Berit Floor Lund

Rigorous simulation models for improved process operation

System analysis and on-line estimation applied to a silicon furnace model

Trondheim, May 2005
Doctoral thesis for the degree of doktor ingeniør
Norwegian University of Science and Technology
Faculty of Information Technology,
Mathematics, and Electrical Engineering
Department of Engineering Cybernetics

Norwegian University of Science and Technology
Faculty of Information Technology, Mathematics, and Electrical Engineering
Department of Engineering Cybernetics
N-7491 Trondheim
Norway

Doctoral Theses at NTNU, 2005:71
ITK Report 2005:3-W

ISBN 82-471-7013-2 (electronic)
ISBN 82-471-7014-0 (printed)
ISSN 1503-8181

Summary

The development of rigorous, mechanistic simulation models for industrial processes is a resource demanding task. These models are often developed for training or engineering purposes. Extended use of such models gives a higher return on the modelling investment, reduces the need to develop and maintain several models, and also increases the chances of survival for the model within the company.

The work undertaken in this dissertation addresses approaches and methods for applying rigorous simulation models to improve the daily operation of process plants. This has been studied using a submerged arc silicon furnace and the existing simulator model Simod as the case. Simod was developed from first principles descriptions of reaction mechanisms and thermodynamics in the process. The model is industrially proven through several years of use.

Silicon furnaces are endothermic chemical reactors with nonlinear, multivariate behavior. The furnaces are to a large extent manually operated. This is a difficult task due to few on-line measurements and the complex behavior.

Two main approaches have been chosen for application of the simulation model to improve daily operation:

- Analysis and extensive simulation studies to characterize and improve the understanding of process behavior
- On-line estimation using the simulation model

Extensive simulation studies and analysis have been carried out in order to characterize the nonlinear and multivariate silicon furnace behavior. The work contributes to the understanding of silicon furnace behavior with compact illustrations of some of the relationships between important inputs and outputs of the process. The analysis especially focuses on the furnace be-

havior around optimal production, and shows how changes in the dominant dynamics can be used to determine the margins to optimality.

The second approach has been to update the Simod model from on-line data. A model which is well adapted to on-line data can provide a means for testing operational strategies, for diagnostics, forecasting, or analysis of historic data.

Emphasis has been put on utilizing as much of the available on-line furnace information as possible, and the dissertation shows how some new measurements can be utilized by the model. In this sense, the model acts as an integration mechanism for on-line process information and provides a possible means for improved interpretation of the information.

A detailed simulation model will in most cases be more detailed than what is needed for estimation. The state vector of the model is not observable, but many of the parameters of the model directly or indirectly affect the mass and energy balances of the model. A parameter estimation strategy for time varying parameters has therefore been chosen.

Many properties of the model as well as the sampling rate of on-line data will determine which estimation methods should be considered. The process sampling rate is 8 hours, which gives ample processing time in the present case. The Simod model architecture and implementation allows for outside control of the model execution, and therefore gradient based methods and iterative methods can be considered in an estimation scheme.

In a large model representing a process with few on-line measurements, many of the candidate variables in the model will be discarded and only a small subset may be included in the estimation algorithm. A parameter ranking method based on orthogonalization of the sensitivity derivative is proposed and applied. The ranking method has been compared to other ranking methods.

A sliding window, nonlinear least squares estimation algorithm has been developed. The estimation criterion contains a term penalizing the deviation from the previous estimate. The weight assigned to this term is updated according to a Kalman-filter like scheme. Applied to industrial data, the estimation algorithm finds estimates for the carbon coverage loss, carbon reactivity, and electrode resistance which lie well within the permissible range of the parameters. The estimation algorithm has been developed to provide flexibility with regards to the number of parameters to be estimated and the number inputs and outputs to be used.

The acknowledgement that the dominant time constant in the response to carbon coverage variations is an indication of the margin to optimality

and over-coking is new knowledge and a contribution to the understanding of silicon furnace behavior. The parameter ranking and sensitivity analysis scheme is a contribution as a method for finding the best parameters for estimation within a large candidate set.

Preface

This dissertation is submitted in partial fulfillment of the requirements for the degree doktor ingeniør at the Norwegian University of Science and Technology (NTNU). The research has been carried out at the Department of Engineering Cybernetics during the period from September 1999 to February 2005, and at the Chemical Engineering Department at Carnegie Mellon University in Pittsburgh from January to June, 2003. The project has funded by the NTNU program "Applied information and communication technology". The project has also received funding from Elkem, SINTEF and the Research Council of Norway.

Returning to studies after having worked for several years is quite a transition. The possibility to attend lectures again, and not least, to have the time to study during office hours felt like a great luxury. It took me a while to change my modus operandi, from being used to responding to numerous interrupts during a day, to reaching the almost meditative state which can be reached when your own thoughts and questions can be followed without any interrupts. But, any doctoral project is a lonely journey, and initial feeling of luxury has been replaced by a number of different emotions during the course of the project. Some of the things I learned are documented in this dissertation. Other things, like months of struggle with tools, interfacing and construction of alternative interfaces to the model, are not directly visible in the dissertation. But the results in the dissertation could never have come about without this effort. Starting a research work is often difficult, to know which end to pick up first, and start working from. Ending it is difficult, also, there are always things that could have been improved.

There are a number of persons who should be thanked for helping me and supporting me in this work.

First of all I want to thank professor Bjarne A. Foss for taking in a doctoral student with quite a few years of work experience. I also want to thank him for his patience and also the impatience which was sometimes needed, and for giving me both encouragement and the criticism along the

way.

I want to thank Elkem for supporting the project and for letting me use the Simod model as a case in the project. There are several persons in Elkem who have contributed greatly to the project. I want to thank Dag Ljungquist and Aasgeir Valderhaug for facilitating the project, and for carrying the project safely through every budget round. This project could however never have been completed without Kjell Ragnar Løvåsen's practical and personal support. Kjell Ragnar has answered an incredible number of questions (some of them several times) regarding the model and the process. He has made necessary changes to the model, and always responded in the same calm, supporting and focused manner. Kjell Ragnar has been the "first line support" in the project and the one I called whenever I needed to throw a new idea at someone. Hopefully the investments and involvement in the project has left both Kjell Ragnar and Elkem with useful knowledge and results, extending beyond what is documented in this dissertation.

I also want to thank professor Birger Erik Ydstie at the Chemical Engineering department of Carnegie Mellon University in Pittsburgh, USA for inviting me to join his research group in the first half of 2003 and for helping me and my family with some of the practicalities of our stay, as well. This was a most interesting period with many fruitful discussions and exchange of ideas regarding the understanding of silicon furnace behavior and control. I also want to thank Erik and professor Larry Biegler for their friendliness, encouragement, and for always being so willing to discuss and help.

Halgeir Ludvigsen at Cybernetica was my on-line, verbal help-system when interfacing Modelfit and Simod. This help was greatly appreciated.

I want to thank my managers at SINTEF for patiently waiting for me to complete my work, and also for supporting the last part of the project financially. I also want to thank all my colleagues at SINTEF and NTNU for answering questions regarding tools and methods. The large research group constituted by the Institute of Engineering Cybernetics at NTNU and SINTEF Applied Cybernetics is maybe one of the largest research groups in control engineering in the world. The value of working in such a large group can not be underestimated.

Finally, I could never have done this without my husband Fredrik's incredible patience and encouragement during all these years. He has also been one of the main sponsors of my doctoral work as he accepted a significant reduction of the family income for several years. Also, I am very grateful to Fredrik and my children Vera and Harald for their support and enthusiasm in moving to Pittsburgh for half a year. It takes a lot of courage

to move to a different school, get new friends and learn a new language.

And last, but not least, I want to thank my mother Gunvor and my late father, Harald, who always told me that education is important.

Trondheim, May 2005 Berit Floor Lund

Publications

The following papers have been published from the project:

Lund, B. F., Foss B. A., Løvåsen, K. R., and Ydstie, B. E. Sensitivity analysis of a dynamic model for submerged arc silicon furnaces. *Proceedings of the Infacon X Congress*, Cape Town, South Africa, 2004.

Lund, B. F., Foss B. A., Løvåsen, K. R., and Ydstie, B. E. System analysis of complex reactor behavior - a case study. In *Proceedings of DYCOPS 7, Cambridge, USA, IFAC* 2004.

Lund, B. F., Berntsen, H. E., Foss, B. A. Methods for parameter ranking in nonlinear, mechanistic models. IFAC World Congress, Prague, Czech Republic, 2005.

The following papers have been submitted for publication:

Lund, B. F., Foss B. A., and Løvåsen, K. R. Analysis and characterization of complex reactor behavior. A case study. Submitted to Journal of Process Control, 2005.

Lund, B. F. and Foss, B. A. Parameter ranking by orthogonalization - applied to nonlinear mechanistic models. Submitted to Automatica as a brief paper, 2005.

Notation and definitions

Mathematical operators

$\mathbb{R}$	-	The set of real numbers
$'$	-	Matrix or vector transposition
$\times$	-	"by" (matrix dimension)
$*$	-	Scalar multiplication
$\langle x, y \rangle$	-	Inner product, $x'y$, between vectors x and y
$\|x\|$	-	Euclidean vector norm, $\|x\| = (x_1^2 + ... + x_n^2)^{1/2}$
$\|x\|^2$	-	$\langle x, x \rangle$
$E(x)$	-	Expected value of random variable x.
$diag\{a, b\}$	-	Diagonal matrix with diagonal entries a and b
$tr(A)$	-	Trace, sum of the diagonal elements of matrix A
$eig(A)$	-	The eigenvalues of A

Symbols introduced in chapter 2 and 3

x	-	Silicon yield, according to (2.1)
C_C	-	Gross carbon coverage
ΔC_C	-	Carbon coverage loss
$C_C - \Delta C_C$	-	Net carbon coverage, $C_C - \Delta C_C = \frac{1+x}{2}$
r_1	-	Carbon type 1 reactivity $(\frac{1}{h\ bar})$
r_2	-	Carbon type 2 reactivity $(\frac{1}{h\ bar})$
$\dot{m}_{C,el}$	-	Electrode carbon mass flow (kg/h)
$\dot{m}_{\mu}$	-	Fuming dust production (kg/h)
$\dot{m}_{\mathrm{Si}O_2}$	-	Quartz feed rate (kg/h)
$\dot{m}_{\mathrm{Si}}$	-	Silicon taprate (kg/h)
$u_{\mathrm{Si}O,tap}$	-	Fraction of SiO fume through taphole
h_{metal}	-	Height set point, Simod tapping controller (m)
h_{top}	-	Height set point, Simod feeding controller (m)

Symbols introduced in chapter 3, continued

i	-	Electrode index, $1, 2, 3$
P	-	Electric power (MW)
$P_{mi,}$	-	Measured power to electrode $i, (MW)$
P_m	-	Total, measured power to electrodes (MW)
P_i	-	Power to electrode i (MW)
R_i	-	Resistance in electrode i
r_e	-	Specific resistance $(\mu\Omega/cm)$
d	-	Electrode diameter (m)
ρ	-	Resistivity $(\Omega\mu m)$
l_i	-	Electrode length (cm)
I_i	-	Electrode current (kA)
ΔP_i	-	Power loss in electrode i
p_{loss}	-	Power loss fraction
ΔP	-	Total power loss in electrodes
P_{ei}	-	Net power, without loss to electrode i
P_{ai}	-	Power to the arc at electrode i
ϕ_i	-	Power fraction to arc
P_h	-	Power to the hearth
P_s	-	Power to shaft
P_c	-	Power to the charge (shaft and hearth)
α	-	Fraction of charge power to the shaft

Symbols introduced in chapters 5 and 6

t	-	Continuous time
T_S	-	Sampling interval
T	-	Discrete time, i.e. actual elapsed time is $t = T_s * T$
k	-	Discrete time index
θ	-	Model parameter vector, n_θ
θ^*	-	True parameter value
$\hat{\theta}_T$	-	Corrected model parameter vector
y_T	-	Process output vector, n_y
$\hat{y}_T$	-	Corrected model measurement vector, n_y
$\bar{y}_T$	-	Predicted model measurement vector, n_y
x	-	State vector
$\hat{x}_T$	-	Corrected state vector at time T
ε_T	-	Vector of model errors at time T, (n_y)
N	-	Sliding window length
J_N	-	Error norm/ minimization criterion using N samples

Symbols introduced in chapter 6, continued

$\frac{dy}{d\theta}$	-	The model sensitivity derivative
ψ_T	-	$\psi_T = \frac{dy_T}{d\theta}$
$\Psi_{T-N+1,T}$	-	$[\psi_{T-N+1}, \psi_{T-N}, ..., \psi_T]'$. Stacked sensitivity derivative from $T-N+1$ to T
$\bar{P}_T$	-	Predicted parameter covariance matrix at T
$\hat{P}_T$	-	Corrected parameter covariance matrix at T
W	-	Measurement noise covariance matrix
V	-	Process noise covariance matrix

Abbreviations and acronyms

COM	-	Component Object Model
DLL	-	Dynamically linked library
MEX-files	-	MEX-files contain dynamically linked subroutines that the MATLAB interpreter can load and execute, Programmed in C or Fortran.

Silicon metal production terminology

Hearth	-	Lower, hottest part of the silicon furnace
Shaft	-	Upper, cooler part of the silicon furnace
SiO_2	-	Silicon dioxide, quartz
SiC	-	Silicon carbide
Carbon coverage	-	Stoichiometric ratio (%) of carbon vs. quartz
Over-coked furnace	-	Condition with net SiC build-up in the hearth
Under-coked furnace	-	Condition with net consumption of SiC in the hearth, no build-up.
Silicon yield	-	The fraction of silicon recovered as metal out of silicon supplied as quartz
Silicon recovery	-	Silicon yield
Optimal carbon coverage	-	Carbon coverage giving maximum silicon yield

Contents

Chapter 1

Introduction

The work undertaken in this dissertation addresses approaches and methods for applying rigorous simulation models to improve the operation of process plants.

There exists no general methodology for addressing all aspects of this problem area, and the topic has therefore been addressed through a case. Also, it is believed that possible approaches as well as possible gains in process operation will depend heavily on the case. The main challenge by addressing the topic through a case is to be specific with respect to the properties of the case that may be general for any other case, for other similar cases and for this case alone.

An important advantage of exploring this area through a case, is that one is forced to address the full range of topics regarding the case, topics that are not necessarily visible on a general level. Another important reason is the drive and motivation provided to the project when a need and interest for the results have been expressed by an organization.

1.1 Motivation

The motivation for bringing the simulation model closer to the process operation is described in the following.

An industrial company runs a large number of similar process vessels in different plants. The main decisions affecting product yield are made manually, based heavily on human observations and judgement regarding the process since there are few on-line measurements within the process vessel. Lack of on-line measurements combined with the process' nonlinear, multivariate behavior and large time constants makes the process behav-

ior difficult to perceive. Different perceptions on how to best operate the process may therefore exist within in the company.

A detailed, dynamic simulation model of the process exists. The model was developed from first principles descriptions of reaction mechanisms and thermodynamics. Emphasis was put on defining inputs, outputs, and parameters to the process which could be recognized by process operation personnel. The simulator has been verified through off-line use in the company over several years. The simulator is relatively well known throughout the company and has proven to be a useful tool to understand and consolidate the knowledge about the process behavior.

In this setting, a natural question is if and how the simulator can be used to aid and improve daily process operation.

1.1.1 The case

The case process addressed in this dissertation is a submerged arc silicon furnace (Schei, Tuset & Tveit 1998). This is a highly endothermic chemical reactor with complex behavior. The simulation model for the silicon furnace is called "Simod", and was originally developed by Sintef, NTNU and Elkem (Foss, Halfdanarson & Wasbø 2000). Simod was originally intended for training and process studies, and has been used by metallurgists and other process personnel in Elkem over the last 4-5 years. Several process models precede Simod in the company, and Simod can therefore be said to hold more than 30 years of modelling efforts and experience.

1.2 Main approach

Two main approaches have been chosen for using the simulation model to improve daily operation:

- analysis and extensive simulation studies to characterize and improve the understanding of process behavior
- on-line estimation using the simulation model

The first approach is to use the model to study and better characterize process behavior. Since the process is largely manually operated, it is important that the operating personnel is able to interpret the process behavior correctly in order to make good decisions about operational actions.

Extensive studies using the simulation model have therefore been undertaken to examine and describe the steady state and dynamic response

to various inputs to the process. Behavior characterization is useful and important if transformed to mental models which can be applied in daily process operation.

The second approach is to update and adapt the process model to on-line data.

An on-line model would be a useful tool to determine quantities that are not measured directly, or uncertain measurements. Also, an on-line model could be used to analyze process behavior, predict future behavior, test the effect of inputs on the model before applying them to the process, and analyze past behavior through applying historic data in estimation.

Simulation models developed for process studies and training are normally much more detailed than models developed for estimation directly. The model formulation is therefore hardly optimal for estimation purposes. A very relevant question is therefore why one wants to consider a large simulation model for estimation purposes and not use a simplified model, or a black box model instead.

Many of the reasons for considering Simod in an estimation scheme lie in the model's existing role in Elkem, and therefore outside the model itself:

- The model exists, is well known and has a relatively widespread use within the company.
- Substantial resources have been spent in development and verification of the model.
- Extended use increases long-term survival chances for the model.

The fact that the model exists and is in use, and that considerable resources have been spent on development and verification of the model, are important reasons for enhancing the use of the model further. The Simod users have themselves raised the question if the model could be updated by on-line data, to use the model to evaluate and plan the daily operation of the process.

The fact that Simod is made from first principles, i.e. has an explicit representation of the physical phenomena within the furnace, makes the model a very important knowledge repository for the company. The main advantage lies in the possibilities for interpreting and analyzing the simulation outcome since quantities and properties can be recognized from the actual process.

Another important reason is that having several models of a highly complex process may decrease the chances of survival for all of them. Focusing

on one model, and using this one as *the* tool within the company, increases the chances of survival of the model. Modelling is a costly activity, and having one model serving more than one purpose, means increasing the return on an often large investment.

The complexity of Simod speaks against its use for estimation purposes. The model is too detailed to be optimal for estimation purposes, and the whole state vector or parameter vector is not observable/identifiable.

The Simod model can also be characterized as computationally expensive. However, computation time in an estimation setting has to be compared to the sampling interval of the on-line measurements used for estimation. In the case process, only 8 hour average or aggregated values exist. This gives ample time for simulation and estimation. The limiting factor is therefore how long a user is willing to wait for an estimation result once a new data set has arrived. One hour calculation time is achievable using Simod. If this is acceptable for a user, this eliminates computation time as a reason for model simplification. Also, necessary computation time can be altered through estimator design. As computers become increasingly more powerful, this aspect becomes less and less important.

A question often raised when presenting Simod for estimation to an external audience, is why not black-box models or neural networks (Sjöberg, Zhang, Ljung, Benveniste, Delyon, Glorennec, Hjalmarsson & Juditsky 1995) are used instead of this complex and computationally heavy model. One important disadvantage of a black box model is that it does not provide the same opportunities for physical interpretation of the results as first principles models do.

A simplified first principles model could of course be an alternative to Simod. This would also provide the advantage of physical interpretation, and could at the same time optimized with respect to observability and/or identifiability. Simplifying the model is however not necessarily a trivial task. A simplified physical model would also provide yet another model that would need to be maintained in the company.

The conclusion of the initial pros and cons is therefore that the use of Simod in an on-line estimation scheme would give significant advantages for the company. The focus of this dissertation is therefore on the utilization of the existing Simod model.

1.3 Theoretical basis for on-line use of the model

There are not many references to on-line use of rigorous, mechanistic simulation process models in the academic literature yet. However, due to the development in computing power, the number of applications of large models on-line should be expected to grow. Examples can however be found within waste water treatment plants (Jumar & Tschepetzki 2001) and also within gas and oil production. Hyllseth & Cameron (2003) reports that a simulation based operator training and support tool has been implemented estimating pressure, holdup and flow in the pipelines of a subsea gas field. Both papers focus on description of the concept, architecture and integration mechanisms of the total solution, and not on the underlying analysis and estimation methods. Diez, Cortes & Campo (2005) describes the use of an on-line lumped model in connection with an off-line CFD (Computational Fluid Dynamics) model applied to improve on-line monitoring of a pulverized coal boiler. On-line simulation is also a topic in discrete event simulation (Davis 1998) with potential applications within manufacturing systems, traffic analysis, military operations, etc.

The aim of on-line adaptation of any model is to minimize the model error, i.e. the deviation between the model's output vector and the actual output vector. This is done using an estimation method which drives the model outputs in the direction of the actual outputs by adjusting variables within the model. These variables can be parameters and/or states of the model.

The methods used to analyze the use of a large model on-line, and to design and validate an estimation solution generally therefore fall into the system identification, observer design, estimation and optimization domains. These domains overlap, and some readers may claim that one domain is a sub-domain of another. Still, all these terms have been included, since they appear in the titles of text books relevant when searching for methods to address this on-line adaptation problem. The terms and their relevance are explained in the following.

According to Ljung (1999, chap. 1), *"System identification deals with the problem of building mathematical models of dynamical systems based on observed data from the system".* System identification in its widest sense, involves choice of model structure, determination of parameters within the model, and possibly state estimation.

Sjöberg et al. (1995) states that finding a suitable model structure is the is the key problem in system identification. Depending on the degree of prior knowledge built into the model structure, different approaches to

finding a model structure have been given the coding white-box, grey-box or black-box models. According to Sjöberg et al. (1995) white-box models can be built when the underlying system is perfectly known and built into the model. In black-box models no physical insight is built into the model. In the grey-box model category there is some physical insight built into the model, but not everything is known. Sjöberg et al. (1995) subdivides the grey-box category into physical modeling and semi-physical modeling. A physical model is a model structure based on descriptions of the physical laws governing the behavior of the system, but the model may contain parameters which need to be estimated from data. In semi-physical models certain known relationships between input and output signals are built into the model, but the models are of black-box character.

Neural networks is an important area within black-box identification. Within this field, regularization is often used to ensure good identifiability despite over-parametrization, and many different regularization techniques exist (Sjöberg, Ljung & McKelvey 1993), (Johansen 1997), (Tikhonov & Arsenin 1977). Johansen (1998) describes how prior knowledge about the system can be incorporated in the formulation of system identification problems in terms of regularization penalty functions and constraints on the parameter space.

Sometimes the term "system identification" is given a narrower definition than the above, and is used to characterize black-box approaches, and distinguish them from first principles based approaches. The system identification domain is to a large extent based on concepts from probability, statistics and signal processing.

Regardless of modelling approach, there will in most practical cases be unknown or uncertain parameters which need to be estimated. Parameter estimation or identification methods are therefore considered as being a part of the system identification domain. State estimation methods are also considered as being a part of the system identification domain.

The definition of an observer given by Levine (1996, chap. 37) is *"An observer for a dynamic system $\mathcal{S}(x, y, u)$ with state x, output y and input u, is another dynamic system $\widetilde{\mathcal{S}}(\hat{x}, y, u)$ having the property that the state $\hat{x}$ converges to the state of x of the process S, independent of the input u or the state x."* The concept of observers was introduced by Luenberger (1966). The original formulation of an observer anticipated that the states (or some of the states) could be observed without error. An introduction to observers is given in Luenberger (1971) and in the textbook by Kailath, Sayed & Hassibi (2000), which covers linear systems. The term "adaptive

observers" characterizes the situation when unknown parameters need to be estimated in addition to the states, see for instance Marino & Tomei (1995). The term "observer" and observer theory is now closely connected to the nonlinear control theory domain.

