimization of operational costs of the complete district heating system. *ENS Journal*, 1323, 1999.

[167] T.C. Park, U.S. Kim, L.H. Kim, W.H. Kim, and Y.K. Yeo. Optimization of district heating systems based on the demand forecast in the capital region. *Korean Journal of Chemical Engineering*, 26(6):1484–1496, 2009.

[168] J.Y. Peng and J.A.D. Aston. The state space models toolbox for matlab. *Journal of Statistical Software*, 41(6):1–26, 2011.

[169] P. Pinson, T.S. Nielsen, H.A. Nielsen, N.K. Poulsen, and H. Madsen. Temperature prediction at critical points in district heating systems. *European Journal of Operational Research*, 194(1):163–176, 2009.

[170] D. Popescu and G. Asachi. Simulation of space heat demand. In *6th International Conference on Electromechanical and Power Systems*, pages 243–246, 2007.

[171] S.J. Qin and T.A. Badgwell. A survey of industrial model predictive control technology. *Control engineering practice*, 11(7):733–764, 2003.

[172] J.B. Rawlings and K.R. Muske. The stability of constrained receding horizon control. *IEEE Transactions on Automatic Control*, 38(10):1512–1516, 1993.

[173] J. Richalet, E. Abu, C. Arber, H. Kuntze, A. Jacubasch, and W. Schill. Predictive functional control: application to fast and accurate robots. In *IFAC 10th World Congress, Munich, PRG*, volume 258, 1987.

[174] J. Richalet and D. O'Donovan. *Predictive Functional Control: Principles and Industrial Applications*. Springer, 2009.

[175] M. Roarty. Cogeneration-combined heat and power (electricity) generation. *Science, Technology, Environment and Resource Group, Parliament of Australia*, 1999.

[176] J.A. Roubos, R. Babuska, P.M. Bruijn, and H.B. Verbruggen. Predictive control by local linearization of a Takagi-Sugeno fuzzy model. In *IEEE International Conference on Fuzzy Systems*, volume 1, pages 37–42. IEEE, 1998.

[177] J.A. Roubos, S. Mollov, R. Babuska, and H.B. Verbruggen. Fuzzy model-based predictive control using Trakagi-Sugeno models. *International Journal of Approximate Reasoning*, 22(1):3–30, 1999.

[178] C. Ryu, Y.B. Yang, A. Khor, N.E. Yates, V.N. Sharifi, and J. Swithenbank. Effect of fuel properties on biomass combustion: Part I. experiments-fuel type, equivalence ratio and particle size. *Fuel*, 85(7):1039–1046, 2006.

[179] L. Saarinen. *Modelling and control of a district heating system*. PhD thesis, Uppsala University, 2008.

[180] S.E. Said and D.A. Dickey. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3):599–607, 1984.

[181] G. Sandou, S. Font, S. Tebbani, A. Hiret, and C. Mondon. Predictive control of a complex district heating network. In *Conference on Decision and Control*, pages 7372–7377. IEEE, 2005.

[182] I. Sarbu. Design of optimal water distribution systems. *International Journal of Energy*, 3(4):59–66, 2009.

[183] I. Sarbu and E.S. Valea. Nodal analysis of looped water supply networks. *International Journal of Energy and Environment*, 3:452–460, 2011.

[184] R. Scharler, I. Obernberger, G. Längle, and J. Heinzle. CFD analysis of air staging and flue gas recirculation in biomass grate furnaces. In *Proceedings of the 1st World conference and exhibition on biomass for energy and industry*, 2000.

[185] A. Schuster, M. Kozek, B. Voglauer, and A. Voigt. Grey-box modelling of a viscose-fibre drying process. *Mathematical and Computer Modelling of Dynamical Systems*, 18(3):307–325, 2012.

[186] B.D. Schutter. Minimal state-space realization in linear system theory: an overview. *Journal of Computational and Applied Mathematics*, 121(1):331–354, 2000.

[187] M. Shimada. State-space analysis and control of slow transients in pipes. *Journal of Hydraulic Engineering*, 118:1287, 1992.

[188] B. Skagestad and P. Mildenstein. *District heating and cooling connection handbook*. NOVEM, Netherlands Agency for Energy and the Environment, 2002.

[189] G.A. Sod. *Numerical methods in fluid dynamics: initial and initial boundary-value problems*, volume 1. Cambridge University Press, 1985.

[190] E. Sousa and I. Sobey. On the influence of numerical boundary conditions. *Applied Numerical Mathematics*, 41(2):325–344, 2002.

[191] K. Steer, A. Wirth, and S. Halgamuge. Control period selection for improved operating performance in district heating networks. *Energy and Buildings*, 43(2):605–613, 2011.

[192] V.D. Stevanovic, S. Prica, B. Maslovaric, B. Zivkovic, and S. Nikodijevic. Efficient numerical method for district heating system hydraulics. *Energy Conversion and Management*, 48(5):1536–1543, 2007.