The term "estimation" emerges from the realm of probability, statistics and filtering theory. The term "state estimation" is in the control engineering domain closely connected to the Kalman filter which has been one of the key concepts within control engineering. Numerous books and papers exist on this topic. One relatively recent text book on the topic is Brown & Hwang (1997). In his original paper, Kalman (1960) developed this optimal filter for state estimation in a discrete time state space system description. The states and the observations are subject to noise assumed to have a Gaussian distribution. The Kalman filter is also used for combined state and parameter estimation, called the "augmented Kalman filter", in which case the Kalman filter state vector is augmented with the parameters and estimated along with the states. Kalman filters are also used in connection with nonlinear models. The nonlinear model is approximated through a first order Taylor linearization of the model. The state estimates are then found through applying the Kalman filter equations on the linearized model, and the solution is called an "extended Kalman filter". The Kalman filter can also be deduced from an observer description, as shown in Kailath et al. (2000).

The book by Graham & Sin (1984) has a relatively extensive treatment of parameter estimation in deterministic and stochastic systems using continuous and discrete models. The focus of the book is adaptive systems, and parameter estimation as an integral part of this. From this point of departure, the Kalman filter is called an "optimum transient observer". "Stochastic observer" is also used as a term for a state estimator. Along this line, an augmented Kalman filter would be called a "stochastic adaptive observer".

Optimization methods are also a key technology within the field of control and estimation. The use of least squares method to determine unknown or uncertain parameters in a model can be regarded as undergraduate curriculum in control engineering. Optimal control is another well-known area within control engineering, see for instance Bryson & Ho (1975).

During the 1990s several research groups have worked with constrained state estimation in finite, moving windows. This is called moving horizon estimation (MHE), and is closely connected to and also inspired by the large research activity in receding horizon control (also called moving horizon con-

trol, or model predictive control, MPC). Allgöwer, Badgwell, Qin, Rawlings & Wright (1999) gives a tutorial and extensive review of receding horizon control and estimation research up to 1999. The publication contains an extensive reference list. The research group of professor Rawlings at the University of Madison, Wisconsin, USA has been one center for activity within receding horizon estimation. Rao, Rawlings & Lee (2001) discusses theoretical and practical challenges concerning linear moving horizon estimation, and also establishes sufficient conditions for stability. A relatively recent publication, Rao, Rawlings & Mayne (2003), addresses constrained state estimation for nonlinear discrete systems. The paper handles the problem through linearization and establishes sufficient conditions for asymptotic and bounded stability.

1.4 Steps in the design of an on-line model

In Levine (1996, chap. 58) and Ljung (1999, chap. 1), the tasks involved in system identification are illustrated as block diagrams. The diagrams are slightly different, but contain many of the same activities. The "system identification process" block diagram in Ljung (1999, chap. 1) starts with prior knowledge as input and uses this for experiment design, choice of model set, and choice of criterion to fit the model structure to actual data. The model is validated, and if it is not accepted, one or more of the previous steps need to be repeated.

The case model in the present work fall into the grey-box category of "physical models" since it has been built from first principles descriptions of mass transportation, reaction mechanisms and thermodynamics. In relation to the block diagrams describing the "system identification process", our model structure has already been chosen since an existing model is used. The focus is therefore not on building the model, but rather to determine which parameters and/or states should be estimated in the model, and to choose a method for doing this. These steps can be regarded either as a part of the model validation procedure of Ljung (1999, chap. 1), or as tasks succeeding the model validation block.

The starting point in the present case is that a large, relatively detailed and computationally heavy model exists for a process with complex behavior.

The steps taken to bring the Simod model on-line have been:

1. Determine which process measurements should act as inputs and outputs in the model.
2. Evaluate the model implementation for estimation purposes.
3. Determine candidate quantities for estimation through process and model analysis.
4. System analysis of the model, steady state and dynamic behavior to inputs and possibly parameters.
5. Sensitivity analysis of candidates for estimation.
6. Estimation algorithm class choice, criterion design, and implementation.
7. Tune and run estimation algorithm using actual process data.

In complex processes, such as the silicon furnace, the available on-line measurements from the process need to be evaluated with regards to independence and dependence to determine which measurements should act as inputs and outputs to the model.

A number of more practical aspects regarding the model implementation should be considered early in the design process, since they may have a significant impact on the design of an estimation scheme. A very important issue is which possibilities the architecture and implementation of the model provide for outside control of the execution. Some estimation methods require rerunning of the model from an arbitrary state vector. If this is not supported by the model implementation, the selection of estimation methods will be limited. Also, if more complex algorithms are to be used for estimation, the interfacing between the model and the estimation algorithm software package may be a challenge. The work load associated with these challenges will most likely vary strongly from model to model, but should not be underestimated.

A large mechanistic model can generally not be assumed to observable or identifiable, and a thorough analysis of the model and process information available is necessary in order to determine which quantities should and could be estimated. It is important to try to assess where the largest uncertainties in the process and model lie, and through this, find candidates for estimation. Important aspects are also the choice of a parameter and/or

state estimation scheme, and to consider how mass and energy balances should be fulfilled during estimation.

A system analysis with regards to the output responses to important inputs and parameters in the model is often necessary. The system analysis is an important basis for interpreting the sensitivity analysis and estimation results.

An important aspect, when the choice of a mechanistic model for estimation has been made, is that the estimation scheme designer is required to have a relatively in-depth understanding of the process behavior. This is required to make the best choices under way in the design, and also in order to interpret and evaluate the estimation results. Obtaining sufficient process knowledge may be a relatively laborious task if the process behavior is complex.

1.5 Dissertation overview

Since the dissertation is clearly case focused, a presentation of the case is given early, in chapters 2 and 3. Chapter 2 contains an overview of the fundamentals of silicon furnace chemistry, furnace instrumentation, as well as aspects of furnace operation. Chapter 3 gives an introduction to the Simod model and discusses how the Simod model can be used for on-line estimation, cf. steps 1-3 in the list above.

Chapter 4 contains a system analysis of the model. A steady state and step response analysis of the model is a natural first step of a sensitivity analysis when a model is to be used for estimation. This corresponds to step 4 in the list above. Such an analysis is also interesting from a process understanding point of view, and chapter 4 is therefore much more extensive than what would be natural if only estimation was the goal. Carbon coverage, i.e. the stoichiometric ratio between carbon and quartz, is the most important input to control the silicon yield. Silicon yield is the silicon fraction recovered as metal through the tap-hole of the hearth, relative to the silicon fraction supplied as quartz. Therefore, the dynamic response to other inputs has been studied at different carbon coverage levels as well. Figure 4.32 summarizes the steady state relationship between important inputs and outputs in a compact presentation.

Chapter 5 contains a sensitivity analysis of the candidate parameters identified in chapter 3, cf. step 5 in the list on page 9. A numeric form of the sensitivity derivative is used in the analysis. Since the model is nonlinear, global results cannot be proven, and the sensitivity derivative is

analyzed locally for many points in the parameter vector space. This requires efficient analysis methods. The parameters have been ranked through successive orthogonalization of the sensitivity derivative. The parameters have also been ranked according to the individual variance contribution of each parameter to the parameter set. A particularly useful form for manual analysis of the sensitivity derivative cross product is also demonstrated. A manual analysis will seldom be sufficient in nonlinear cases, but should still be included to gain initial insight into the problem. The chapter also analyses the effect of different output configurations. Chapter 4 provides an important reference for interpretation of the results in chapter 5.

In chapter 6 the estimation criterion has been designed and an estimation algorithm has been selected, cf. step 6 in the list on page 9. The algorithm is based on a sliding window parameter estimation scheme. The criterion contains simple boundary constraints on the variables and updating of the parameter estimate covariance. An SQP algorithm is used as search algorithm. The estimation scheme has been applied to data from two different silicon furnaces, cf. step 7, and the results are presented in chapter 7.

The results are discussed within each chapter. A conclusion is given in chapter 8.

1.6 Contributions of the dissertation

Figure 4.32 gives a compact presentation of the steady state relationship between the most important input and output variables of the furnace. An earlier version of the figure was published in (Lund, Foss, Løvåsen & Ydstie 2004*a*).

The analysis of the mechanisms ruling the change in the silicon furnace behavior is a contribution. This was published in (Lund, Foss, Løvåsen & Ydstie 2004*b*). The acknowledgement that the dominant time constant in the response to variations in carbon coverage is an indication of the margin to optimality and over-coking, is new knowledge about silicon furnace behavior, see chapter 4. This result has been submitted for journal publication (Lund, Foss & Løvåsen 2005).

The sensitivity analysis methods and the way they are applied in chapter 5 provide a useful scheme for parameter selection in large models. The use of successive orthogonalization in connection with parameter ranking in mechanistic models is a contribution and has been published in Lund, Berntsen & Foss (2005). The formulation of the additional variance con-

tribution as a proposition and a proof in chapter 5 is a contribution in connection with parameter ranking. This, and a more extensive presentation of the sensitivity analysis in chapter 5, has been submitted as a journal publication in Lund & Foss (2005).

Chapters 6.5 and 6.6 contain minor contributions to the practical application and implementation of sliding window, nonlinear least squares parameter estimation schemes.

Chapter 2

Process Description

The purpose of this chapter is to give an introduction to silicon metal production, silicon furnace chemistry and operation.

2.1 Silicon metal

Silicon (Norwegian: "Silisium") is the second most abundant element in the earth's crust after oxygen (Schei et al. 1998). In nature silicon is found in oxidized form, SiO_2, more commonly known as quartz, or in other crystallized forms such as flint.

Technically, silicon in solid state is not a metal, but a semiconductor. The term "silicon metal" is used to characterize relatively pure silicon as opposed to for instance ferrosilicon which has an intended content of iron.

Silicon metal is a greyish, very shiny, light and rather brittle material, see figure 2.1. Due to its mechanical properties, the metal is not used in consumer goods directly, but it is an important raw material for several other industries. Examples are production of silicones in the chemical industry and as an important additive in the casting industry.

Production of silicon metal basically implies removal of the two oxygen atoms in the quartz molecule. The production of high purity silicon metal was a prerequisite for the revolution in electronics and computing. Figure 2.2 shows silicon metal, a silicon micro-chip and solar-cells which is another well-known silicon metal product.

According to (Elkem 2004*a*), Elkem in 2003 produced 190.000 metric Tons of silicon metal. This is approximately 20% of the world's production. Elkem has silicon metal furnaces in Salten, Fiskå, Bremanger, Meråker, Thamshavn, all in Norway, and in Alloy in West Virginia in the United

Figure 2.1: Quartz rock and silicon metal.

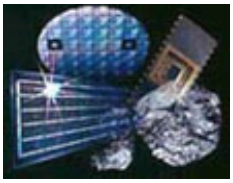

Figure 2.2: Silicon metal products (Elkem 2004*b*).

States.

2.2 Silicon metal production plants

Figure 2.3 gives an overall view of a silicon smelting plant. The raw materials of the process, quartz (SiO_2) and carbon materials, are fed at the top of the furnace and silicon metal is tapped at the bottom. The produced metal goes to solidification and crushing.

The furnace requires large amounts of electrical energy to convert quartz into silicon metal. The energy is supplied through consumable carbon electrodes. The electric power supply typically lies between 10 and $40MW$ and relates to the physical dimensions of a furnace. The height of a furnace is in the range $2.2 - 3.5m$, the internal diameter at the top is $5 - 12m$, and the internal diameter at the bottom is $5 - 9m$. According to Schei et al. (1998, chap 3.2.1), a typical medium-sized silicon furnace would be $20MW$ with a pot diameter of $7m$, a pot depth of $2.7m$, electrode diameter of $1.25m$, and a distance between the electrode centers of $2.6m$.

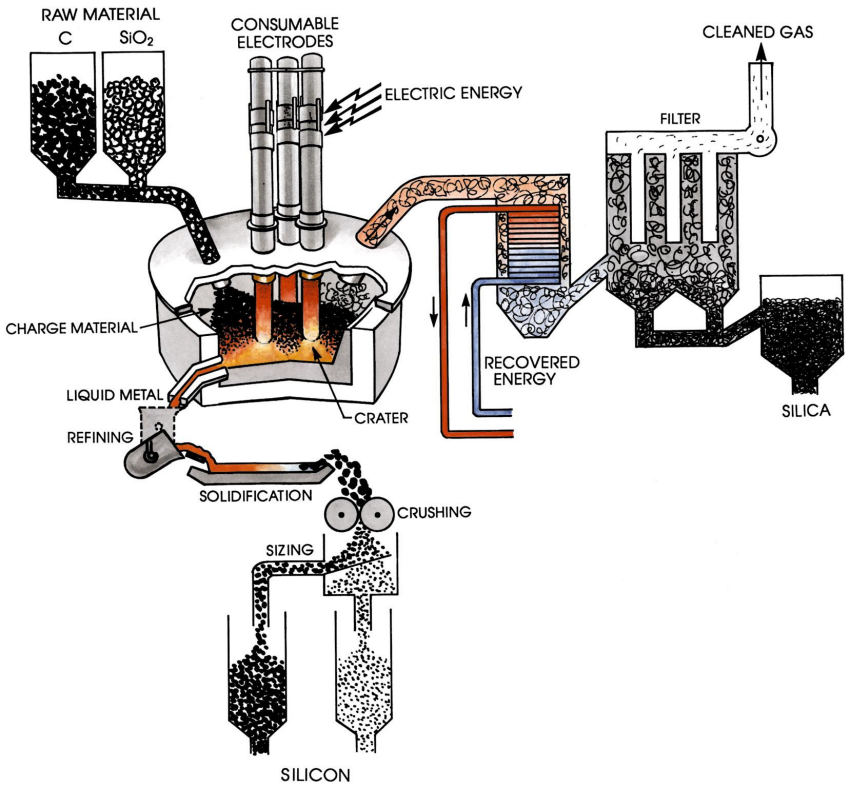

Figure 2.3: A silicon smelting plant (Schei et al. 1998).

The off-gases from a silicon furnace are mainly silicon oxide, SiO, and carbon monoxide,CO. Both components oxidize at the furnace top and in the off-gas system. Figure 2.4 shows the top of a silicon furnace with the sliding gates open. The bright light in figure 2.4 is caused by oxidizing SiO gas. The oxidized SiO fume generates a fine dust called condensed silica fume (Schei et al. 1998), or fuming dust. This dust is captured in a filter. Elkem sells silica dust under the trade name Elkem Microsilica®. Microsilica has proven to be a valuable additive in the production of concrete (Elkem 2002) and has become a valuable by-product of silicon production.

According to Schei et al. (1998, chapter 8.6), only about 32% of the energy consumed in the process remains as chemical energy in the product. Some of the remaining 68% is captured by the water cooling system of the furnace, but the largest fraction of the energy is in the SiO and CO off-gas

Figure 2.4: The top of a silicon furnace (Elkem 2004*b*).

escaping over the furnace top. Heat recovery systems are installed in some plants, and an overview over heat recovery systems in Norway is given in Kolbeinsen, Lindstad, Tveit, Bruno & Nygaard (1995). Larger heat recovery systems may include a turbine and generator for electric energy production in addition the production of hot water for heating of buildings and other purposes.

2.3 Silicon furnace chemistry

A principal sketch of a silicon furnace is shown in figure 2.5. Approximately half of the electric energy supplied through the carbon electrodes is released through electric arcs in a gas filled cavity formed around the electrode tips in the lower part of the furnace. This part of the furnace is referred to as "crater" or "hearth", see figure 2.3. The rest of the electrical energy is supplied through ohmic conduction in the charge surrounding the cavity of the hearth and in the lower part of the "shaft", which is the upper, cooler part of the furnace.

The gross reaction for the whole furnace can be written as (Schei et al.

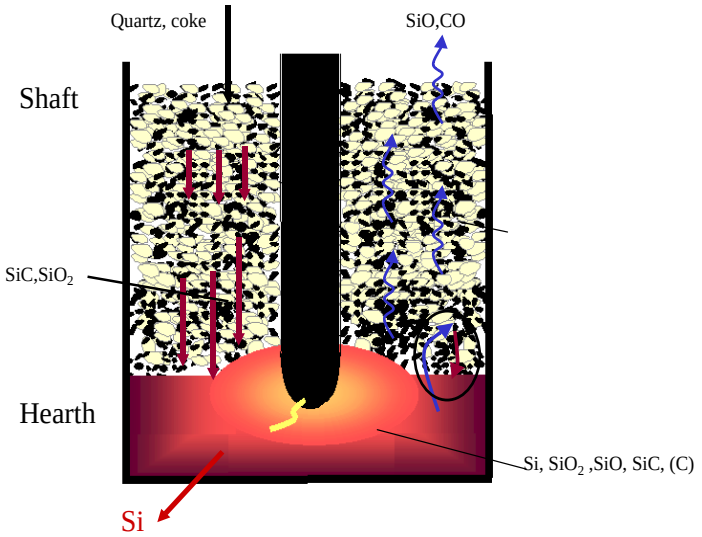

Figure 2.5: Silicon furnace with main reactions and mass flows.

1998):

$$SiO_2+(1+x)C=xSi+(1-x)SiO+(1+x)CO \qquad x \in [0,1] \qquad (2.1)$$

where x is defined as silicon yield, which typically lies around 0.8. As can be seen from reaction (2.1), a higher silicon yield requires a higher fraction of carbon to the reaction.

In reality, a number of reactions take place in order for Si to be formed. The production of silicon oxide gas, SiO, is the "engine" of the reaction chemistry in the furnace. Since SiO_2 is a very stable chemical compound, the formation of SiO gas is a highly endothermic reaction between silicon metal and melted SiO_2:

$$Si+SiO_2=2SiO \qquad (2.2)$$

The reaction mainly takes place in the hearth. Some of the SiO is consumed within the hearth to form silicon metal according to the reaction:

$$SiC+SiO=2Si+CO \qquad (2.3)$$

This reaction requires a sufficiently high partial pressure of SiO as well as a temperature of at least 2000°C, which is obtained by the electric arc (Schei

et al. 1998). Most of the silicon carbide required in the reaction (2.3) is formed in the shaft in an exothermic reaction between SiO rising from the hearth and the carbon feed in the shaft, see reaction (2.4).

$$SiO+2C=SiC+CO \tag{2.4}$$

SiC is a solid, and travels down to the hearth where it is consumed to produce Si according to reaction (2.3). If sufficiently high amounts of unreacted carbon reaches the hearth, reaction (2.4) also takes place in the hearth. If the reaction (2.3) consumes less SiC than what is produced by reaction (2.4), silicon carbide builds up in the hearth. In this situation, the charge is "over-coked".

When the amount of SiC in the hearth is sufficiently high, a particularly hard and nonreactive form of SiC, denoted $SiC\alpha$, may form and deposit in the hearth

$$SiC = SiC_{\alpha} \tag{2.5}$$

causing tapping problems, and in extreme cases, a long-term shut-down of the furnace.

The furnace is normally run under "under-coked" conditions, which means that all the generated SiC is consumed, and there is no net build-up.

Some of the SiO that rises up through the shaft condenses, and travels down with the other solids to the hearth as $Si{\cdot}SiO_2$ condensate. The transportation and condensation of SiO is the main source of heating of the shaft. In the hearth, or lower part of the shaft, the $Si{\cdot}SiO_2$ condensate splits into Si and SiO_2 again.

The heating of the materials in the shaft through SiO condensation provides sufficient heat to melt the quartz. This is another important function of the shaft.

The above reactions are the dominant reactions of the silicon furnace under normal operating conditions, i.e. when the SiO partial pressure is sufficiently high. When the SiO partial pressure is lower, for instance during start-up of the furnace, other reactions dominate (Schei et al. 1998).

2.4 Silicon furnace operation

The operation of a silicon furnace is difficult mainly due to the lack of on-line information about the metallurgical conditions of the furnace, but also due to nonlinear behaviour and long time constants. The information normally available in a silicon furnace is shown in figure 2.6. The measurements are

described in the following in connection with electrical and metallurgical operation.

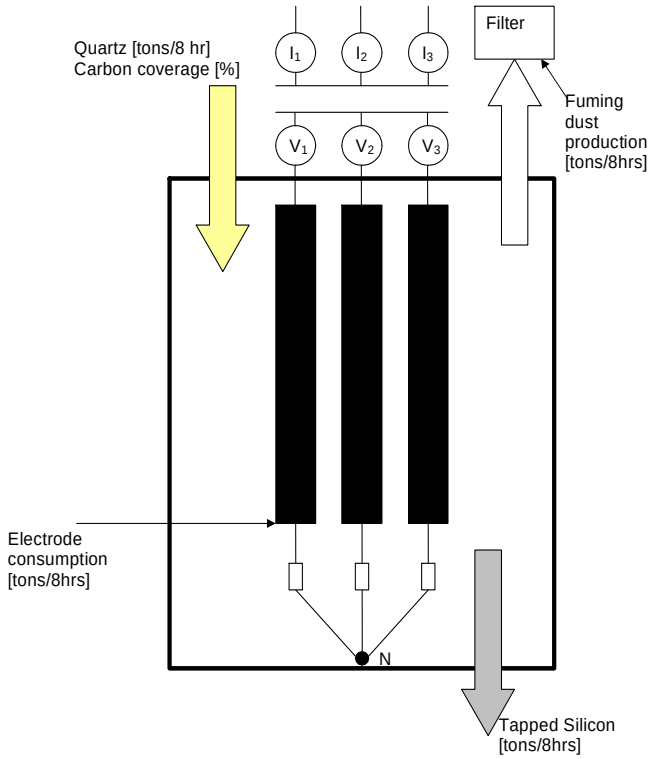

Figure 2.6: Measurements normally available in a silicon furnace.

2.4.1 Instrumentation and control of the electrical system

Electrical power is supplied through three consumable carbon electrodes using alternating voltage, see figure 2.6. The two horizontal, parallel bars in figure 2.6 illustrate the furnace transformer. The electrode currents are measured on the primary side of the furnace transformer, whereas the voltages to each electrode are measured on the secondary side. Based on the available measurements on the individual electrodes, it is possible to calculate the total power supplied through each electrode as well as the total resistance "seen" by each electrode.

The aim of the electrical control is to supply a stable amount of power as far down in the furnace as possible and to maintain a stable electrical

arc between each electrode tip and the metal bath in the hearth. The voltage can only be changed in steps in the furnace transformer, while the current is changed indirectly by altering the resistance to each electrode. The resistance seen by the electrodes is influenced by the length of the electric arc, the composition of materials in the shaft and walls surrounding the cavity in the hearth, and by the temperature and composition of the atmosphere in the cavity. From an electrical point of view, the resistances represented by the arc and charge are in parallel.

Valderhaug (1992) gives an extensive review of work undertaken up to 1992 on models of the resistivity of the charge in silicon furnaces, and on the resistivity of the different components in the silicon furnaces. Based on this, Valderhaug (1992) states that quartz is nearly a non-conductor up to 1800K. Above this temperature the resistivity decreases, but is still much larger than for that of carbon. A high carbon content in the charge therefore lowers the resistance between the electrode and the furnace wall (electric ground) in the shaft. Also, the size distribution of the carbon affects the resistance since a high content of small carbon pieces creates a higher number of inter-particle contacts (Valderhaug 1992). A longer arc gives a higher resistance in the arc. Also, a higher temperature of the hearth lowers the resistance.