[193] V.D. Stevanovic, B. Zivkovic, S. Prica, B. Maslovaric, V. Karamarkovic, and V. Trkulja. Prediction of thermal transients in district heating systems. *Energy Conversion and Management*, 50(9):2167–2173, 2009.

[194] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions On Systems Man And Cybernetics*, 15(1):116–132, 1985.

[195] S. Takriti, J.R. Birge, and E. Long. A stochastic model for the unit commitment problem. *IEEE Transactions on Power Systems*, 11(3):1497–1508, 1996.

[196] P. Tatjewski. *Advanced control of industrial processes: structures and algorithms*. Springer Verlag, London, 2007.

[197] P. Tatjewski. Supervisory predictive control and on-line set-point optimization. *International Journal of Applied Mathematics and Computer Science*, 20(3):483–495, 2010.

[198] J.W. Taylor and P.E. McSharry. Short-term load forecasting methods: An evaluation based on european data. *IEEE Transactions on Power Systems*, 22(4):2213–2219, 2007.

[199] W.D. Timmons, H.J. Chizeck, F. Casas, V. Chankong, and P.G. Katona. Parameter-constrained adaptive control. *Industrial & Engineering Chemistry Research*, 36(11):4894–4905, 1997.

[200] P. Tkalich. Derivation of high-order advection-diffusion schemes. *Journal of Hydroinformatics*, 8(3):149–164, 2006.

[201] J. Unger, M. Kozek, and S. Jakubek. Reduced order optimization for model predictive control using principal control moves. *Journal of Process Control*, 2011.

[202] I. Vallios, T. Tsoutsos, and G. Papadakis. Design of biomass district heating systems. *Biomass and Bioenergy*, 33(4):659–678, 2009.

[203] L.J.P. Van der Maaten. An introduction to dimensionality reduction using matlab. Technical report, Maastricht University, 2007.

[204] K. Veijonen, T. Vainikka, P. amd Järvinen, and E. Alakangas. Biomass co-firing - an efficient way to reduce greenhouse gas emissions. *EUBIONET - European Bioenergy Networks*, 2003.

[205] M. Verhaegen and V. Verdult. *Filtering and system identification: a least squares approach.* Cambridge University Press, 2007.

[206] H.K. Versteeg and W. Malalasekera. *An introduction to computational fluid dynamics: the finite volume method.* Prentice Hall, 2007.

[207] L. Wang. *Model predictive control system design and implementation using MATLAB.* Springer Verlag, 2009.

[208] Y. Wang and S. Boyd. Fast model predictive control using online optimization. *IEEE Transactions on Control Systems Technology*, 18(2):267–278, 2010.

[209] L. Wasserman. *All of nonparametric statistics.* Springer, 2005.

[210] N.P. Waterson and H. Deconinck. Design principles for bounded higher-order convection schemes–a unified approach. *Journal of Computational Physics*, 224(1):182–207, 2007.

[211] F. Wernstedt, P. Davidsson, and C. Johansson. Simulation of district heating systems for evaluation of real-time control strategies. In *First European Simulation and Modelling Conference*, 2003.

[212] F.M. White. *Fluid mechanics.* McGraw-Hill, Boston, 1999.

[213] X. Wu, L. Liu, Q. Zhu, W. Du, B. Wang, and J. Zhang. U-model-based adaptive control for a class of stochastic non-linear dynamic plants with unknown parameters. *International Journal of Modelling, Identification and Control*, 13(3):135–143, 2011.

[214] Y.B. Yang, C. Ryu, A. Khor, N.E. Yates, V.N. Sharifi, and J. Swithenbank. Effect of fuel properties on biomass combustion. part II. modelling approach-identification of the controlling factors. *Fuel*, 84(16):2116–2130, 2005.

[215] D. Yeh and G.T. Yeh. Computer evaluation of high order numerical schemes to solve advective transport problems. *Computers & fluids*, 24(8):919–929, 1995.

[216] J. Yen, L. Wang, and C.W. Gillespie. Improving the interpretability of TSK fuzzy models by combining global learning and local learning. *IEEE Transactions on Fuzzy Systems*, 6(4):530–537, 1998.

[217] C. Yin, L. Rosendahl, S.K. Kær, S. Clausen, S.L. Hvid, and T. Hille. Mathematical modeling and experimental study of biomass combustion in a thermal 108 MW grate-fired boiler. *Energy & Fuels*, 22(2):1380–1390, 2008.

[218] C. Yin, L.A. Rosendahl, and S.K. Kær. Grate-firing of biomass for heat and power production. *Progress in Energy and Combustion Science*, 34(6):725–754, 2008.

[219] P.C. Young. *Recursive Estimation and Time-Series Analysis.* Springer Verlag, New York, 2011.

[220] I. Zang. A smoothing-out technique for min—max optimization. *Mathematical Programming*, 19(1):61–77, 1980.