The position of the electrodes is often determined by electrical conditions, such as maintaining certain resistance or current conditions in the furnace. The electrical conditions can also be altered to through a "quartzing" operation. Quartzing implies adding quartz close to the electrodes at the top of the furnace. The quartz melts and runs down along the electrode walls. This makes the electrical system "see" a larger resistance in the charge. The electrodes will be lowered by the electrical control system to decrease the resistance in the arc to maintain a chosen total resistance or current in the furnace. This is a stabilizing action since this will cause less energy to be supplied to the shaft, and possibly lower the temperature of the shaft, which again may lower the resistance.

Valderhaug (1992) gives detailed information about how electrical measurements are made and has developed several models in the form of equivalent circuits for the electrical conditions connected to the electrodes. One of the main problems connected to the electrical measurements and control of the electrodes is that currents are induced between the electrodes. Valderhaug (1992) suggest a decoupling strategy for electrical control of the electrodes of a ferrosilicon furnace. More recent work in the field of decoupled current control of ferrosilicon furnaces is provided by Hauksdottir,

Gestsson & Vesteinsson (2002). The paper also cites previous work made by the same authors.

For some furnaces, information about electrode lengths and arc energy fractions may exist, but this type of on-line information is not commonly provided.

2.4.2 Metallurgical measurements and operation

In general there are few on-line measurements of the metallurgical conditions of the furnace. However, mass flow measurements in and out of the furnace exist. The quartz feed to the furnace is weighed and also the silicon production out of each furnace, see figure 2.6. The fraction of tapped, unconverted SiO_2, slag, is also registered. For some furnaces, the fume production is weighed individually. In other situations, furnaces share a filter, and information about individual fume production is not available.

The electrode carbon consumption can be calculated from electrode slip data, electrode geometry, and electrode carbon properties.

At an overall level, the purpose of furnace operation is to maximize the furnace silicon yield, i.e. maximize the metal production and minimize the silica fume flow, and in addition minimize the specific energy consumption (MWh/ton silicon), (Valderhaug 1992). This involves determining the carbon coverage, i.e. the stoichiometric ratio of carbon vs. quartz, the mixture of carbon materials (coal, coke, wood-chips) and deciding the electrical operating point of the furnace. These tasks are normally done by a metallurgist.

The silicon yield is mainly determined by the carbon coverage to the furnace. A 100% silicon yield can never be reached since there is always some loss of silicon through SiO escaping over the furnace top. Maximizing silicon yield through optimizing the carbon coverage at any time is a challenging task since the maximum silicon yield lies at the brink of over-coking. These challenges form an outset for the analysis in chapter 4.

The tasks of the operators include tapping of silicon, charging of raw materials, and stoking at the furnace top. Charging of raw materials may be done manually or automatically. Stoking is a manual operation and implies stirring the raw materials at the furnace top to counteract the formation of gas channels through the charge in the shaft due to the overpressure of SiO in the hearth.

Tapping of silicon may be done continuously through an open tap-hole, giving a near constant silicon metal level in the hearth. Discontinuous tapping is also used, giving higher metal level variations in the hearth. There

are several tapping channels through the furnace lining since metal may be unevenly distributed in the hearth. These tapping channels can be plugged and opened by the operators. Tapping channels often become clogged, and keeping the channels open, or opening channels is one of the tasks of the operator. Tapping channels are opened using a steel rod, oxygen lance, or an electrical graphite electrode (Schei et al. 1998). Sometimes blasting is necessary to open the tapping channels.

Chapter 3

Initial evaluation of Simod for estimation

This chapter evaluates the Simod model for use in an on-line estimation scheme. This corresponds to steps 1-3 in the list on page 9 describing the design steps for an estimation algorithm using an existing, rigorous model.

An introduction to the Simod model is given in section 3.1. The first step in the analysis is to examine the actual inputs/outputs provided as on-line measurements of the process to see if and how these measurements can be applied to the model, see section 3.2. Any model is a simplification of the actual process, and the model can generally only utilize information about phenomena that are actually represented in the model. The section also discusses which outputs are the most important in the adaptation of the model to real data.

The next step is to determine which variables within the model should be altered for the adaptation between model and actual outputs to take place, see section 3.3. The section discusses general aspects such as state versus parameter estimation, and in addition where the largest uncertainties regarding the process and model lie. The purpose of the discussion is to find candidate variables for estimation.

In section 3.4 the preliminary conclusions regarding estimator design are summarized in a figure illustrating the information flow between real data, the model and an estimation algorithm.

The many practical aspect of using an existing, rigorous model for estimation are treated in section 3.5.

3.1 The Simod model, an overview

The Simod model is a nonlinear, dynamic representation of the mass balances, chemical reactions and thermodynamic behavior of the silicon furnace. The model is one-dimensional, which means that no gradients are assumed in the horizontal direction. The model is based on the equations described in the confidential report by Foss et al. (2000). The main mass transportation mechanisms in the model are:

- Solid materials; carbon, quartz and silicon carbide travelling downwards.
- Liquid phase materials; melted quartz, condensate SiO (melted quartz and Si) travelling downwards.
- Gas phase components; SiO and CO travelling upwards.

Once these balances were been developed, the Simod model was written on a Differential Algebraic Equation (DAE) form:

$$\dot{x} = f(x,y,z) \tag{3.1}$$
$$\dot{y} = g(x,y,z)$$
$$h(x,y,z) = 0 \tag{3.2}$$
$$x(t_0) = x_0 \tag{3.3}$$
$$y(t_0) = y_0 \tag{3.4}$$

In (3.1) it is assumed that the dynamic states of the model can be divided into slow and fast modes, x and y. The variable z is an auxiliary variable. The dynamics of the gas, liquid and solid phases have a large span in time constants resulting in a very stiff model. The stiffness problem is handled using the following approach (Foss & Wasbø 2001):

- The gas dynamics is ignored.
- y and z are computed prior to the integration by exploiting the structure of the model equations.
- The dominant dynamics (dynamics of the solids) is integrated using a numerical integration scheme for non-stiff systems.
- The furnace model is divided horizontally into compartments.

- Homogenous process conditions are assumed within each compartment.

Ignoring the gas dynamics means that the hold-up of the gas species is ignored. This is done by associating y with the hold-up or concentration of gas species, and by converting the differential equations for the fast dynamics, (3.1), into algebraic equations.

The distributed nature of the process and model has been approached using finite volumes, each with homogeneous conditions, and flows of energy and mass between them. The hearth is defined by one volume which can vary in height. The shaft can be divided into a number of compartments, typically ten. The top shaft volume can vary in height to represent variations in charging. The feeding operation in real furnaces has in the model been simulated using a discrete on-off controller with a dead-band that can be specified by the user. The tapping operation of silicon from the hearth is simulated using another on-off controller maintaining a specified level in the hearth. The feed rate at the top and tap rate at the bottom can also be set via the simulator user interface with the controllers switched off.

The solution scheme for the algebraic equations and the integration scheme for the dynamic part of the model have proven efficient and robust through approximately 5 years of use Foss & Wasbø (2001). Extensive work has been undertaken by Elkem personnel to verify the model's behavior vs. the dynamic process.

The model can be parametrized to represent furnaces of different physical dimensions. The thermochemical data used in the model is provided by a commercial software package.

Any model is a simplification of the reality. Different phenomena in the actual system are either elaborated or simplified depending on the purpose of the modelling activity. The emphasis in the Simod model is on the thermodynamic and chemical phenomena in the furnace. The electric power supply exists as a power input to the model, and a power loss parameter is used so that the model power input can be compared to the actual electrical power input to the furnace. However, no electrical properties such as resistance of the charge or the arc are represented in the model.

Another important simplification of the model vs. the real process is the spacial discretization and the assumption of homogeneous conditions within each vertical compartment of the furnace. In an actual process, the conditions may for instance be very different around each electrode tip. Silicon may build up in one part of the hearth, and being tapped in another.

In addition to this comes the fact that many properties of the model are not measured on-line and constant values are therefore set in the model. Many of these properties appear in the parameter list of the model and will be treated later in this chapter. Therefore, despite its relatively high level of detail, the Simod model is still an idealized version of a silicon furnace.

An overview over the Simod state vector, output variable vector, input vector and parameter vector is given in the following. However, Simod is a proprietary model and all details can not be revealed.

3.1.1 The state vector and output variable vector of Simod

The state vector and output variable vector of Simod contain of the following main groups of quantities:

- Total amount of the chemical compounds in each volume of the furnace model
- Energy in each volume
- Aggregated mass flows in and out of the furnace
- Aggregated electric power supply to the furnace
- Height of the hearth volume, and the top volume of the shaft

In addition, there exist auxiliary states and variables such as the effective carbon reactivity within each volume and conversion rates of the reactions in all volumes.

All these quantities can be accessed over an interface to the model, and are therefore available for use in an estimation scheme.

3.1.2 Input vector of Simod

The input vector to Simod includes:

- Furnace electrical power supply
- Charge rate of SiO_2
- Carbon coverage
- Fraction of carbon II
- Silicon taprate

The model allows for two different carbon types to account for different reactivity by different reduction materials in reaction (2.4), and the fraction of carbon II is therefore an input. If the silicon taprate is controlled externally, and not by the internal controller, the taprate value is given in the input vector.

3.1.3 Parameter groups in Simod

The number of parameters in the model partly depends on the number of shaft elements used in the model. With ten elements in the shaft, there will be 258 parameters. The parameters can be accessed via the model interface. The following parameter groups exist:

- Boundary defining parameters
  - Loss parameters
  - Properties of inputs
  - Physical dimensions
- Internal parameters
  - Internal mass and energy flow parameters
  - Chemical reaction and melting parameters
  - Internal control mechanism parameters
- Simulator control parameters

The boundary defining parameters define the model versus its surroundings. Parameters in this group include the physical size of the furnace and properties of the inputs to the furnace. Other boundary defining parameters are the losses associated with the power to the furnace and also the loss of carbon at the furnace top due to combustion and escape of carbon dust. The vertical energy distribution is represented by the fraction of the power going to the hearth, cf. section 3.2.2. Also, there are parameters to determine the amount of energy supplied back to the charge from the burning of carbon monoxide and silicon oxide at the furnace top. This has an impact on the temperature conditions in the furnace as well as in the furnace off-gas. The taprate of slag (i.e. melted, unreacted quartz) from the hearth is a parameter in the model.

By internal parameters are meant reaction rates, melting temperatures etc. These quantities have typically been determined from physical property data regarding the specific chemical components. There is a large number of internal heat transfer parameters in the model describing the properties of the heat transfer between elements in the shaft and hearth. Other internal parameters are the set-points and dead-band of the charging and tapping controllers of the model.

In addition to these parameters, there are control parameters for switching the feeding and taprate controllers on and off, and parameters to control where the model reads its input data. These parameters are not relevant for on-line updating.

3.2 Inputs and outputs of the silicon furnace. Relevance for Simod

The description of the state, output variable, and parameter vectors give an indication of which phenomena are represented in Simod. By comparing these with the description of the on-line information in section 2.4, the following measurements are directly applicable to Simod:

- Quartz feed rate
- Carbon coverage rate (i.e. stoichiometric ratio of carbon to quartz)
- Taprate of silicon
- Fuming dust production
- Electrode carbon consumption
- Individual power and currents supplied through each electrode

The top five measurements in the list above represent the mass flows in and out of the furnace. Provided the measurements are correct, they should account for the total material balance of silicon. In addition, the following on-line information can be used for indirect calculation of parameters/disturbances to Simod, if available:

- Electrode length (conducting part)
- Energy fraction to arc/charge

The data sampling interval may be different for the different measurements, and also between plants, but here it is assumed that measurements are available as 8 hour averages or aggregated values. Shorter sampling intervals exist, however, especially in the electrical system.

In the following, the input/outputs above have been described in more detail with emphasis on the uncertainties connected to each input/output.

3.2.1 Mass flow measurements used by Simod

The quartz feed rate is regarded as a fairly accurate measurement. Any inert pollutants of the quartz materials are analyzed off-line, and net weights are available.

The carbon coverage rate is measured in per cent of a 100% stoichiometric ratio of carbon versus quartz. A bias must be subtracted from the carbon coverage rate since some carbon dust will be burnt off at the furnace top. The amount may vary, but a typical figure is 13% of the carbon feed supplied at the furnace top. The raw material weights are considered to be fairly accurate.

The taprate of silicon is associated with noise. Tapping is made into a ladle. The ladle is weighed when it is full. If a ladle has been filled across shifts, the silicon production is split between the shifts by recording the starting time and stopping time for the filling and by assuming a constant taprate. The assumption of a constant taprate may be a source of noise in the measurement. Also, there may be more serious tapping problems causing silicon to build up inside the furnace, to be tapped later. This can be regarded as a process disturbance.

The fuming dust production measurement can be used by Simod if the measurement represents one furnace only.

3.2.2 The use of power supply and distribution measurements in Simod

The electric power supply to a furnace is measured as the total power before the transformer, and as individual power to the electrodes after the transformer. Total electric power is an input in the Simod model. Simod also has a parameter expressing power loss as a fraction of the actual supplied power. If the total power is measured before the transformer, then the loss parameter must include loss over the transformer as well as the loss within the electrodes. If the power measurements after the transformer are used, only the losses in the electrodes need to be considered. This opens up a pos-

sibility to express the power loss fraction in terms of information regarding electrode properties. The main advantage of this is that the model will be exposed to the variations in on-line data regarding the electrodes.

Simod also has a parameter for the fraction of the net power going to the hearth. This fraction contains both the energy supplied through the arc as well as the energy supplied through ohmic conduction of the hearth walls. Since on-line information about the energy fraction going to the arc exists in some furnaces, it is possible to express the energy fraction to the hearth as a function of the energy fraction to the arc.

In the following the power loss and energy fraction to the hearth are represented as a function of the electrode lengths, currents and resistivity.

Power loss as a function of electrode length and current

The power loss in the electrodes is caused by the chemical reactions within the electrodes and also by evaporation of carbon. In the following, the power measurement after the transformer is considered. The power loss parameter in Simod, p_{loss}, is given as the fraction between power loss, ΔP, and the total power supplied, P_m

$$p_{loss} = \frac{\Delta P}{P_m} \tag{3.5}$$

with typical values in the range $0.05 - 0.1$.

When the measured power to each electrode, P_{mi}, is used, then the transformer loss does not need to be considered, and the total power supplied is

$$P_m = \sum_{i=1,2,3} P_{mi} \tag{3.6}$$

The total power loss is

$$\Delta P = \sum_{i=1,2,3} \Delta P_i \tag{3.7}$$

Substituting (3.6) and (3.7) into (3.5) gives the following expression for the power loss fraction

$$p_{loss} = \frac{\sum\limits_{i=1,2,3} \Delta P_i}{\sum\limits_{i=1,2,3} P_{mi}} \tag{3.8}$$

The loss in each electrode, ΔP_i, can be expressed as

$$\Delta P_i = R_i I_i^2$$

assuming that the loss resistance is in series with the electrode. The electrode current, I_i, is measured, and the focus is therefore the electrode resistance, R_i, as the unknown parameter. The resistance depends on the material properties of the electrodes in addition to the geometry of the electrode. On-line information about the electrode length exists, and the cross sectional area of the electrode can be considered constant. Under the assumption that the current passes through the whole electrode, and otherwise has homogenous material properties, the resistance can represented as

$$R_i = r_e l_i$$

where r_e has the unit $\mu\Omega/cm$ and l_i is the length of electrode i in cm. If the the material in all electrodes can be assumed to have the same properties, then r_e is the same for all electrodes. The total power loss, ΔP, is then

$$\Delta P = \sum_{i=1,2,3} \Delta P_i = r_e \sum_{i=1,2,3} l_i I_i^2 \tag{3.9}$$

Substituting (3.9) into (3.8) gives the following expression for the power loss fraction

$$p_{loss} = r_e \frac{\sum\limits_{i=1,2,3} l_i I_i^2}{\sum\limits_{i=1,2,3} P_{mi,}} \tag{3.10}$$

leaving the specific resistance r_e as the only unknown parameter.

From an estimation point of view, this is a preferred variable for estimation since r_e more rightfully can be considered a slowly varying parameter than p_{loss}. The reason for this is that r_e under normal conditions will depend on more slowly varying electrode material properties of the electrode, whereas the actual p_{loss} varies with varying currents and electrode lengths.

Since the electrode has a constant diameter, d, the resistivity, ρ (in $\Omega\mu m$), relates to the specific resistance, r_e, to the cross sectional area as a proportionality constant

$$\rho = R_i \frac{\pi \left(\frac{d}{2}\right)^2}{l_i} = r_e \pi \left(\frac{d}{2}\right)^2$$

The resistivity will vary with temperature. Abnormal conditions, such as electrode breakage, may alter the resistivity significantly.

Energy fraction to arc and hearth

Simod has a parameter specifying the energy fraction to the hearth, a, defined as

$$a = \frac{P_h}{\sum_{i=1,2,3} P_{ei}} \tag{3.11}$$

where P_h is the energy supplied to the hearth, and $P_{ei} = P_{mi} - \Delta P_i$ is the net power supplied to electrode i. P_h includes both the arc power and the power supplied through conduction in the hearth walls. Since on-line information about the energy fraction to the arc exists in some furnaces, an alternative expression for (3.11) is derived.

The arc power fraction ϕ_i is defined as

$$\phi_i = \frac{P_{ai}}{P_{ei}} \tag{3.12}$$

where P_{ai} is the power to the arc of electrode i. The sum of the arc power and charge power for all electrodes must equal the sum of the hearth and shaft powers in the model

$$\sum_{i=1,2,3} P_{ei} = \sum_{i=1,2,3} (P_{ai} + P_{ci}) = P_h + P_s \tag{3.13}$$

where P_s is power supplied to the shaft. Therefore,

$$P_h = \left(\sum_{i=1,2,3} P_{ei} \right) - P_s \tag{3.14}$$

The power to the shaft, P_s, is only a fraction, α, of the total power supplied by conduction in the charge

$$P_s = \alpha \sum_{i=1,2,3} P_{c_i} \tag{3.15}$$

where P_{ci} is the power supplied to the charge by conduction from electrode i. Also, using (3.12) and since $P_{ci} = P_{ei} - P_{ai}$, then P_{ci} must equal

$$P_{ci} = (1 - \phi_i) P_{ei} \tag{3.16}$$

Equations (3.15) and (3.16) can be used to express (3.14) as follows

$$\begin{aligned} P_h &= \sum_{i=1,2,3} P_{ei} - \alpha \sum_{i=1,2,3} (1 - \phi_i) P_{ei} \\ &= (1 - \alpha) \sum_{i=1,2,3} P_{ei} + \alpha \sum_{i=1,2,3} \phi_i P_{ei} \end{aligned}$$

The energy fraction to hearth, a, in (3.11) can then be expressed in terms of ϕ_i as

$$a = 1 - \alpha(1 - \frac{\sum_{i=1,2,3} \phi_i P_{ei}}{\sum_{i=1,2,3} P_{ei}}) \tag{3.17}$$

In this expression the fraction of energy to the hearth is a function of α, and the estimation of α is therefore an alternative to estimation of a.

3.2.3 Independent and dependent variables

In a complex process vessel such as the silicon furnace, it may be difficult to determine the inputs and the outputs of the model. Instead it is necessary to determine the independent and dependent variables. Generally, independent variables should be used as inputs to the model, and the dependent variables could be used as outputs for comparison in a parameter estimation criterion. The variables to be considered are the inputs/outputs presented in the previous section, i.e. the quartz feed rate, the taprate of silicon, the microsilica production rate, the power supply and the vertical power distribution.

The classification of the variables made in the following is based on qualitative knowledge about the process and the model and with discussions with Elkem personnel.

Considerations regarding variables

The carbon coverage rate is determined manually and should therefore be considered an independent variable. Electrode carbon is measured, and should be considered an independent variable.

The quartz feed rate is strongly related to the electric power level of the furnace. Manual tuning also shows that it is relatively easy to fit the model quartz feed rate to the measured process quartz feed rate by altering the power loss fraction. Since it can not be claimed that the power level is dependent on the quartz feed rate, the supplied power is considered an independent variable the quartz feed rate a dependent variable.

The vertical power distribution is not measured as needed in the model. However, there is reason to believe that the variations in the power to the arc and charge also cause variations in the silicon taprate and the microsilica fume generation. The power fraction to the hearth can therefore be calculated from logged data according to (3.17), but then α is unknown, and must estimated. The vertical power distribution depends on the position of

the electrode as well as properties of the hearth and charge not modelled in Simod, and should therefore be considered an independent variable.

The silicon taprate is in the model either controlled by a level controller for metal in the hearth, or in an input variable to the model. If the hearth metal level controller is used to determine the taprate, silicon is tapped every time the metal level exceeds a certain level. In practice this resembles continuous tapping through an open tap-hole. In this situation, the taprate depends on the production of silicon within the hearth, which again depends on carbon coverage, power supply etc., and the silicon taprate generated by the controller is a dependent variable.

The tapping controller can be switched off, and an actual, measured taprate can be given. In this situation the taprate is an independent variable. The risk by exposing the model to the measured silicon taprate directly, is that the noise in the measurement causes the simulated hearth to run dry or to overflow. In this situation, there is no guarantee that the model functions correctly.

The fuming dust production is a dependent variable since it will vary as a result of the power supplied to the furnace, carbon coverage, energy fraction to the hearth, etc.

Classification of variables

The conclusion to this is that the independent variables which should be used as inputs to the model are:

- Carbon coverage
- Electrode carbon
- Electric power level
- Vertical power distribution

Dependent variables used as outputs

- Quartz feed rate
- Fuming dust rate

A variables that can be considered both independent or dependent is:

- Silicon taprate

The dependent variables are candidates for use in an estimation criterion, and therefore the quartz feed and fuming dust rates are the primary candidates. If the actual silicon taprate is not used as an input to the model, but instead controlled by the internal controller, then the silicon taprate should also be included as an estimator output. In other words, since the chosen output vector contains all the measurements needed for the total mass balance of silicon, the purpose of our estimator is to make the model silicon mass balance match the real silicon mass balance of the furnace. This is a sensible purpose since the overall aim of the furnace operation is to control the silicon mass balance.

3.3 Estimation by updating what?

Now that the quartz feed rate, the silica fume rate and possibly the silicon taprate have been established as estimator outputs, the next step is to decide which variables are to be corrected within the model to drive the estimator outputs in direction of the actual outputs, cf. step 3 in the list on page 9.

In state estimation, the state vector of the dynamic model is corrected directly. This generally requires that the state vector is observable from the model and the output data at hand. In parameter estimation, the deviation between actual and simulated outputs is used to correct certain parameters in the model. Parameter and state estimation techniques can be applied to the same model. If there are unknown but constant parameters in the model, parameter estimation or identification is often performed off-line using logged data. If there are unknown, but time varying parameters, the parameters can be estimated using recursive estimation techniques. Parameter and state estimation can also be combined. A classic technique is the augmented Kalman filter in which the state vector of the Kalman filter is augmented with the time varying parameters, and estimated along with the states (Brown & Hwang 1997).

In the following, the application of state and parameter estimation to the Simod model is discussed.

3.3.1 State estimation

An observability analysis of Simod model was performed in Ljungquist (2002). The main conclusion from this analysis was that the state vector of Simod is not observable.

One very important aspect in addition to observability is that state estimation implies direct correction of the state vector, i.e. the components

and temperatures calculated by the model. In such a situation it is no longer guaranteed that the mass and energy balance of the model would be fulfilled. Fulfilling the mass and energy balances could be entered as an additional constraint to the estimation problem. However, considering the complexity of the Simod model, it would be difficult to predict the model's behavior when correcting the state vector directly. Other approaches to correction of the mass and energy balance should therefore be sought first.

3.3.2 Parameter estimation

In the Simod case, several of the parameters express uncertainties associated with the mass and energy inputs of the process and model. Adjustments of these parameters affect the mass and energy balance of the furnace, and therefore indirectly the state vector of the model. Adjusting these parameters instead of the state vector directly would allow the model itself to fulfill itsmass balance and energy balances.

In the following, the parameter groups of Simod listed in section 3.2 are described more closely to find individual parameters which may be candidates for on-line estimation.

Reaction and melting parameters

Even though there may be uncertainties regarding some of the reaction and melting parameters, we generally want to avoid adjusting the reaction rates, melting points etc. The reason for this is that these parameters have been determined based on deep physical and chemical insight, and unless uncertainty has been expressed explicitly by the modellers, these parameters are left untouched.

One exception is the reactivity of the carbon materials fed to the furnace. The carbon materials consist of a mix of coke, charcoal and wood-chips. The porosity, size distribution and individual properties associated with these materials affect the resulting reactivity of the mix, which will vary over time. Assessing the effective reactivity of the carbon materials is generally difficult, and many models and methods exist for lab analysis. The most recent work in this field is documented in the dissertation by Myrhaug (2003).

The reactivity is in the Simod model represented as a lumped parameter and is a clear candidate for on-line estimation.

Uncertainties in inputs and loss parameters

Many of uncertainties associated with inputs to the process were explained in connection with the inputs in section 3.2. The loss of power, loss of carbon coverage and the energy fraction to hearth are all time varying and obvious candidates for on-line estimation.

Other losses in the process are direct fuming of SiO from the hearth. The average direct fume loss from the hearth is therefore a candidate parameter for estimation.

The heat loss on the surface of the furnace has been considered for estimation. The conclusion is however that the effect of altering this parameter is non distinguishable from estimating the electrical power loss to the furnace. This loss is therefore kept constant, and the power loss into the furnace instead is considered a candidate instead.

The active reaction cross section is a parameter in the model. It is a quantity with some uncertainty associated with it, but is assumed to not vary much over time. This parameter is therefore set constant.

The carbon electrodes are consumed, and the carbon supply from the electrodes is a parameter in the model. This parameter is however set from on-line information, see section 3.2.2.

Set-points for internal controllers in the model

The silicon metal level in the hearth is assumed to change over time since the tap-holes of the furnace are opened and plugged, and also located in different places. This parameter is therefore a candidate for estimation.

Normally the charge level set-point should not change over time, but there may be variations in the feeding, or other disturbances affecting this level. The charge level (height of the top element of the shaft) is therefore also a candidate for estimation.

3.3.3 Resulting candidate parameter set

From the above discussion, the initial candidate parameter set is:

1. Average specific resistance , r_e
2. Reactivity of carbon materials, r_1
3. Loss of carbon coverage, ΔC_C
4. Direct fume rate from hearth, $u_{\mathrm{Si}O,tap}$

5. Electric energy fraction to the shaft charge (indirectly energy fraction to the hearth), a

6. Charge level set-point, h_{top}

7. If measured silicon taprate is not used as input: Metal level set-point in hearth, h_{metal}

3.4 Summary. Preliminary estimation scheme using Simod.

Figure 3.1 summarizes the considerations in this chapter and gives an overview of the principal information flow between the process, the Simod model, and a parameter estimation algorithm. In the figure, the taprate of silicon is considered an output from the model, and is compared with the actual silicon taprate.

Note that the figure includes all the candidate estimation variables. A final set is determined after further analysis in the succeeding chapters.

3.5 Implementation aspects

A prerequisite for bringing any model on-line is that it is actually possible to interface the estimation tools to the model. This may seem like a trivial task, but is probably the single factor that is most likely to destroy the chances of using a large simulation model for on-line estimation. Using the model for other purposes than intended originally, often implies that the original interface to the model may not be well suited. One important issue is how easy it is to separate the model from its original user interface, and run it from another user interface or other tools. This depends on the architecture chosen for the original model.

Other important issues regarding the on-line use of large simulation models, are the possibilities to expose the model to external inputs, to alter the parameter vector within the model on-line, and the possibilities to start and restart the model from an arbitrary state vector. These aspects significantly affect the choice of possible methods.

The Simod model has been programmed in C++ and is part of a component based simulator structure in which tidy interfaces between the components has been a requirement. The components have a COM-interface, and several components comprise the normal Simon simulator structure,

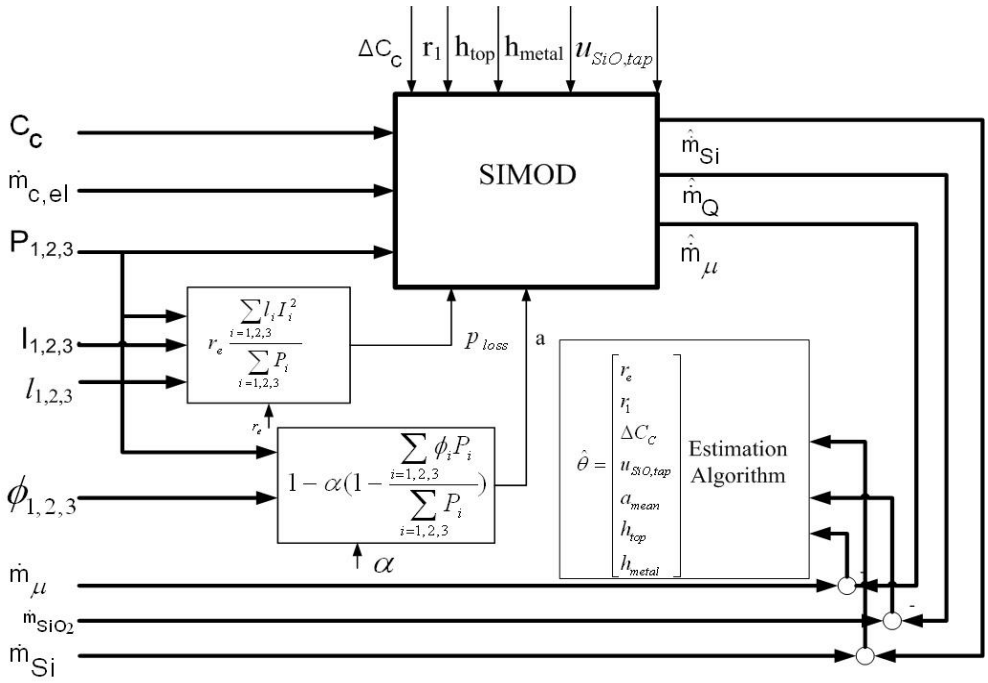

Figure 3.1: On-line Simod concept including all candidate parameters.

see figure 3.2. Commands and data are exchanged between the model and the other components using methods provided by their interfaces. The user interface of the application is shown in figure 3.3. From a user application program, a model can be selected, uploaded, parameters and events can be specified, and the simulation can be started and stopped. The results of the simulations can be plotted, and stored in files.

As a part of the work undertaken for this dissertation, a Matlab-interface based on MEX-functions was developed for the Simod model in order to gain access to the possibilities offered by Matlab for experimentation and analysis of dynamic models, see figure 3.4. This means that the application program in figure 3.2 was replaced by Matlab, and the simulator kernel with a Matlab script. The Matlab script contains calls to MEX-functions which again contain calls to one or more of the methods provided by the model and solver interfaces.

In the Simod case, the model has a very tidy interface, which made interfacing to Matlab a realistic task. This may however not always be

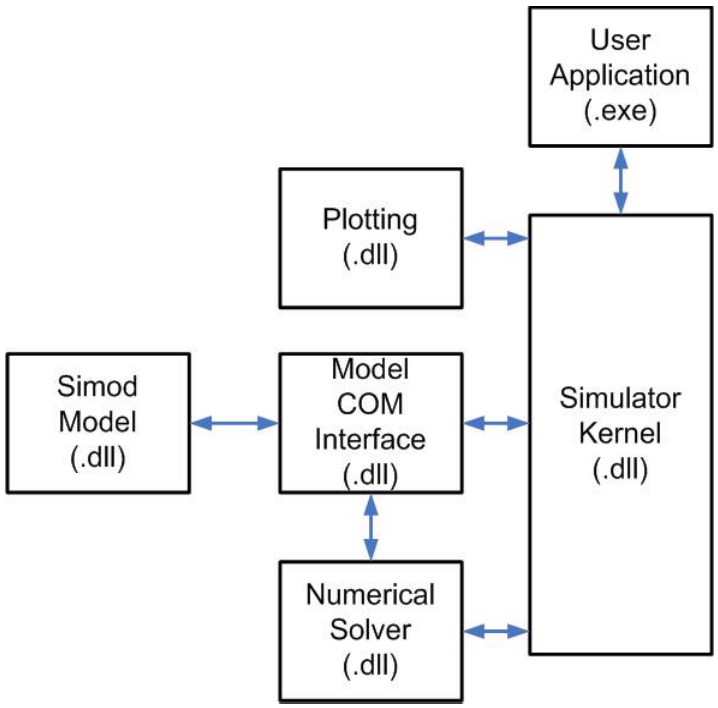

Figure 3.2: The modular structure of the Simon simulator.

case, depending on the architecture and implementation of the model. In any case, the workload associated with making a new interface to a model should not be underestimated. Practical problems such as lack of debugging tools, which can work across different platforms, may make the job time consuming.

One should also be prepared for the need to make at least minor changes to the model. Different uses of the model may cause a need for some new functionality. If the model is large and complex, this is best made by someone with first hand knowledge to the model software. In the Simod case, some changes to Simod were carried out by Elkem.

The experiments and results of chapters 4 and 5 have been obtained using the Matlab interface to the model. Further considerations regarding choice of software for estimation are made in chapter 6.

Figure 3.3: The user interface of Simon.

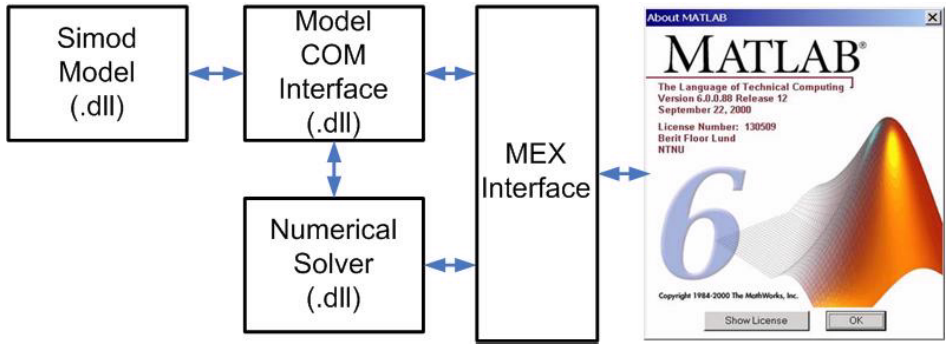

Figure 3.4: The MEX interface enables running Simod from Matlab.

Chapter 4

System analysis

This chapter contains a system analysis of the submerged arc silicon furnace behavior through simulation studies using the Simod model. The focus of the study is on the metallurgical phenomena in the furnace, i.e. on chemical reactions and thermodynamic phenomena.

The study of steady state relationships and dynamic behavior of a model is a natural part of a sensitivity analysis of a model which is to be used for estimation, and this chapter corresponds to the fourth step of the list on page 9. The present study is however much more extensive than what is normally regarded as necessary when estimation is the goal. As the simulations were made, the understanding of the behavior and making compact presentations of the multivariate relationships became an interesting topic in itself. It was seen that the results could contribute to better understanding of the furnace behavior, and this explains the rather comprehensive nature of this chapter. Providing good mental models for the relationship between variables is important since silicon furnace operation depends so heavily on human observations, judgement, and decisions.

This chapter gives a detailed description and analysis of the silicon furnace dynamics with emphasis on variations in the carbon coverage, as this is the most important input to determine silicon yield. Since carbon coverage has such a profound impact on the furnace behavior, section 4.2 is the most comprehensive. The studies of the response to other inputs such as carbon reactivity, section 4.3, overall power level, section 4.4, and vertical power distribution, section 4.5, have be made considering the carbon coverage level of the furnace as well.

4.1 Method for studying the silicon furnace behavior

The model has been simulated to steady state for many input values and combinations of input values to determine the steady state behavior. The resulting curves are presented in different ways to best illustrate the relationship between the variables involved.

For studies of the dynamic behavior of the model, a positive perturbation has been made around a steady state operating point. The size of this perturbation is chosen as 1% of what is considered a natural variation range for the input. This perturbation is assumed to be small enough to represent a linear approximation to the process behavior at the operating point.

For the carbon coverage input, the time constants and gains in the linearized model have been estimated using a least squares method, "lsqcurvefit" in Matlab.

In the real process, measurements of the mass transport of carbon, quartz feed rate, and silicon taprate are normally made as averaged or aggregated values on an 8 hour basis. In this chapter, hourly values are used as outputs.

The studies have been made using a MEX interface to the model, cf. figure 3.4, enabling the execution and control of the model from Matlab scripts.

4.2 Steady state and dynamic response to carbon coverage

This section has three parts. The first part shows the steady state response in various outputs to different levels of carbon coverage using a nominal case. The parameters and inputs used in the nominal case represent a realistic furnace, and are shown in table 4.1.

The carbon coverage input has been simulated to steady state for the gross carbon coverage 93 to 99%. It has been assumed that all carbon is supplied from the top, and none from the electrodes. Under this assumption, a carbon coverage loss of 3% is a realistic figure. The net carbon coverage values supplied to the reactions of the furnace are therefore 90, 91, 92, 93, 94, 95, and 96%. Net carbon coverage values have been used in all tables and figures in this chapter.

Electrode carbon would normally constitute approximately 10% of the supplied carbon to the furnace, and the carbon loss should in that case have

Table 4.1: Nominal input values

Input/parameter name	Symbol	Value
Gross carbon coverage,	C_C	93..99%
Loss of carbon	ΔC_C	3%
Reactivity carbon	r_1	0.56
Net load	$(1 - p_{loss})P$	22 MW
Electrode carbon	$\dot{m}_{C,el}$	0 kg/h
Reaction constant, reaction (2.5)	k_{SiC_α}	0
Feed level controller on		Yes
Tapping controller on		Yes

been set to approximately 13%. Only one type of carbon has been used, and its reactivity has been set to 0.56. The k_{SiC_α} is the reaction constant of reaction (2.5).

No loss of electric power has been assumed, and the 22 MW is therefore a net value.

4.2.1 Steady state gain from carbon coverage to silicon taprate and fuming dust

The steady state gain for carbon coverage to silicon taprate is plotted in figure 4.1. The steady state gain is relatively constant up to 96%. Above 96.5%, silicon carbide starts to build up in the hearth. In a real furnace, the build-up of silicon carbide would cause severe tapping problems. In the simulator, the build-up of SiC gives a higher hearth and a shallower shaft. This causes a negative integrating effect on the taprate in the model, and no steady state gain can be defined above 96.5%, and this region is in this sense "unstable".

The maximal carbon coverage rate one can use and still be in the under-coked operating region, is referred to as "optimal carbon coverage". Above this carbon coverage rate, the furnace will be over-coked. The optimal, net carbon coverage rate lies at approximately 96% for these simulations. The region around 96% is denoted "non minimum phase". This will be demonstrated and explained later in this section.

Figure 4.2 shows the steady state gain from carbon coverage to fuming dust production. As the carbon coverage increases, the SiO gas produced in the hearth will meet more carbon in the shaft and more SiO will be consumed to produce SiC. The steady gain is constant and negative in the under-coked region.

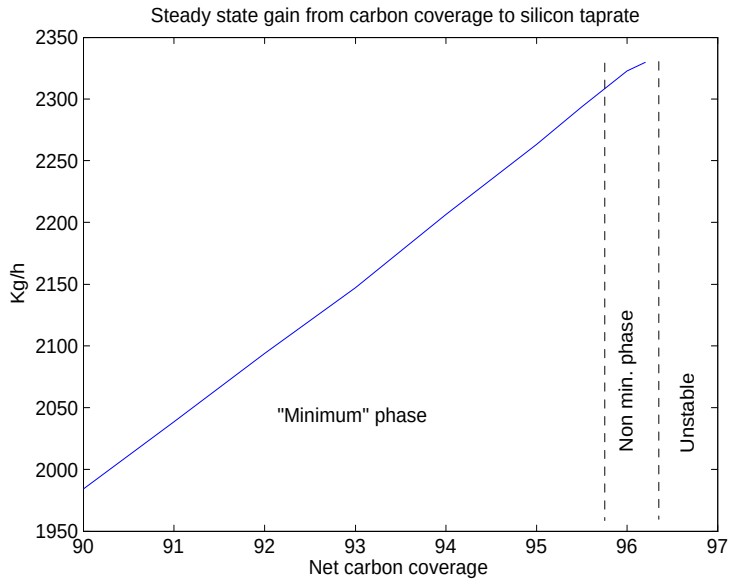

Figure 4.1: Steady state gain from carbon coverage to silicon taprate.

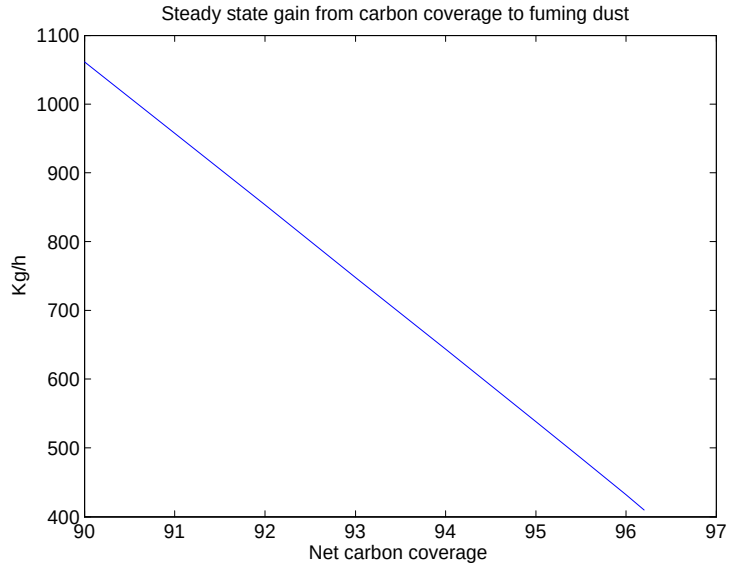

Figure 4.2: Steady state gain from carbon coverage to fuming dust production in the nominal case.

4.2.2 Steady state profiles

In the following the steady state profiles down through the furnace for silicon carbide, silicon oxide and temperature are shown. Element 1 is the shaft top, and element 11 is the hearth in these figures.

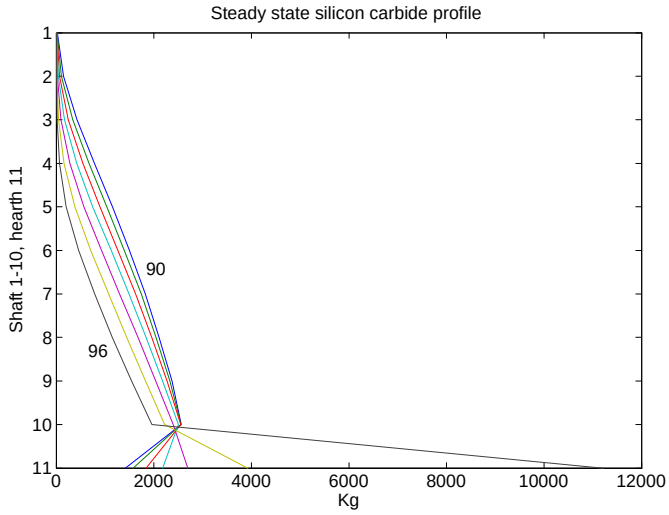

Figure 4.3: Steady state SiC profile for carbon coverage 90 ..96, $r_1 = 0.56$.

The steady state amount of silicon carbide down through the furnace for different carbon coverage values is shown in figure 4.3. The steady state SiO gas profiles are shown in the figure 4.4. Low carbon coverage gives an overall higher SiO partial pressure in the furnace. The steady state temperature profiles in the furnace for net carbon coverage values from 90% to 96% are plotted in figure 4.5.

Figure 4.5 shows that the variation in the hearth temperature (element 11) is small, but that the hearth temperature for 96% carbon coverage seems to cross over the other curves and be the highest. The flat area corresponding to the elements 9 to 6, or 4, corresponds to the condensation temperature of SiO gas. It should also be observed that 96% carbon coverage gives the lowest temperature at the furnace top, whereas 90% gives the highest. The explanation lies in figure 4.4 since the flow of hot SiO gas from the hearth and condensation of SiO gas are the main sources of heating of the shaft. Therefore, the lower flow of SiO at higher carbon coverage values gives a larger temperature difference between the hearth and the furnace top.

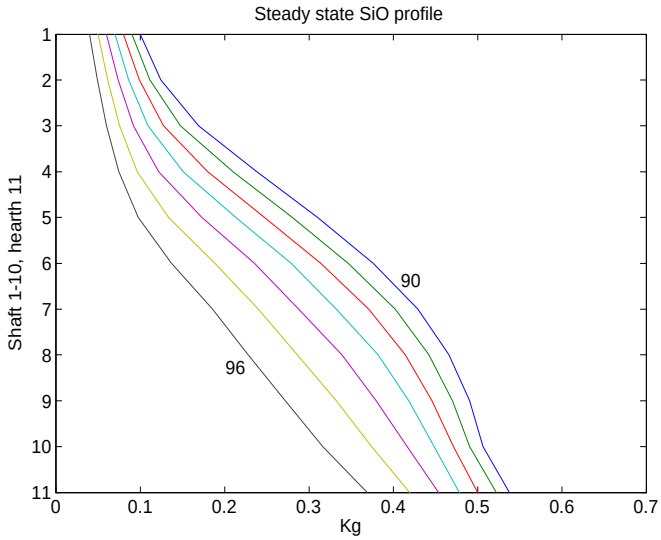

Figure 4.4: Steady state profiles of SiO for different carbon coverage values. r1=0.56.

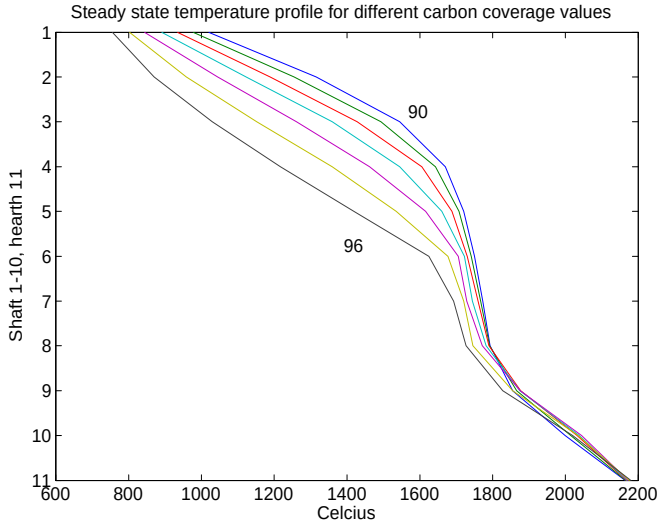

Figure 4.5: Steady state temperature profiles for carbon coverage 90..96. Nominal case.

4.2.3 Dynamic response to carbon coverage

In the following, it has been investigated how the dynamic response to carbon coverage changes at different carbon coverage levels.

The following method has been chosen. At each of the steady state net carbon coverage levels, 90, 91, 92, 93, 94, 95, and 96%, a small perturbation of the carbon coverage level of 0.1% carbon coverage has been made at time 1 hour. A step of 0.1% has been chosen since is represents 1% of the carbon coverage variation range which is assumed to be 10%. The net taprate response has been recorded, i.e. the steady state level has been subtracted from each response. All of these net responses have been plotted together in figure 4.6, so that the dynamic behavior at different carbon coverage levels can be easily compared. Carbon coverage level 97% has also been included. This, however, gives an over-coked furnace, and no steady state taprate could be reached before the step. Therefore, the taprate value before the step has been subtracted from the response to give a net response for this value.

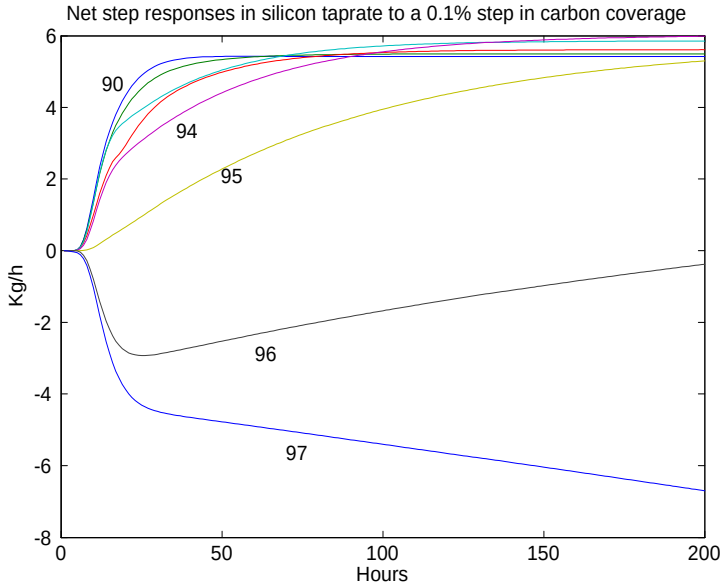

Figure 4.6: Net step responses in taprate Si to a 0.1% step in the carbon coverage at different carbon coverage values. Nominal case.

Figure 4.6 shows that the responses are much faster for small carbon

coverage values than for the higher values. The dominant time constant at a carbon coverage value of 90% lies at approximately 8 hours, whereas for 95% it is approximately 100 hours. There also seem to be some nonlinear effects in the responses corresponding to 92% and 90%. At 96% carbon coverage there is an inverse response in the silicon taprate. This was indicated as a non-minimum phase region in figure 4.1. Since only the response up to 200 hours is included, the figure does not reveal the fact that the response corresponding to 96% has approximately the same steady state gain as the other curves. For 97% there is a net negative integrating effect, and no steady state gain can be defined for this input value.

The net step responses in the fuming dust production corresponding to the responses in figure 4.6 are plotted in figure 4.7. The gain is negative, since increased carbon coverage values will cause more silicon to be recovered as tapped metal and less as dust. The responses seem to be more linear than the silicon taprate responses in figure 4.6, and there is less or no transport delay.

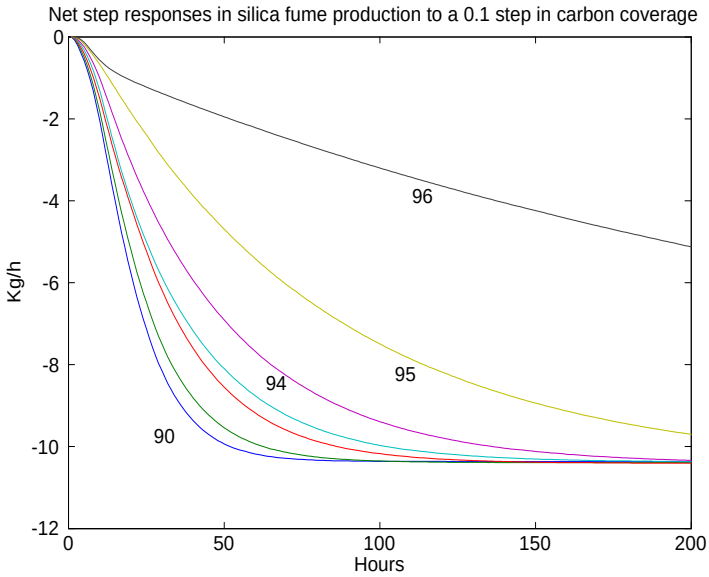

Figure 4.7: Net step response to a carbon coverage perturbation in silica fume production at different carbon coverage levels. Nominal case.

Quantification of the change in dynamic behavior in silicon taprate

The change in dynamic behavior observed in figure 4.6 is quantified using the following approximation

$$H(s) = \frac{Ke^{-\tau s}(1 + T_1 s)}{(1 + T_2 s)(1 + T_3 s)(1 + T_4 s)} \tag{4.1}$$

The gain, transport delay, and time constants have been estimated using "lsqcurvefit" in Matlab, and the results are shown in table 4.2. All the time constants are in hours.

A third order model has proven to give a sufficient approximation. Even though the carbon to taprate response is said to be minimum-phase for small carbon coverage values in figure 4.1, a small transport delay has still been included to capture some of the propagation time for the materials down through the furnace. This delay, τ, is estimated to approximately 2.5 hours for all responses.

Table 4.2 shows that the gain is relatively constant, which is expected according to figure 4.1. The two right-most columns, T_3 and T_4, are the fastest poles of the response. The responses seem to have a double pole in the range $2.5 - 5.0$ hours.

Table 4.2: Zeros and poles in (4.1)

Carbon Coverage	K	τ	T_1	T_2	T_3	T_4
90	54.4	2.5	*	*	4.9	4.9
91	55.5	2.5	31.2	34.6	2.7	7.8
92	56.2	2.5	28.5	34.2	0.5	14.1
93	59.3	2.5	17.1	33.3	2.7	2.7
94	61.7	2.5	16.9	46.7	3.7	3.7
95	58.2	2.5	5.1	78.4	5.0	5.0
96	56.0	2.5	-306.7	688.8	3.1	3.1

T_1 represents a zero and T_2 a pole. For 90% carbon coverage, these two cancel out. For higher carbon coverage values, the time constant of the pole, T_2, increases. The time constant of the the zero, T_1, decreases and finally goes negative. For a carbon coverage value of 96%, the zero is faster than the pole, and since the time constant is negative, the result is an inverse response. The negative integrating effect observed for 97% carbon coverage can be interpreted as the time constant for the pole T_2 having become infinitely large. The development of the dominant time constants

in the pole, T_2, and zero, T_1, have been plotted against the carbon coverage level in figure 4.8.

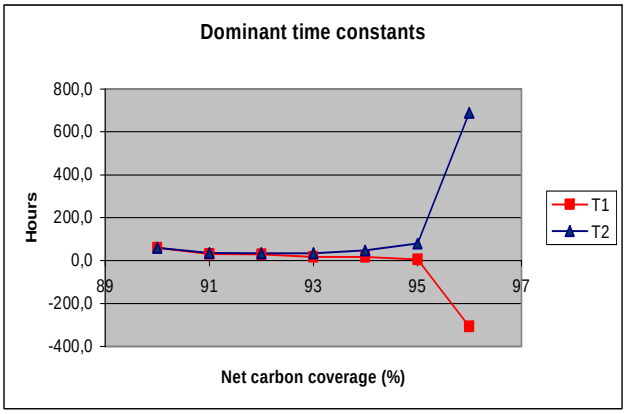

Figure 4.8: Development of the dominant time constants T_1 and T_2 for increasing carbon coverage.

4.2.4 Inclusion of formation of SiC_α

In the previous figures of this chapter, the reaction constant $k_{\mathrm{Si}C_\alpha}$ of reaction 2.5 was set to zero, see table 4.1. In the simulations it was therefore assumed that all the silicon carbide formed could react with SiO to form Si. The formation of SiC_α, according to reaction (2.5), is represented as a threshold effect in the model. The reaction comes into effect in the model when a certain level of unreacted carbon in the hearth is reached.

To study the effect on the net dynamic response when this effect is included, the reaction constant k_{SiC_α} has been set to 0.2. The net step responses from carbon coverage to silicon taprate are shown in figure 4.9. Since all other input and parameter values are the same as in table 4.1, the responses can be compared to those in figure 4.6.

The responses corresponding to carbon coverages 90% to 94% are identical to those in figure 4.6. The reaction creating SiC_α starts when a certain level of carbon in the hearth is reached, at 95% carbon coverage. Since SiC_α in the model is a nonreactive form of SiC, SiC_α starts to build up in the hearth, and the downwards integrating effect occurs.

The main effect of including this reaction is therefore that the slowest modes appearing with $k_{\mathrm{Si}C_\alpha} = 0$ shown in figure 4.6, are replaced with an

integrating effect that occurs at a lower carbon coverage level than without this reaction. The inclusion of this reaction and the way it is modelled and parametrized therefore significantly affects how the model behaves at high carbon coverage levels.

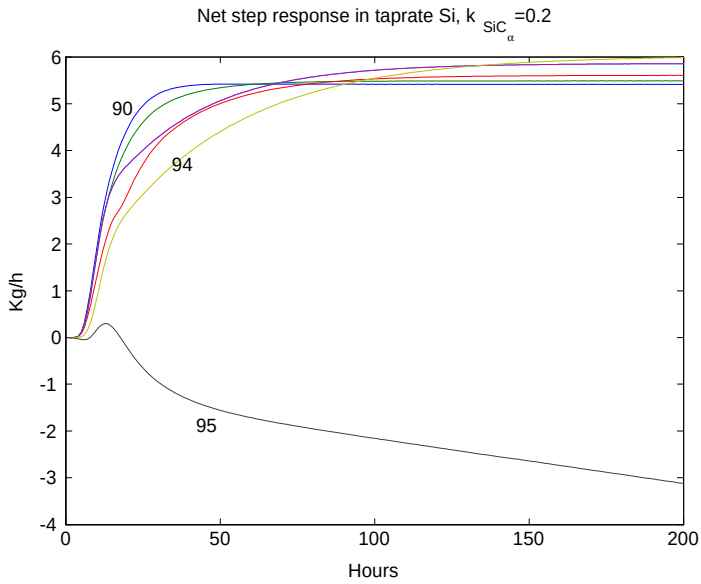

Figure 4.9: Net step responses when SiC_α generation is included. $k_{Si\,C_\alpha} = 0.2$.

4.2.5 Analysis of the variations in carbon coverage response

The changing dynamic behavior of the carbon coverage to silicon taprate responses shown in figure 4.6 and identified in table 4.2, has a dramatic effect on process operation. At low silicon yield it is possible to have a "fast" process, and the effects of altering the carbon coverage can be observed within a shift, at higher carbon coverage rates the response is very slow.

In the following, the causes of the changing dynamic response of the furnace has been analyzed.

Literature

The literature describing the causes of changing dynamics in reactors and integrated plants is the basis for the following analysis.

Changing dynamic behavior exists in many chemical reactors or integrated process arrangements. In Kuhlmann & Bogle (1997) input multiplicity in a van de Vusse is examined. Input multiplicity means that several inputs can give the same steady state solution in the output. This requires a sign change in the gain from the input to the output (Jacobsen 1994). Kuhlmann & Bogle (1997) points out that in some cases there exists a connection between models with input multiplicities and models with non-minimum phase behavior. Both characteristics are caused by competing physical effects on one output variable (Jacobsen 1994). The steady state gain curve shown in figure 4.1 has the sign change in common with a process with input multiplicities. But since silicon carbide builds up, there is a negative integrating effect for large inputs, and no steady state value for the gain exists. If excess silicon carbide could have been "tapped" from the process, the steady state gain curve could possibly be similar to that of the van de Vusse process in Kuhlmann & Bogle (1997). However, this is not possible in practice.

Changing dynamics in a plant may be caused by interconnections in process plants, see for instance Morud & Skogestad (1996). Morud & Skogestad (1996) classifies different process interconnections and gives an overview over the effect each type has on the plant dynamics. On the overall level, the paper distinguishes between external and internal interconnections. By external interconnections are meant interconnections between subsystems associated with different processing equipment. Internal interconnections mean interconnections between phenomena within one process vessel. Within the internal interconnection "class", there exist recycling, parallel paths and series interconnections. Series interconnection is the simplest kind of interconnection and implies that there is a one-way flow between subsystems of material and/or energy. In a system with recycle, mass and/or energy flow is fed back in the process. Recycle generally moves the poles of the process. In which way, generally depends on whether the feedback is positive or negative. In most systems, recycle leads to positive feedback, according to Morud & Skogestad (1996). The third form of internal interconnections is parallel paths, or process "feed-forward". The existence of parallel paths creates or moves the zeros of the process. If the effects of the parallel paths have opposite signs, they are competing effects which may give unstable zero dynamics, or inverse responses for linear systems. Jacobsen (1999) pointed out that recycle also can move control relevant zeros of a plant, and uses reactor and distillation column arrangement as an example.

Process interconnections and plant dynamics

Figure 4.10 shows the mass flows of carbon, quartz, silicon carbide, silicon and silicon oxide into, out of, and within the silicon furnace. The upper box is the shaft, and the lower the hearth. The chemical components are color-coded, and the reactions (2.2), (2.3), and (2.4), see page 17, occur where arrows of different colors meet.

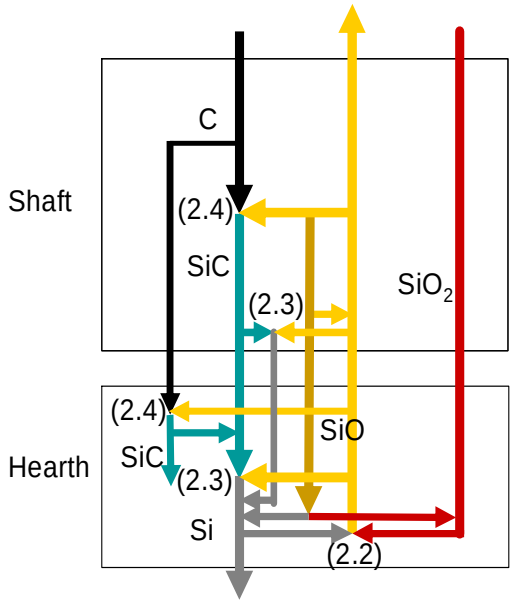

Figure 4.10: Component flows between shaft and hearth.

The red, right-most arrow indicates the flow of quartz. The quartz mainly passes down through the shaft, where it melts, and enters a reaction with silicon, indicated with the grey arrow, and reacts to form silicon oxide, which is the yellow arrow.

SiO is consumed in reactions in the hearth and the shaft. In the shaft, some SiO condenses, and flows down through the shaft and to the hearth again, indicated by the brown arrow. Some SiO escapes over the furnace top.

Carbon is represented by the black arrow. Carbon meets SiO in the shaft to form SiC which is indicated with the green arrow. If carbon reaches the hearth, SiO will be consumed in the hearth to form SiC. SiC is consumed in the reaction with SiO to form Si in the hearth. If the shaft is sufficiently

hot, this reaction may also take place in the lower part of the shaft. If the production of SiC is larger than the consumption, SiC builds up in the hearth. This is indicated with the green arrow which ends in no reaction in the hearth.

Figure 4.10 is complete in the sense that it covers both the under-coked and the over-coked situation. The figure reveals that there may exist several types of interconnections between phenomena and the different types may not be easily distinguished. In order to determine which are the dominant phenomena in the silicon furnace, and through this possibly explain the changes in the dynamic behavior, it is necessary to quantify the steady state conversion rates of the different reactions at different carbon coverage levels.

Steady state conversion levels of the reactions

The upper curve in figure 4.11 shows the steady state conversion rate of the silicon producing reaction (2.4) and the lower curve, the SiO producing reaction (2.2) in the hearth. The difference in generated SiO is approximately 4 kmol/h (twice the increase in the reaction) from the lowest to the highest carbon coverage level, whereas the difference in consumed SiO in the Si producing reaction is almost 10 kmol/h. Since the SiO consumption in the hearth increases more than the production, the SiO level in the shaft will be lower for higher carbon coverage levels.

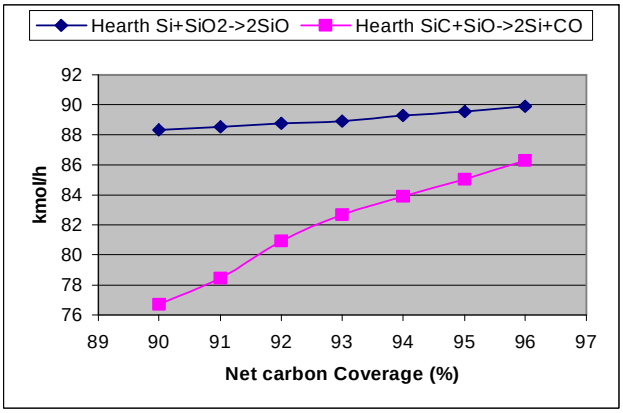

Figure 4.11: Steady state conversion rates for SiO and Si producing reactions in hearth.

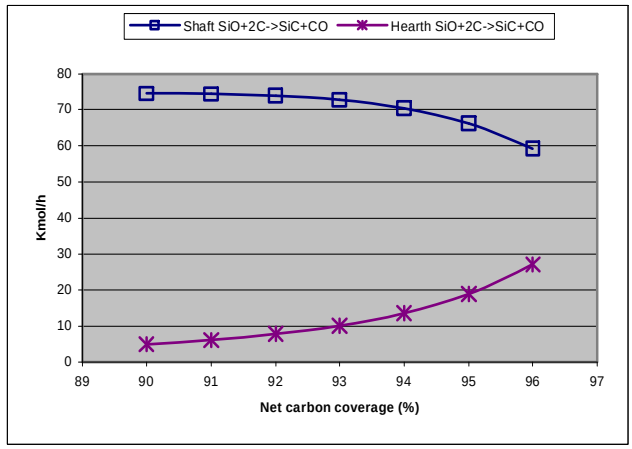

Figure 4.12: SiC generation in the shaft (upper) and hearth (lower).

Figure 4.12 shows the conversion rates for reaction (2.4). The upper curve corresponds to the shaft reaction, and the lower, the hearth. For low carbon coverage rates, most of the SiC is formed in the shaft and little in the hearth. For higher carbon coverage rates, less SiC is formed in the shaft and more in the hearth. The explanation lies in the increased SiO consumption in the hearth due to increased production of Si, as seen in figure 4.11. In addition, the increased production of SiC in the hearth reduces the SiO in the shaft further.

The last reaction which needs to be quantified is the recycling of SiO through condensation. The simulations show that most of this recycling takes place within the shaft itself, and that very little condensate reaches the hearth (0.01-0.02 kmol/h). This is a very small rate compared to the rates of the reactions in the figures above, and the "material recycle" of Si through SiO/SiC therefore dominates. The effect of recycle through condensation of SiO in the shaft is therefore ignored in the following analysis.

Analysis of the dominant mechanisms

By ignoring the recycling of SiO through condensation in the shaft, figure 4.10 can be redrawn emphasizing the response from carbon to tapped silicon, represented by figures 4.13 and 4.14.

Figure 4.13 shows the main reactions when the furnace is run under-coked, and with carbon coverage far below the optimal carbon coverage

value. The reactions to the left of the dashed line take place in the shaft, and the reactions to the right, in the hearth. The reaction creating silicon in the hearth, is the upper right reaction. Any surplus of silicon is tapped. Silicon is "recycled" through a reaction with quartz in the hearth. This reaction forms SiO which is used to form Si in the hearth. The SiO surplus rises up through the shaft and reacts with carbon to form SiC. Unreacted SiO escapes over the furnace top as fuming dust.

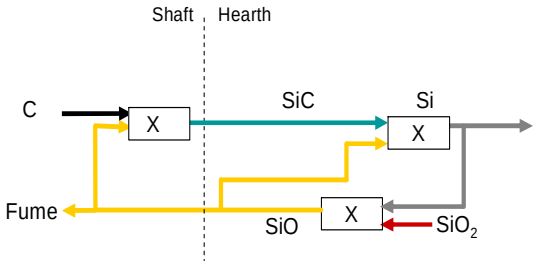

Figure 4.13: Component flows and reactions when carbon coverage is far below optimal value in the under-coked region.

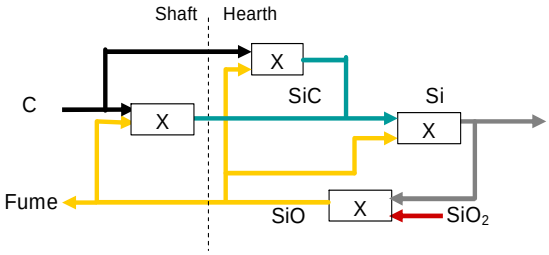

Figure 4.14: Component flows and reactions in the hearth and shaft at close to optimal carbon coverage, still in the under-coked region.

Figure 4.14 shows the component flows and reactions in the under-coked operating region, but with a carbon coverage close to, or at, the optimum value. In this situation, some unreacted carbon enters the hearth. SiC is therefore also formed in the hearth, indicated by the top, middle reaction. According to figure 4.6, higher carbon coverage gives a slower carbon to taprate response.

Parallel and possibly competing phenomena may be the cause of the moving zero. The two SiO consuming reactions in the hearth producing SiC

and Si are obvious candidates since they are exposed to the same partial pressure of SiO, and compete about the SiO. SiC production in the hearth is insignificant for low carbon coverage values according to figure 4.12, but amounts to almost half of the SiC production at 96% carbon coverage. In the low carbon situation, there is a relatively large surplus of SiO in all parts of the furnace. For a high carbon coverage level, the general level of SiO in the furnace goes down, and more unreacted carbon will enter the hearth. In addition, the SiC generating reaction in the hearth will take a more and more significant part of the SiO in the hearth.

One possible reason for the slower response, i.e. the large time constant in the dominant pole at high carbon coverage levels, is that the carbon in the hearth needs time to react to SiC, whereas for low carbon coverage values, the SiC is formed in the shaft and comes down to the hearth "ready" to enter the Si producing reaction. Another possible reason for the slower response at high carbon coverage rates, is the decreased amount of SiO gas in the furnace, see the steady state profiles for SiO in figure 4.4. Since the partial pressure of SiO (above an equilibrium pressure) drives the SiC and Si generating reactions, less free SiO in the furnace will clearly give a slower response in all reactions. It is also possible that the high production of SiC in the hearth also affects the time constant of the pole.

4.3 Steady state and dynamic effects of variations in the carbon material reactivity

In the previous section, the effect of changing the carbon coverage was studied, keeping the reactivity of the carbon constant. In this section, the steady state and dynamic effect of changing the reactivity have been studied.

The following reactivity values have been used, $r_1 = 0.3, 0.4, 0.5, 0.6, 0.7$. Note that in section 4.2, $r_1 = 0.56$ was used, see table 4.1.

4.3.1 Steady state responses to carbon reactivity

For the studies of steady state behavior, simulations to steady state were made for all 35 combinations of carbon reactivity values and carbon coverage values.

Steady state gain in silicon taprate and yield vs. carbon coverage and reactivity

The steady state silicon taprates are shown in figure 4.15 for different reactivity values.

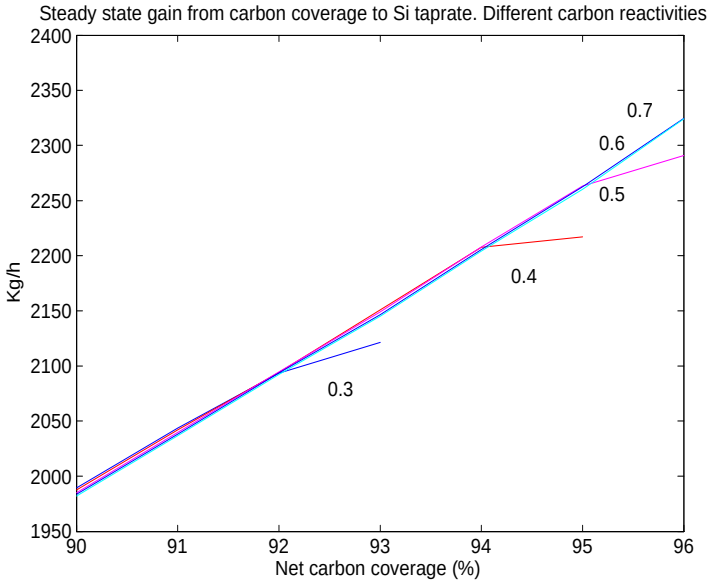

Figure 4.15: Steady state gain from net carbon coverage to silicon taprate.

Figure 4.15 shows that for reactivity 0.3, a carbon coverage between 93 and 94% (the grid is somewhat coarse) causes over-coking. For reactivity 0.4, over-coking occurs above 95%. For reactivities 0.5 and 0.6, the furnace can be run at 96% net carbon coverage and still not be over-coked.

The steady state gain in figure 4.15 shows low steady state sensitivity to reactivity for low carbon coverage rates. For low carbon coverage values, the difference in hourly taprate for the reactivities lies between 0 and 7 kg/hour. The steady state taprate is however very sensitive to the reactivity around the optimal carbon coverage for each reactivity rate.

Figure 4.16 shows the silicon yield versus carbon coverage and for different reactivity rates. The silicon yield is the fraction of silicon recovered as metal relative to the silicon supplied as quartz. The curves show the same tendency as for the steady state silicon taprate in figure 4.15, i.e. that the sensitivity in the yield to reactivity is low for low carbon coverage rates,

but higher close to optimal carbon coverage.

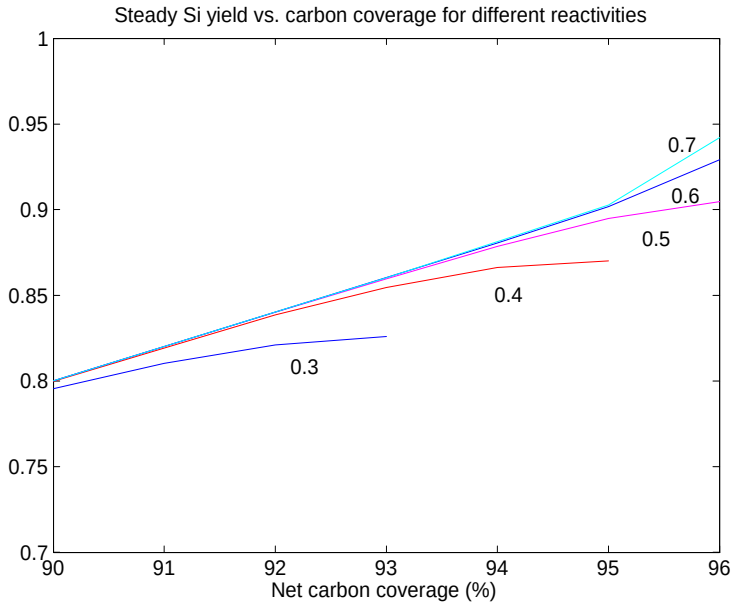

Figure 4.16: Silicon yield vs. carbon coverage for different reactivities.

Steady state temperatures vs. carbon coverage and reactivity

Reaction (2.4) creating silicon carbide is an exothermic reaction. When the reactivity of carbon increases, this reaction is shifted higher up in the shaft, and a higher shaft temperature can be expected. Figure 4.17 shows the steady state temperature at the top of the shaft as a function of both carbon coverage and reactivity. The highest temperature at the top of the shaft is obtained with the lowest carbon coverage value, 90%, and the highest reactivity value, 0.7. The figure verifies that the top of the shaft becomes hotter when the reactivity increases. It also shows that the temperature at the top of the shaft is sensitive to the carbon coverage level.

The lowest temperature at the shaft top can be obtained at the highest carbon coverage, 96%, at reactivities 0.5, 0.6 and 0.7. The lower reactivities at 96% carbon coverage will give a shallower furnace due to large amounts of SiC in the hearth.

Figure 4.18 shows the steady state temperatures in the hearth as a function of carbon coverage and reactivity. The lowest hearth temperature oc-

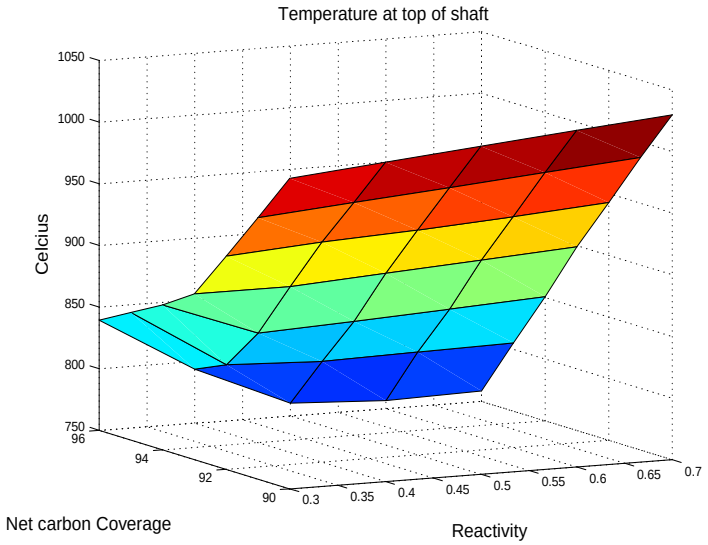

Figure 4.17: Steady state temperature at top of shaft.

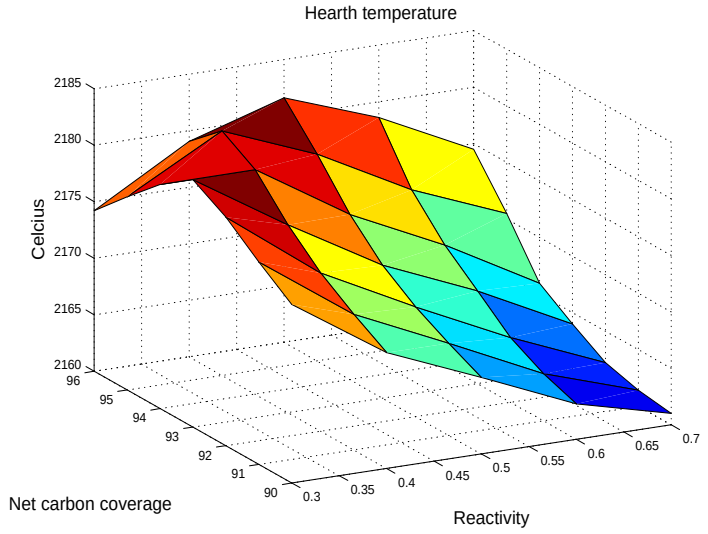

Figure 4.18: Steady state hearth temperature as a function of reactivity and carbon coverage.

curs for the highest reactivity and the lowest carbon coverage value. Under these conditions, the temperature difference between the top at the bottom is smaller than when the carbon coverage is high. The highest hearth temperature occurs at the optimal carbon coverage and is less dependent on the reactivity. At optimal carbon coverage the flow of SiO to the shaft is much smaller than for a very low carbon coverage value, see figure 4.4. This gives a reduced transportation of heat out of the hearth at close to optimal carbon coverage. Instead, the SiO is consumed for production of SiC in the hearth. This is a exothermic reaction. Without going into a quantitative analysis, this is most likely explanation for the higher hearth temperature at higher carbon coverages.

4.3.2 Dynamic response to variations in carbon reactivity

In the following, the dynamic effects of variations in the carbon reactivity are studied. Changes in the reactivity may occur when the mix between carbon materials is changed. Disturbances in the size distribution of the carbon materials may however also occur due to segregation (stratification) in the raw material silos.

Perturbation of the carbon reactivity

Assuming the natural variation range for carbon reactivity lies between 0.3 and 0.7, a 1% perturbation gives a step size of 0.004. Figure 4.19 shows the net responses in the silicon taprate to a positive perturbation in the reactivity at carbon coverage levels at time 1 hour for the net carbon coverage values 90, 91, 92 and 93%. The net responses to the same perturbation for carbon coverages 94, 95 and 96% are shown in figure 4.20. Note that only the reactivity has been perturbed for each steady state carbon coverage rate, the carbon coverage has been kept constant in figures 4.19 and 4.20.

All of the responses are mainly transient effects with little or no steady state effect. This is consistent with the steady state gain curves of figure 4.15. The transient is however slower for higher carbon coverage levels, see figure 4.20.

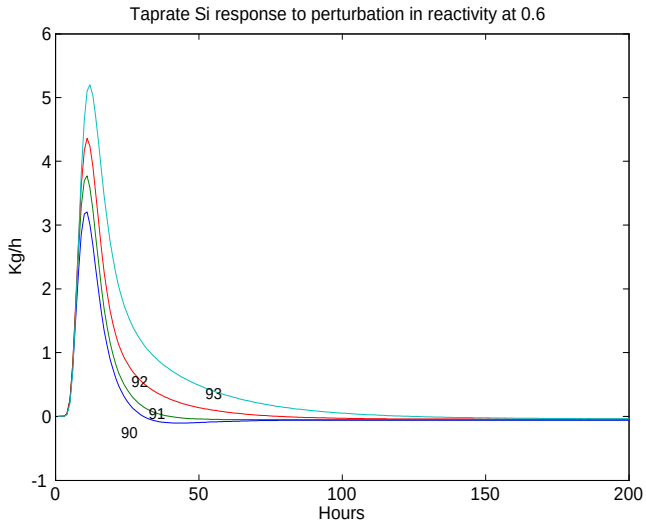

Figure 4.19: Net step response in taprate to a positive perturbation in reactivity at carbon coverages 90-94%.

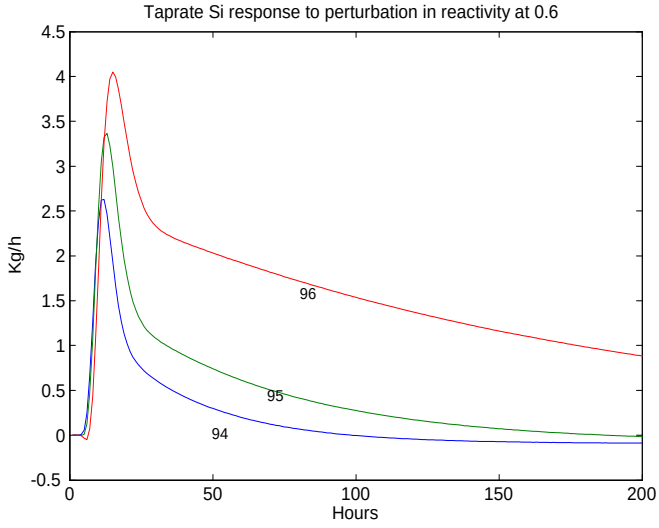

Figure 4.20: Net response in taprate to a positive perturbation in carbon reactivity to carbon coverages 94, 95, 96%.

Perturbation of carbon coverage at different reactivities

In figure 4.21 a perturbation of the carbon coverage level has been made at carbon coverage 91% for different reactivities. The net response is faster for higher carbon reactivities. For the highest reactivity 0.7 there is a small overshoot. For the lowest reactivity, 0.3, the step causes a slow response with a small inverse response. Since the step-response for reactivity 0.7 has an

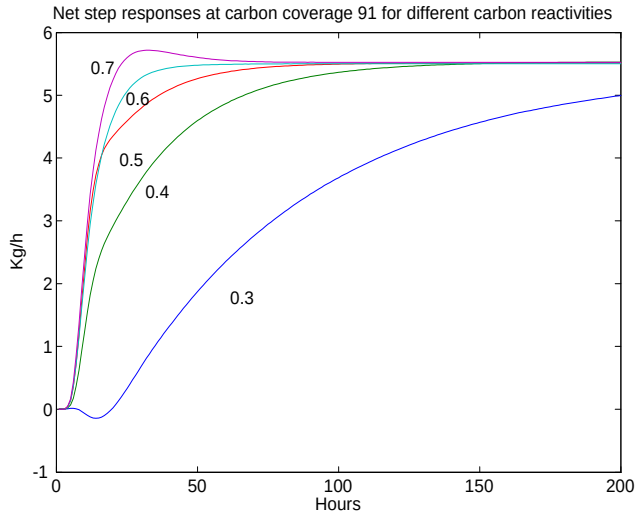

Figure 4.21: Net responses in silicon taprate to a 0.1% step in carbon coverage for different reactivities. Carbon coverage is 91%.

overshoot not observed in the nominal case, the response to this reactivity is examined further, cf. figure 4.22. Even at 96% carbon coverage there is no inverse response, although the response is slow.

Analysis of the responses to reactivity variations

The transient effects observed in figures 4.19 and 4.20 are typical for a system with competing phenomena with a fast positive response and then a slower negative response. The initial increase in the taprate seems to have the same time constant for all carbon coverage levels, whereas the negative response seems to have an increasing time constant for increasing carbon coverage levels. A larger time constant for the negative response lets the output reach a higher level before the decrease starts. Also, the slower, negative response seems to have a higher order for carbon coverage

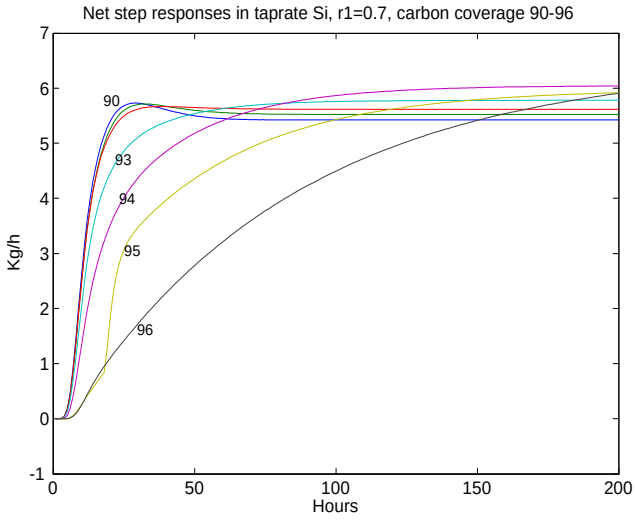

Figure 4.22: Net step response in silicon taprate to a 0.1% step in carbon coverage, $r_1 = 0.7$.

levels 94-96% than for 90-93%. For carbon coverage 90-93% there is no or very little unreacted carbon reaching the hearth, and there is some silicon production in the shaft, according to the model. A possible explanation of the responses in figure 4.20 is that an increase in the reactivity will increase the silicon production in the shaft further. Silicon metal produced in the shaft flows down into the hearth relatively quickly in the model. An increased consumption of SiC in the shaft will give less SiC in the hearth. This will subsequently decrease the Si production in the hearth and cause the negative response. The steady state effect is close to zero since the carbon coverage has not been changed.

Figure 4.21 shows that increased reactivity will give a faster response in the silicon taprate to a step in the carbon coverage. This figure along with figures 4.6 and 4.22 also show that the slow response occurs for lower carbon coverage rates if the reactivity is low. The reason for this is that low reactivity will give carbon in the hearth for low carbon coverage values. Therefore, a slow response to a positive carbon coverage step is an indication that the furnace is close to maximum silicon yield, or optimal carbon coverage, see figure 4.16. On the other hand, a fast response is an indication that the process is run with large margins to optimal carbon coverage. The speed of the response either in SiO fume or silicon taprate to carbon

coverage therefore expresses the margins to over-coking.

All the figures of this section indicate that an increase in reactivity should be associated with an increase in carbon coverage both to increase the silicon yield. Another reason is to avoid an excessively high temperature level at the furnace top which may damage equipment as well as give unpleasant working conditions for the operating personnel.

4.4 Steady state and dynamic response to the energy supply to the furnace

In the following, the steady state and dynamic responses to variations in the electrical power supply to the furnace are shown.

Figure 4.23 shows the steady state silicon production vs. carbon coverage for different load levels. The figure shows that the production increases almost linearly with the load level, except at high power and carbon coverage levels where the production is lowered.

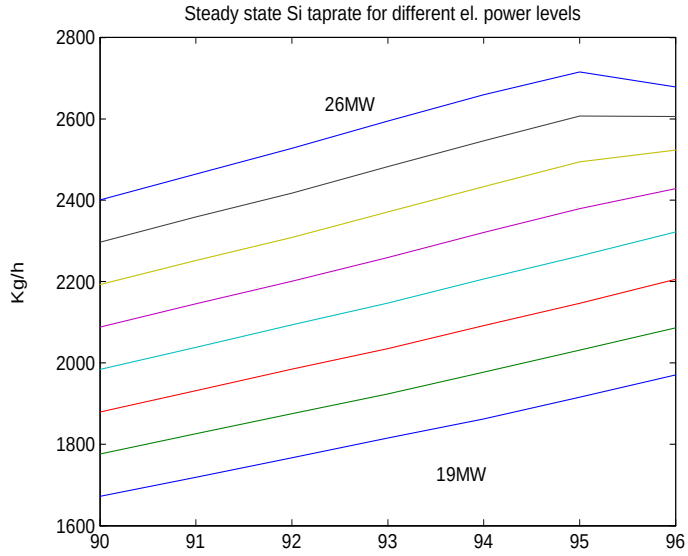

Figure 4.23: Steady state silicon production vs. net carbon coverage for electric power levels, $19MW, ..., 26MW$. Carbon reactivity, $r_1 = 0.5$.

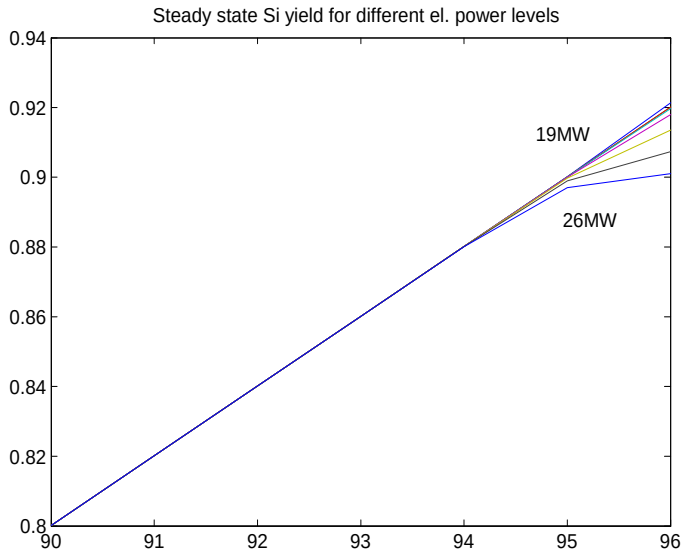

Figure 4.24: Silicon yield vs. carbon coverage for electric power levels $19MW, ..., 26MW$. Carbon reactivity, $r_1 = 0.5$.

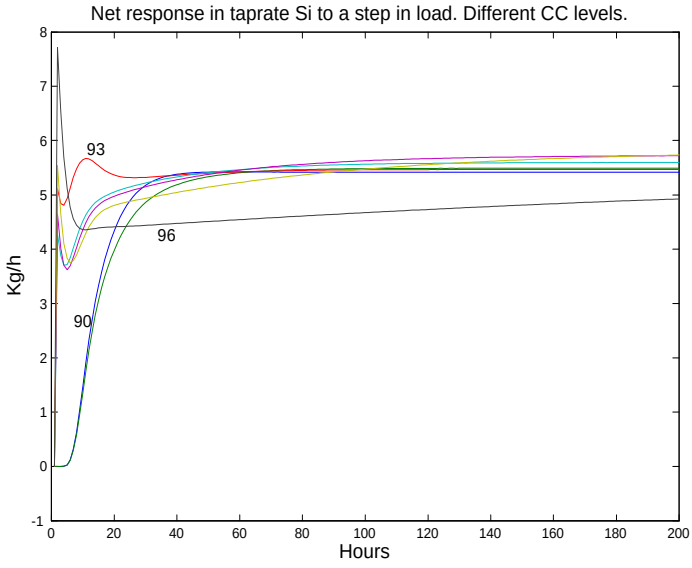

Figure 4.25: Net response in tapped silicon for step in load for different net carbon coverage values. Carbon reactivity $r_1 = 0.56$.

Figure 4.24 shows the corresponding silicon yield. This means that increasing the power level primarily increases the mass flow through the furnace, and has little effect on the silicon yield, except at high carbon coverage values where the achievable silicon yield is lowered when the power is increased.

Figure 4.25 shows the net step responses in the silicon taprate to a 1% step at time 1 hour for the nominal load, $22MW$. Each curve corresponds to a different carbon coverage level. At low carbon coverage 90 and 91%, the response seems close to a first or second order step response with no overshoot. For higher carbon coverage levels (92% and above), the response is more complex. From this we can conclude that the variations in load will cause considerable disturbances to the silicon production, and that the direction of the dynamic response varies depending on the carbon coverage level in the furnace.

4.5 Effects of variations in vertical energy distribution

A parameter determines how the total electric power is to be distributed between the hearth element and the shaft elements in Simod, cf. (3.11). The parameter expresses the fraction of energy supplied to the hearth, and has a nominal value of 0.9. Referring to the actual furnace, this fraction consists of both the arc energy and energy supplied by ohmic conduction to the hearth. The remaining fraction is distributed evenly between the shaft elements. This parameter was also discussed in section 3.2.2.

In the actual furnace, one generally wants as high a power fraction as possible to be supplied to the hearth, and as little as possible in the shaft. The energy distribution in the actual furnace is altered by changing the vertical placement of the electrodes. The composition of the charge also affects the energy distribution.

It is of interest to study the consequences of altering the energy distribution in the furnace through model studies. This can be done by changing the energy fraction to the hearth parameter. A narrow band, 0.89 to 0.91, has been chosen around the nominal value, 0.9. Other parameters are as specified in table 4.1.

4.5.1 Steady state response to changes in the electric energy distribution

Since there is very little difference in the steady state gain from carbon coverage to taprate for the different energy fraction to hearth levels, a mean silicon taprate value has been subtracted, and the net deviations from the mean taprate have been plotted in figure 4.26. There is a very small effect of changing the energy fraction to the hearth for the selected values, but the highest energy fraction lies above the mean for all carbon coverage values. The curves corresponding to 0.9 and 0.89 cross each other between carbon coverage 92 and 93%.

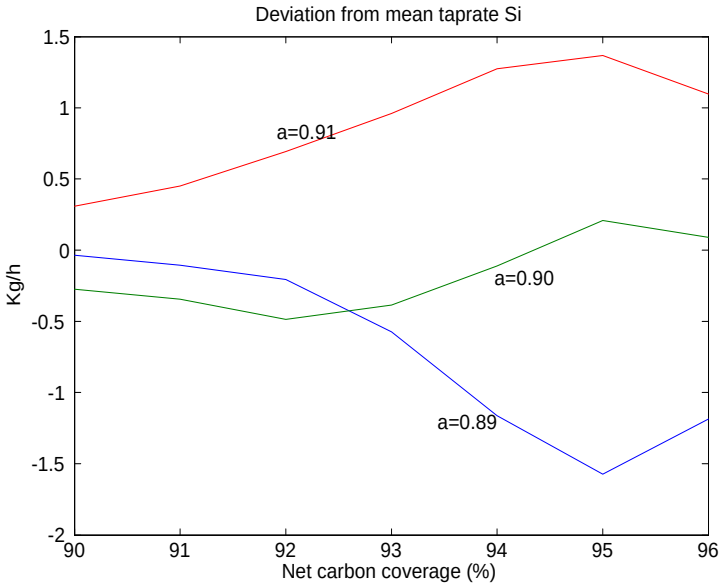

Figure 4.26: Deviation from mean in steady state si production for energy fractions to hearth at different carbon coverage levels.

4.5.2 Dynamic response to changes in the vertical energy distribution.

In the following, the dynamic effect of altering the electric energy fraction parameter in the simulator model is studied. Due to the nonlinear effects of the model, this has been done using a positive perturbation of 1% of the chosen interval, 0.89 to 0.91, i.e. 0.00018.

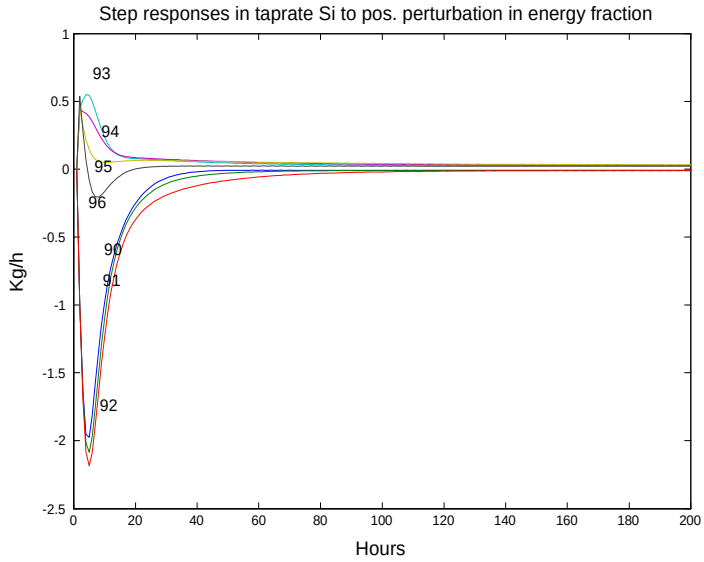

Figure 4.27: Net step response to a positive step in energy to hearth at $a = 0.90$

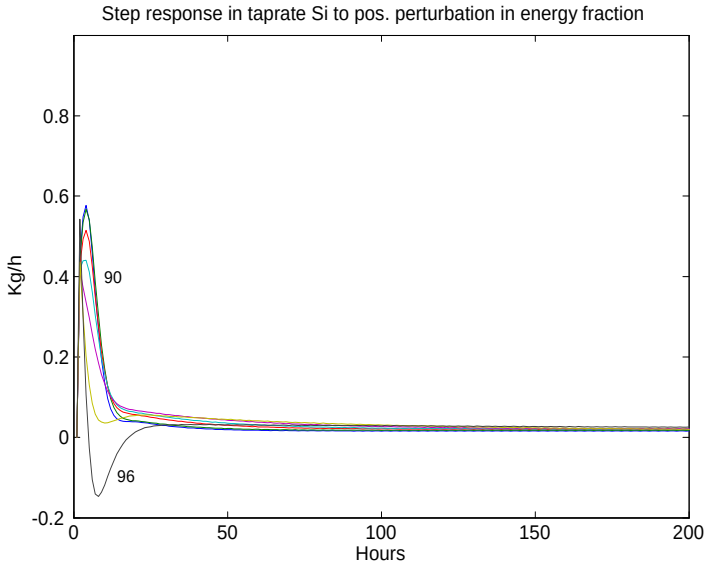

Figure 4.28: Net step response in the taprate of silicon to a positive perturbation in the energy fraction to the hearth at $a = 0.91$.

A positive step in the energy to the hearth implies a corresponding negative step to the energy supply in the shaft. Figure 4.27 shows that for carbon coverages 90 to 92%, an increase in the energy to the hearth will lead to a negative transient response in the silicon taprate. For higher carbon coverages, 93 to 96%, the small increase in the energy fraction to the hearth will lead to a positive transient response taprate of silicon and a decrease in the fume production.

In figure 4.28, the same perturbation has been made around a steady state energy fraction value of 0.91. In figure 4.28, all the transient responses are positive.A similar perturbation around $a = 0.89$ has also been made. In this case, carbon coverage values 90-94% gave a negative transient whereas 95 and 96% gave a positive transient response.

4.5.3 Analysis

The change in direction of the transient responses in figures 4.27 and 4.28 is further analyzed.

There is a qualitative shift in the taprate response in figure 4.27. For an energy fraction of 0.9, this shift takes place between 92 and 93%. The same shift in behavior can also be seen in SiC production in the hearth, i.e. decreased production for lower carbon coverage values. One fundamental difference between these two carbon coverage values for the nominal case is that for 92%, the combination of partial pressure of SiO and temperature gives silicon production in the lower elements of the shaft. This production has a negative response to a positive step in the energy fraction to the hearth. For carbon coverage 96%, there is no silicon production in the shaft.

Examination of the conditions in the model for $a = 0.91$ and $a = 0.89$ confirms this result. Figure 4.29 shows the steady state silicon production in the shaft for different carbon coverage values and energy fraction values. From the curves is can be observed that for a high energy fraction to hearth, 0.91, there is no silicon production in the shaft for any of the carbon coverage values. For the lower energy fractions there is silicon production in the shaft for the low carbon coverage values.

From an energy distribution perspective it is therefore useful to divide the under-coked region into two subregions depending on the energy distribution in the furnace. In one subregion an increase in the energy fraction to the hearth will lead to a rapid and significant increase in the fume production and a corresponding decrease in the silicon production. This region is characterized by a relatively low energy fraction to the shaft and/or a low

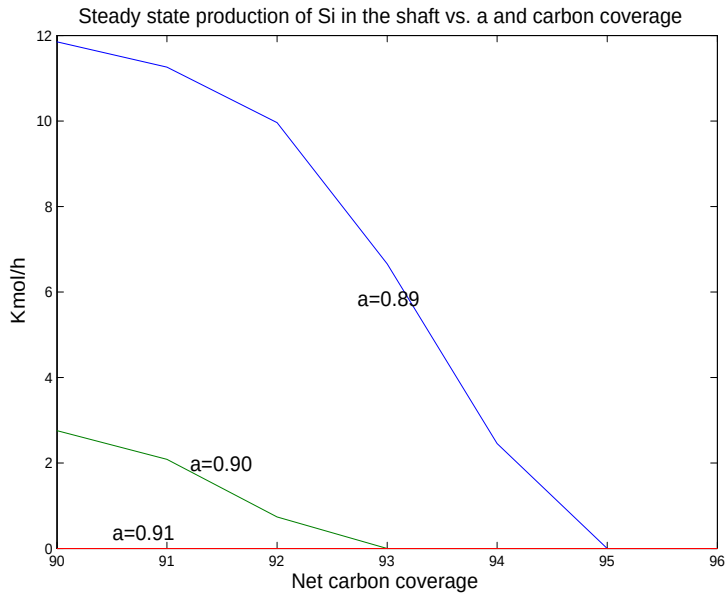

Figure 4.29: Steady state production of silicon in the shaft for different energy fractions to the hearth and carbon coverage values.

carbon coverage. This will give a warm shaft with a relatively high partial pressure of SiO and subsequent Si production in the shaft, and could therefore be called the "hot shaft" region.

In the other subregion of the under-coked operating region, the "cool shaft" region, the energy fraction to the hearth is higher and/or the carbon coverage is higher. This gives a cooler shaft with less SiO and no silicon production in the shaft. An increase in the energy fraction to the hearth will in this region lead to an increased taprate of silicon and a decrease in the SiO fume production. In figure 4.29, the steady state production of silicon in the shaft has been plotted for different energy levels and for different carbon coverage levels.

The effects have been summarized qualitatively in figure 4.30 showing the steady state gain between carbon coverage and silicon taprate. An increase in the energy to the hearth fraction, will move the boundary for the "hot shaft" region to the left, i.e. to lower carbon coverage values. A decrease in a will move the boundary towards higher carbon coverage values.

The effects described in both regions are highly transient, but has a very small steady state effect as can be seen in figure 4.31. The steps shown in

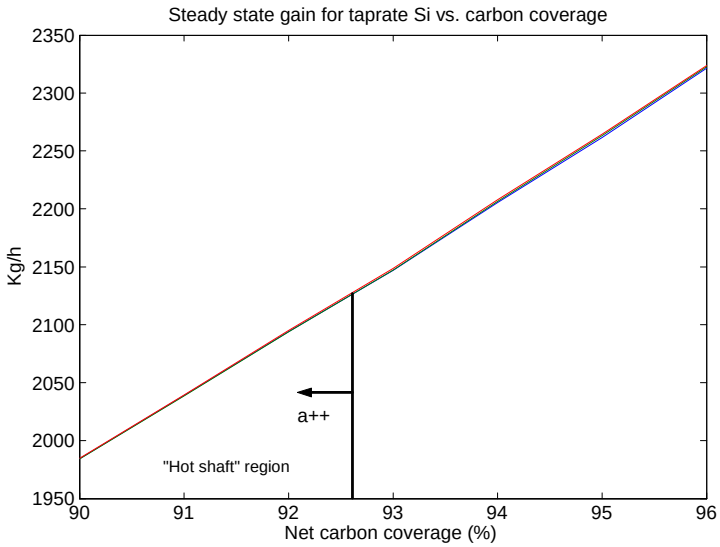

Figure 4.30: Increasing the energy to hearth fraction reduces the "hot shaft" region.

figure 4.31 are meant to illustrate the effect of a step of 0.01 from 0.89 to 0.9 in the ”hot shaft” region. There is a significant negative transient in the taprate of silicon and a very small steady state negative effect (-0.23kg/h in taprate).

A step from the ”hot shaft” region to the ”cool shaft” region, from 0.9 to 0.91 at carbon coverage 90, gives a transient which first goes negative, then positive and gives a very small positive steady state effect (+0.58 kg/h). A step within the ”cool shaft” region will give a positive transient and a small steady state effect of +1.5 kg/h. The transient is much smaller within this region.

Process operation implications

From an operation perspective it is well known that the energy should be supplied as far down in the furnace as possible. Too little energy to the hearth causes slag formation and problems such as clogging of the tap-hole.

The energy fraction has also been varied within a very narrow range to ensure sound model behavior. The results have still been included to provide documentation of the experiments made on the model, and the conclusions

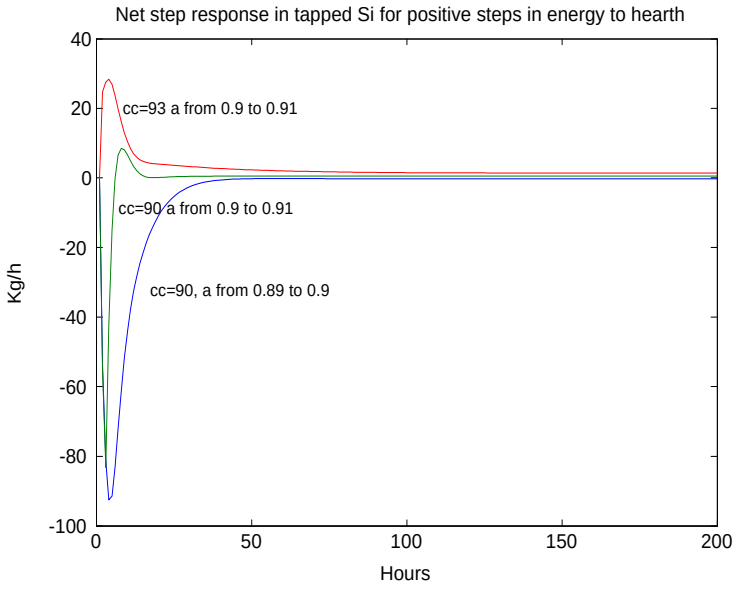

Figure 4.31: Net step response for 3 combinations of carbon coverage and energy fraction

drawn from the analysis. The effects observed and described here have not been verified by real operation, besides the fact that the model has been generally validated through use.

4.6 Interpretation for process operation and control

This section summarizes the results of the chapter in sections 4.6.1 and 4.6.2. In section 4.6.3 application of the results in operation is discussed.

4.6.1 Summary of steady state results

Figure 4.1 showed that increased carbon coverage increases the silicon taprate and silicon yield. Figures 4.15 and 4.16 showed that in the under-coked situation, an increase in the reactivity has little effect on the steady state taprate and silicon yield. Increasing the reactivity gives a potentially increased maximum taprate and maximum silicon yield. Reaching this potentially higher yield has to be done by increasing the carbon coverage. The exception is

when the silicon yield is at a maximum for one reactivity, in which case an increase in reactivity will give a significant increase in the steady state silicon yield as well. On the other hand, decreasing the reactivity may cause the furnace to enter an over-coked state with subsequent production problems when the furnace is operated at maximum silicon yield. A decrease in reactivity should therefore be associated with a decrease in carbon coverage, to be on the safe side.

Figure 4.23 showed that increasing the power increases the silicon taprate though increasing the material conversion rate through the furnace, but with very little effect on the silicon yield, see figure 4.24. However, at large carbon coverage values, the silicon yield may be decreased when increasing the electric power level.

Figure 4.32 gives a qualitative interpretation of the steady state relationship between the inputs carbon coverage, reactivity and the electric power level on the outputs silicon taprate, quartz consumption and silicon yield in an under-coked operating situation. The figure is slightly revised compared to the first version published in Lund et al. (2004*a*).

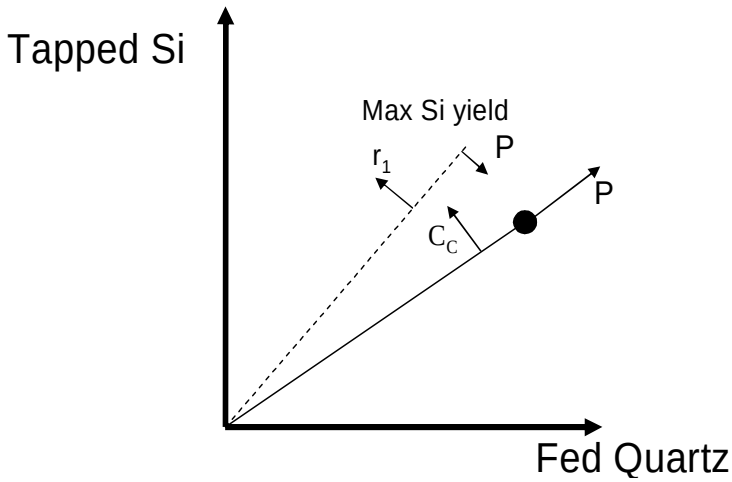

Figure 4.32: Qualitative steady state quartz vs. silicon relationship.

The axes are fed quartz and tapped silicon, respectively. The slope of a line within this "plane" will indicate the silicon yield. The circular point indicates an actual quartz feed rate vs. silicon production. The slope of the solid line out to this point gives the actual silicon yield. Increasing the carbon coverage will increase silicon yield, i.e. make the line steeper, as

indicated with the small C_C arrow attached to the solid line.

The left-most, dashed line indicates the theoretical maximum silicon yield. The arrow marked "r_1" indicates that an increase in the reactivity increases the theoretical maximum yield, which is known from figure 4.16.

Increasing the power will increase both quartz consumption and silicon production, but the yield will mainly be the same. Increasing the power will therefore mainly stretch the solid line, indicated with the attached arrow marked "P". Increasing the electric power to the furnace decreases the theoretical maximum yield somewhat at high carbon coverage values, cf. figure 4.24. This is indicated with the "P" arrow attached to the theoretical maximum yield.

4.6.2 Summary of the results of the dynamic responses

The responses in figures 4.22 and 4.6 were obtained with a carbon coverage positive step of 0.1%. In manual control of the furnace, larger steps are used, and a simulation using negative steps of 1% has therefore been included in figure 4.33 for demonstration. The steps start in a steady state situation corresponding to a net carbon coverage of 96%. The simulations are made with nominal parameters, cf. table 4.1. The inverse response at carbon coverage 96% in figure 4.6 can be recognized in figure 4.33 as the carbon coverage is stepped down from 96 to 95%. It can also be observed that the dominant time constant decreases as the carbon coverage decreases, i.e. it takes shorter time for a new steady state value to be reached when reducing from 91% to 90% than from 95% to 94%. The conclusion is that the same dynamic behavior is recognized in figure 4.33 as can be observed for the smaller steps in figure 4.6.

The most important result of the studies of the dynamic responses in this chapter is that the dominant time constant of the dynamic response in silicon taprate (or SiO fume) to a step in carbon coverage is an indication of the margins to critical carbon coverage (maximal silicon yield). A very slow response indicates that one is close to or at the maximum silicon yield, whereas a fast response indicates that one has a large margin to over-coking. The same, dynamic behavior can be found in the fuming dust responses. This is a very important result since one of the main problems in silicon furnace operation is to know what the margins are to over-coking. The recognition of this effect has been verified by leading metallurgists in Elkem.

The changes in the dynamic behavior can partly be observed in the fuming dust measurement, see figure 4.7. However, there is no change of sign in the fuming dust steady state gain, and no inverse response in this

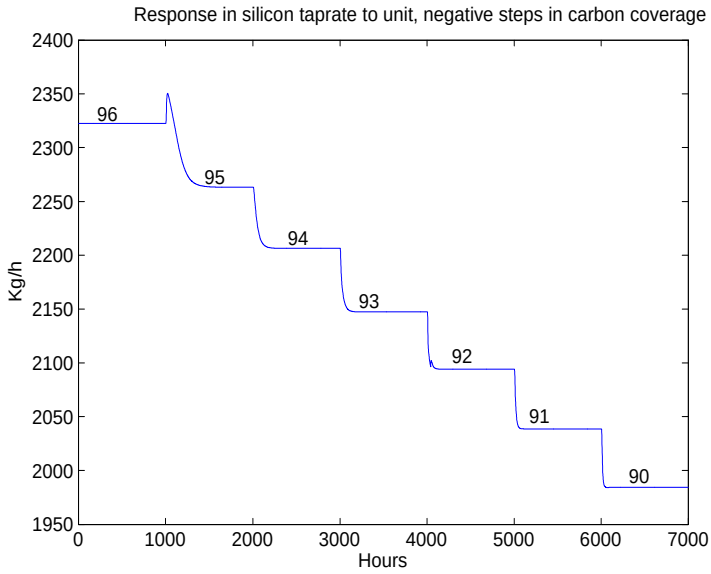

Figure 4.33: Step responses to 1% negative steps in net carbon coverage from 96%.

output at high carbon coverage values. The reason is that for high carbon coverage values the silicon is "stored" inside the furnace as silicon carbide. This effect takes place at the expense of the silicon taprate, not of the fuming dust production.

The inclusion of the reaction producing SiC_α in figure 4.9 drastically changes the response to increased carbon coverage close to maximum silicon yield. The reaction is modelled as a threshold mechanism, and comes into effect when the amount of carbon in the hearth exceeds a certain level. There are however uncertainties regarding the correctness in the representation of this mechanism. Also, uncertainties have been expressed regarding the reactivity of SiC_α. Therefore, the main analysis was made without this reaction.

4.6.3 Application of results in operation

The aim of silicon furnace operation is, regardless of the level of automation, to:

- optimize silicon yield

- avoid silicon carbide build-up in the hearth
- keep a high enough temperature in the hearth
- prevent an excessively high temperature at the furnace top

The implementation of the results of this chapter can be made at different levels of automation:

- Improved manual interpretation of furnace behavior
- Determine maximal carbon coverage through model-based estimation
- Automate carbon coverage determination

These points will be treated as separate sections in the following.

Improved manual interpretation of furnace behavior

The knowledge about the connection between margins to over-coking and furnace dynamics can be applied to manual operation by perturbing the carbon coverage level and observe the silicon furnace response. Even though this connection has been verified qualitatively by Elkem metallurgists, it is still a question how well these observations can be made in a real furnace.

It was shown in section 4.3.2 that a change in the reactivity mainly lead to a transient response in the taprate. The reactivity parameter used in the model is a lumped quantity affected by many properties of the carbon materials and the effective reactivity to the furnace may be difficult to determine. When the carbon material composition (reactivity) to the furnace is changed, the direction of the transient response in the taprate or fuming dust production can in principle used to determine if the effective reactivity of the new material composition is lower or higher than in the composition used previously. The carbon coverage must be held constant. The transient responses to a positive step in the reactivity were shown in figures 4.19 and 4.20.

Since there are many disturbances connected with taprate measurements, the transient response due to reactivity variations may be easier to observe through the fuming dust measurement. Figure 4.34 shows the responses in the fuming dust production corresponding to the taprate responses in figure 4.19.

The analysis in this chapter also showed that variations in the electric power and power distribution to the furnace cause fast variations in the

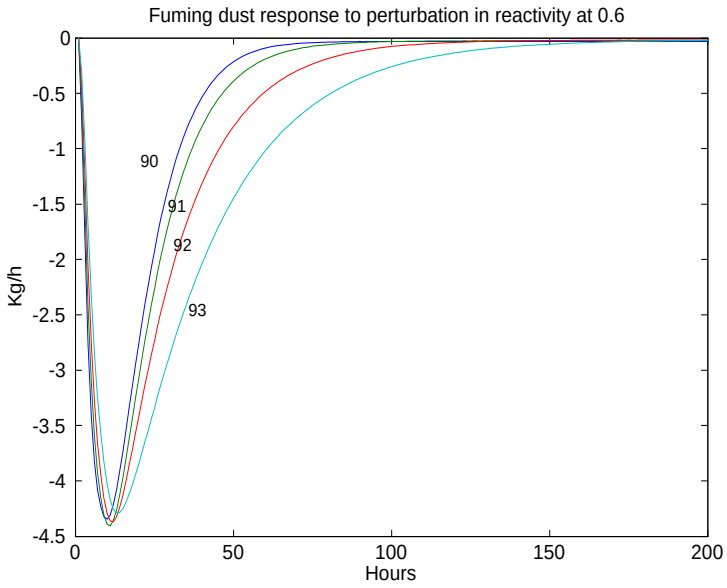

Figure 4.34: Net fuming dust response to a positive perturbation in the carbon reactivity.

silicon taprate. The response in the taprate to electric energy variations is also highly nonlinear and depends on the carbon coverage level. The analysis also showed that variations in the electric energy distribution in the furnace (energy fraction to hearth parameter) cause significant transients in the taprate of silicon. The direction of the transient varies depending on the carbon coverage level. These variations may or may not obscure the possibilities of making an unambiguous observation of the furnace's response to carbon coverage or carbon reactivity variations.

Another angle to this problem is to see if the electrical control applied to furnaces today is able to reduce or keep the variations in both the total power to the furnace and the power to the hearth sufficiently small.

To study this, a simplified equivalence diagram for the voltage between the electrode and the outer shell of the furnace (electric ground) has been drawn in figure 4.35. A similar diagram can be found in Schei et al. (1998). Current is conducted in the charge in the shaft and hearth and in the electric arc. For simplicity, the charge resistance R_c is shown as one resistance. The resistance in the arc, R_a can be viewed as being in parallel since the outer shell of the furnace is electrical ground. The voltage V is between the

electrode and the outer shell.

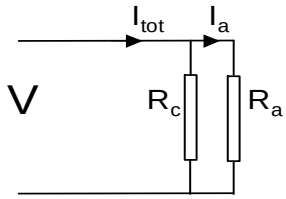

Figure 4.35: Equivalence diagram for the resistance between the electrode and ground.

The question is if both the total energy supply to the furnace, P, and the power to the arc, P_a, can be stabilized simultaneously. To study this, P and P_a should be expressed in terms of the main disturbances as well as the main free variables.

The charge resistance, R_c, varies depending on the composition and other properties of the charge and can therefore be considered a disturbance. The arc resistance, R_a, can be changed by raising or lowering the electrode. This can be done continuously, and R_a can therefore be considered a free variable. The voltage, V, can in principle be considered a free variable and used as an input, although it can only be changed in rather coarse steps over the furnace transformer. Expressing both the total power and the arc power in terms of the independent variables and the main disturbance gives

$$P = VI_{tot} = \frac{V^2}{R_t} = \frac{V^2}{\frac{R_c R_a}{R_c + R_a}}$$
$$P_a = VI_a = \frac{V^2}{R_a}$$

where R_t is the total resistance. If both R_a and V are free variables, it will in principle be possible to control P and P_a individually. However, since the voltage only can be changed in rather coarse steps in the furnace transformer, this input does not provide continuous control, and the only control input to be used in most situations is the resistance R_a. Therefore, it can be seen from the equations above that this single variable can not be used to keep both P and P_a constant.

The aim of electrical control is often to maintain a constant R_t, or at least keep it below a maximum value. In essence this implies control of the total power to the furnace, at the expense of variations in the arc power. The priority given to the total furnace power over the arc power is sensible since variations in the total power has a steady state effect on the taprate,

and also affects the theoretical maximum for the silicon yield, see figure 4.23.

The charge resistance, R_c, can also be altered by adding quartz close to the electrode. The quartz will melt and flow down along the electrode walls, and the resistance seen between the electrode and the furnace wall will increase significantly. This will cause the electrode to be lowered by the electrode control system in order to maintain a constant R_t. This is however not part of the regular electrical control of the furnace. This was also mentioned in section 2.4.1.

The analysis this far has not taken into account that the power supply to a silicon furnace is a 3-phase system, and the analysis has therefore been made on an overly idealized model. Still, even under these assumptions it has been shown that due to the coarse steps for voltage control, the total power to the furnace and the arc power can not simultaneously be kept constant. The electrodes are connected in a Δ-configuration (Valderhaug 1992) which means that disturbances in the voltage of the individual electrodes will affect the voltage of the other electrodes. In addition to this, currents flow in the charge between the electrodes. Currents will also vary due to inductive coupling between the electrodes.

The conclusion is therefore that even under the idealized conditions described in figure 4.35, it seems that variations in the power to the furnace or the arc and subsequent transient behavior must be tolerated due to the limitations in the voltage control.

Determine maximal carbon coverage through model-based estimation

A model based estimator can be used as an alternative or supplement to manual interpretation of the information from the process. The model should include both a material and energy balance. Even if it proves possible to make good manual observations of the furnace dynamics, a model based estimator can be a valuable addition to determine if the furnace is operated far from or close to its optimum. A model based estimation algorithm based on Simod is described in chapter 6 and the estimation results in chapter 7.

Automatic carbon coverage determination

Attempts on automatic control of carbon coverage has been made. Ruszkowski & Ydstie (2002) describes an inventory based, passive controller for the

carbon coverage rate. A passive controller of the carbon content can be obtained if on-line information about the total carbon content (carbon, silicon carbide, carbon monoxide) is provided. The method has been tested on a dynamic model for the silicon furnace. The model assumes a constant temperature profile through the furnace.

Chapter 5

Sensitivity analysis

The purpose of an estimation algorithm is to drive the model's outputs in the direction of the actual outputs. It is therefore crucial to find a set of model variables which can do this effectively.

This chapter analyses the sensitivities in the model outputs to the candidate parameters listed on page 37 to provide a basis for determining which parameters should be selected for estimation.

Referring to the list of the steps in the design of an on-line model on page 9, the sensitivity analysis is the fifth step.

The contents of the chapter are as follows. Section 5.1 gives a general introduction to parameter sensitivity analysis and defines the background and scope for the methods chosen here. The applied methods are presented in sections 5.2, 5.3, and 5.4.

The methods have been applied to the Simod sensitivity derivative, and the results are presented in sections 5.5 to 5.9. A summary and discussion of the results are given in section 5.10.

5.1 Introduction

Analysis of the sensitivity derivative of a model is a necessary step when designing model based parameter estimation algorithms. When existing simulation models are to be used for estimation, there may exist a high number of candidate parameters, while only a few may actually be included for estimation. It is therefore important that the sensitivity analysis provides the information necessary to rank the parameters internally and to determine how many of them should be included for estimation.

This work addresses methods for efficient manual and automated sensitivity analysis for nonlinear models on the general form

$$\hat{y} = g(\theta, u)$$

where $\hat{y} \in \mathbb{R}^{n_y}$ is the model output vector, $\theta \in \mathbb{R}^{n_\theta}$ the parameter vector, $n_y \geq n_\theta$ and $u \in \mathbb{R}^{n_u}$ is a measured input vector which may be present. The sensitivity derivative, S, of the model outputs to the parameter vector is defined as

$$S = W^{-\frac{1}{2}} \frac{d\hat{y}(\theta, u)}{d\theta} \in \mathbb{R}^{n_y \times n_\theta} \tag{5.1}$$

where $W^{-\frac{1}{2}} \in \mathbb{R}^{n_y \times n_y}$ is a positive definite, diagonal weighing matrix. The focus in this chapter is on situations when the sensitivity derivative is on a numerical form, as opposed to situations when an analytical expressions for S can be found. It is paramount that S has been scaled according the natural or permissible variation range of the outputs in y and the parameters in θ since the scaling directly affects the parameter ranking outcome.

Since the model is nonlinear, global identifiability can generally not be proven. To increase the probability that the model is identifiable over the whole parameter space, the sensitivity derivative must be checked for as many parameter vector values as possible. If the model is also nonlinear in the input vector u, then u will affect the sensitivity derivative, and S needs to be calculated for a wide range of values of u as well. This requires efficient automated sensitivity analysis methods.

Each column of the sensitivity derivative will express all of the outputs' sensitivity to a single parameter and can therefore be viewed as the sensitivity direction for the corresponding parameter. In sensitivity analysis, the following properties of the sensitivity derivative are of interest:

- The size of the elements in each column.
- The differences in the norm of the columns.
- Degree of linear dependence between columns.

Large elements in a column indicate that these outputs are very sensitive to changes in the parameter corresponding to the column. With many large elements, the column will represent a "long" vector, with a large norm. If there are large differences in the norms of the columns, this is an indication that the model outputs are much more sensitive to some parameters than others. Linear dependence between the columns of the sensitivity derivative

indicates that parameters or combinations give similar "signatures" on the output vector, and it may be difficult to identify these parameters.

Many linear transformations of S or $S'S$ can reveal these properties. Eigenvector transformation of $S'S$, or singular value transform of S, can be used for manual parameter ranking through ordering the eigenvalues of $S'S$ according to size, and inspecting the corresponding eigenvectors to determine which parameters are the most significant contributors to this particular direction. Li, Henson & Kurtz (2004) determines parameter ranking by principal component analysis to the steady state sensitivity derivative and computes the minimum distance between each sensitivity vector and the eigenvectors (principal direction).

The parameters can also be ranked based on calculating the condition number of subsets of S. A large condition number,

$$\kappa = \sqrt{\frac{\max eig(S'S)}{\min eig(S'S)}}$$

indicates either weak individual sensitivity or linear dependence within S. As an automated method this will give a trial and error method requiring the design of a selection algorithm. A selection algorithm may be designed to pick the parameters sequentially, according to the smallest condition number, or to check all possible subsets of parameters. Ranking by condition number has been used by Weijers & Vanrolleghem (1997).

Since many, maybe most, of the candidate parameters typically will be excluded from the final set, there is actually no need to combine the original directions into eigendirections before removing them. In this chapter parameter ranking is based on successive orthogonalization of S. The method utilizes Gram-Schmidt decomposition, and the columns are picked one by one selecting the vector with the largest Euclidean norm from the set of non-selected columns. The selected direction is removed from the remaining non-selected directions by projection and subtraction before a new column is selected.

The factorization method appears in the literature as QR factorization with column pivotation (QRcp) (Golub & Van Loan 1996).

The column permutations can be made based on different criteria. Kanjilal, Ballav & Saha (1995) applies QR decomposition to determine the most dominant direction in a linear regression matrix. The columns are selected so that the inner product between the column and the output vector is maximized. The selected direction is removed by projection both from the non selected columns of the regression matrix and also from the output vector

before a new choice is made. This method has in Kanjilal, Saha & Koickal (1999) been given the name "m-QRcp" and is applied to select variables in a multilayer neural network.

The advantage of successive orthogonalization is that no selection algorithm needs to be designed, since the selection order is an integral part of the method. This makes implementation particularly simple. This method is described by example in section 5.3 and can be used without assumptions about the probability distribution of the model error.

Even if automated methods are the focus here, it is still useful to inspect the sensitivity derivative manually to gain insight into important features of the problem, or to seek explanation for the automatic ranking results. A particularly useful form for manual inspection is presented in section 5.2.

For the present model, the model error $\varepsilon = y - \hat{y}(\theta)$ can be assumed to be Gaussian. Then the trace, $tr(S'S)^{-1}$, can be used as a measure of the variance of the parameter set. Since orthogonalization of S gives a triangular form of $S'S$, the variance contribution associated with each additional parameter included in the set has a particularly simple form, see section 5.4. This has been used to determine how many of the candidate parameters are to be included by deciding if there are "good" and "less good" subsets within the ranked candidate set, identified as leaps in the additional variance.

The parameters are in this case time varying. They can however be considered constant within a reasonable time frame, and it is therefore possible to use a time window of outputs in the estimation. This will improve the identifiability properties of the problem. Based on the variance properties of the problem, a minimum window length can be found, cf. section 5.7.7.

The additional variance contribution of individual parameters also provides a method for ranking the parameters. Ranking by successive orthogonalization has been compared to ranking based on smallest variance contributions and to ranking by condition number of the subset combinations $S'S$, cf. section 5.7.4.

5.2 Manual inspection of $S'S$

The following is valid for any sensitivity derivative $S \in \mathbb{R}^{n_y \times n_\theta}$, $n_y \geq n_\theta$, but for pedagogical reasons, an $n_\theta = 3$ matrix, $S = \begin{bmatrix} a & b & c \end{bmatrix}$, is used. $a, b, c \in \mathbb{R}^{n_y}$ are column vectors.

Cross multiplication of S gives

$$S'S = \begin{bmatrix} \|a\|^2 & \langle a,b\rangle & \langle a,c\rangle \\ \langle b,a\rangle & \|b\|^2 & \langle b,c\rangle \\ \langle c,a\rangle & \langle c,b\rangle & \|c\|^2 \end{bmatrix}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product and $\|\cdot\|$ the Euclidian norm of a vector. The squared norms on the diagonal of $S'S$ give a measure of the total sensitivity of each individual parameter. A large diagonal element indicates that variations in the corresponding parameter have a large "gain" on the model outputs. Large inner products in the off diagonal elements indicate that the two vectors have large elements in the same places (linear dependence), and/or a large norm of the vectors involved. To separate the information about the length of each individual parameter from the information about linear dependence, the diagonal elements of $S'S$ are extracted and $S'S$ is written on the form

$$S'S = D'C'D$$

with $D'CD$ equal to

$$\begin{bmatrix} \|a\| & 0 & 0 \\ 0 & \|b\| & 0 \\ 0 & 0 & \|c\| \end{bmatrix} \begin{bmatrix} 1 & \frac{\langle a,b\rangle}{\|a\|\|b\|} & \frac{\langle a,c\rangle}{\|a\|\|c\|} \\ \frac{\langle b,a\rangle}{\|b\|\|a\|} & 1 & \frac{\langle b,c\rangle}{\|b\|\|c\|} \\ \frac{\langle c,a\rangle}{\|c\|\|a\|} & \frac{\langle c,b\rangle}{\|c\|\|b\|} & 1 \end{bmatrix} \begin{bmatrix} \|a\| & 0 & 0 \\ 0 & \|b\| & 0 \\ 0 & 0 & \|c\| \end{bmatrix} \quad (5.2)$$

In (5.2) the first and last matrices contain information about the "strength" of each parameter, as sensitivity vector length. The off diagonal terms of the middle matrix can be recognized from the Cauchy Schwarz inequality, provided $\|x\|, \|y\|, > 0$, then

$$-1 \leq \frac{\langle x,y\rangle}{\|x\|\,\|y\|} \leq 1 \quad (5.3)$$

(5.3) can be recognized as the cosine of the angle between the vectors. Strong linear dependence gives values close to -1 or 1, whereas near orthogonality gives values close to zero.

The condition number of the middle matrix, C, in (5.2) is known as "collinearity index", (Brun, Reichert & Kunsch 2000), (Belsley 1991). Brun et al. (2000) uses the collinearity index for reduction of the parameter set by removing columns and combinations of columns from S and calculating the corresponding collinearity index. This is combined with inspection of different norms of the columns of the sensitivity derivative in order to decide

on which parameter set to use. A similar analysis was also made in a very initial stage of this the identification problem design, (Lund et al. 2004*a*). The method is however largely manual, and not very well suited for analysis of a large number of sensitivity derivative matrices.

5.3 Ranking by successive orthogonalization of S

In the following, parameter ranking is made utilizing the directions already present in the sensitivity derivative S. S is as defined in section 5.2.

Successive orthogonalization of S involves ranking the columns of S according to their "strength" and linear independence, simultaneously, and is carried out as follows: From the non-selected set of columns in S, select the column with the highest norm, form a unit vector, and remove this direction (by projection and subtraction) from the non selected set of columns. The procedure is repeated until all columns have been selected. The order of selection is stored in a permutation matrix. The selection procedure is demonstrated using the $S = \begin{bmatrix} a & b & c \end{bmatrix}$ matrix as an example.

Example 5.1 *Assume column vector a in S has the largest norm and therefore corresponds to the parameter with the highest individual sensitivity. This direction is selected, a unit vector,* $q_1 = \frac{a}{\|a\|}$*, is formed and removed from b and c by subtracting the projection of b and c onto* q_1*, giving*

$$\widetilde{b} = b - \frac{q_1' b}{q_1' q_1} q_1 = b - (q_1' b) q_1$$
$$\widetilde{c} = c - (q_1' c) q_1$$

$\widetilde{b}$ *and* $\widetilde{c}$ *which are now orthogonal to* q_1*. Assuming b originally was pointing in nearly the same direction as a, then a large component would have been subtracted from b when forming* $\widetilde{b}$*. Assume therefore that* $\widetilde{c}$ *has the larger norm, and is selected to form the second unit vector,* $q_2 = \frac{\widetilde{c}}{\|\widetilde{c}\|}$*. The* q_2 *direction is removed from vectors of the remaining set, which is now only* $\widetilde{b}$*, to form the new vector* $\bar{b}$

$$\bar{b} = \widetilde{b} - (q_2' \widetilde{b}) q_2$$

and set $q_3 = \frac{\bar{b}}{\|\bar{b}\|}$*. The selection order and decomposition can be summarized into* $SE = QR$*, where*

$$S \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} & & \\ q_1 & q_2 & q_3 \\ & & \end{bmatrix} \begin{bmatrix} q_1' a & q_1' c & q_1' b \\ 0 & q_2' c & q_2' b \\ 0 & 0 & q_3' b \end{bmatrix} \tag{5.4}$$

The procedure above is an "economical" QR decomposition since Q has n_θ columns. A full QR decomposition would produce a Q matrix with n_y columns.

The result (5.4) can be further refined by extracting the diagonal of R, $R = D\bar{R}$ which gives the QDR form of the cross product $E'S'SE$

$$E'S'SE = \overline{R}'D'Q'QD\overline{R} = \overline{R}'D'D\overline{R}$$

since $Q'Q = I$, the expression now has the triangular form

$$E'S'SE = \begin{bmatrix} 1 & 0 & 0 \\ \frac{q_1'c}{q_1'a} & 1 & 0 \\ \frac{q_1'b}{q_1'a} & \frac{q_2'b}{q_2'c} & 1 \end{bmatrix} D'D \begin{bmatrix} 1 & \frac{q_1'c}{q_1'a} & \frac{q_1'b}{q_1'a} \\ 0 & 1 & \frac{q_2'b}{q_2'c} \\ 0 & 0 & 1 \end{bmatrix} \tag{5.5}$$

with

$$D'D = D^2 = diag \begin{bmatrix} (q_1'a)^2 & (q_2'c)^2 & (q_3'b)^2 \end{bmatrix}$$

The elements of D^2 contain the squared lengths of the orthogonal fractions of the original sensitivity vectors.

The triangular form of $S'S$ in (5.5) is a Cholesky form. This connection between the results of Gram-Schmidt orthogonalization of a matrix S and the Cholesky form of $S'S$ is pointed out in Strang (1988), page 196.

The parameter selection procedure described in the example above is easy to implement in Matlab or from scratch. The QR algorithm in Matlab can be used, but needs to be run in "economy" mode, and with a permutation matrix as one of the return variables. The Matlab QR algorithm may return negative values on the diagonal of R, this is corrected by multiplication with -1 in R and Q. The diagonal of the returned R needs to be extracted.

Simple implementation of the algorithm can be made using an already existing QR decomposition algorithm, for instance in Matlab, as shown in algorithm 5.1.

Algorithm 5.1 *Matlab code for QR-based parameter ranking*

```
function [E,Q,D,R,covest]=qdr(S)
% function [E,Q,D,R,covest]=qdr(S)
% This function performs an "economy" QR decomposition
%(SE=QDR) according to the at each step
% most dominant column vector (highest norm).
%Q is a orthonormal matrix. The reordering of the columns result
% in a pivotation matrix E. We want D(i,i) > 0.
```

```
% The 0 in qr(S,0) is very important, otherwise Q will automatically be
% augmented to a quadratic matrix. We want Q to have the same dimen-
sions as
% S.
[Q,rr,ee]=qr(S,0);
[nyy,npp]=size(S);
E=zeros(npp);
invD=zeros(npp);
D=zeros(npp);
for (i=1:npp)
E(ee(i),i)=1;
if(rr(i,i)<0)
rr(:,i)=rr(:,i)*(-1);
Q(:,i)=Q(:,i)*(-1); %
end;
D(i,i)=rr(i,i);
invD(i,i)=1.0/D(i,i);
end;
R=invD*rr;
```

5.4 Variance contribution of additional parameters

Berntsen (1977) pointed out that a triangular form of $S'S$ gives a particularly simple expression for the variance contribution of individual parameters in the set. This is explained in the following.

The Cramer Rao inequality, (Ljung 1999), states that there is a lower limit to the covariance of any unbiased estimator of θ

$$cov(\hat{\theta}) \geq M^{-1} \tag{5.6}$$

where M is the Fisher information matrix. If it can be assume that the model error, $\varepsilon = y - \hat{y}(\hat{\theta})$, with $\varepsilon, y \in \mathbb{R}^{n_y}$, is stochastic with Gaussian properties and covariance W, then the Fisher information matrix is

$$M = \frac{d\hat{y}(\hat{\theta})}{d\theta}' W^{-1} \frac{d\hat{y}(\hat{\theta})}{d\theta} = S'S$$

The trace, or the sum of the diagonal elements can be regarded as a measure of the variance of the parameter set

$$var\hat{\theta} = tr(cov(\hat{\theta}))$$

Equation (5.6) can be used as a lower limit to the variance of the set

$$var\hat{\theta} \geq trM^{-1}$$

If the model error can be considered Gaussian, then

$$var\hat{\theta} \geq tr(S'S)^{-1}$$

Berntsen (1977) demonstrated that if $(S'S)^{-1} = UD^{-2}U'$ where U is unit upper diagonal, then the individual variance contribution of each parameter is

$$var\hat{\theta}_i = \frac{\|u_i\|^2}{d_i^2}$$

where u_i is column i in U. In the following, the results from Berntsen (1977) have been formulated as a proposition based on successive orthogonalization of S instead of LU factorization of $S'S$ used in the explanation in Berntsen (1977). The basic idea is that if one parameter is added to a set of $i-1$ parameters, the new sensitivity derivative is formed by appending the the $i'th$ parameter's sensitivity derivative

$$S_i = [S_{i-1}|s_i]$$

Performing successive orthogonalization on S_i in the column order 1 to i gives

$$S_i = Q_i \begin{bmatrix} D_{i-1} & 0 \\ 0 & d_i \end{bmatrix} \begin{bmatrix} \bar{R}_{i-1} & \bar{r}_i \\ 0 & 1 \end{bmatrix}$$

D_{i-1} and $\bar{R}_{i-1}$ are given by $S_{i-1} = Q_{i-1}D_{i-1}\bar{R}_{i-1}$. The appended $i'th$ column will only appear in the $i'th$ column of $\bar{R}_i^{-1}$, as is shown in the proposition below. This property of the inverse is also used in the Levinson-Durbin algorithm (Söderström & Stoica 1989) for updating of the parameter estimate.

Proposition 5.1 *Define the matrix* $S_i = [s_1|...|s_i], S_i \in \mathbb{R}^{(N*n_y)\times i}, N*n_y \geq i$ *and* $rank(S_i) = i$. $(S_i'S_i)$ *has the Cholesky factorization*

$$S_i'S_i = \bar{R}_i'D_i^2\bar{R}_i$$

where $D_i = diag[d_1, ..., d_i]$ *and* $\bar{R}_i$ *is upper unit triangular. Then the inverse*

$$(S_i'S_i)^{-1} = U_iD_i^{-2}U_i'$$

where $\bar{R}_i^{-1} = U_i = [u_1, ..., u_i]$ *and* $i \geq 1$ *has the property*

$$tr(S_i'S_i)^{-1} = tr(S_{i-1}'S_{i-1})^{-1} + \frac{\left\|u_i^2\right\|}{d_i^2}$$

Proof. $i = 1 : S_1 = s_1$ *gives* $R_1 = 1$, $U_1 = 1$ *then*

$$tr(S_1'S_1)^{-1} = \frac{1}{d_1^2}$$

$i = n : S_n = [S_{n-1}|s_n]$. *Assuming there exists an orthogonal form* $S_{n-1} = Q_{n-1}D_{n-1}\bar{R}_{n-1}$, *then*

$$S_n = Q_n \begin{bmatrix} D_{n-1} & 0 \\ 0 & d_n \end{bmatrix} \begin{bmatrix} \bar{R}_{n-1} & \bar{r}_n \\ 0 & 1 \end{bmatrix}$$

and $U_{n-1} = \bar{R}_{n-1}^{-1}$ *then* U_n *given by* $U_n\bar{R}_n = I$ *equals*

$$U_n = \begin{bmatrix} U_{n-1} & -U_{n-1}\bar{r}_n \\ 0 & 1 \end{bmatrix}$$

where $\bar{r}_n$ *only affects the* $n'th$ *column of* U_n. *Setting* $u_x = -U_{n-1}\bar{r}_n$ *then*

$$\begin{aligned} tr(S_n'S_n)^{-1} &= tr(U_nD_n^{-2}U_n') \\ &= tr \begin{bmatrix} U_{n-1} & u_x \\ 0 & 1 \end{bmatrix} \begin{bmatrix} D_{n-1}^{-2} & 0 \\ 0 & d_n^{-2} \end{bmatrix} \begin{bmatrix} U_{n-1}' & 0 \\ u_x' & 1 \end{bmatrix} \\ &= tr \begin{bmatrix} U_{n-1}D_{n-1}^{-2}U_{n-1}' + u_xd_n^{-2}u_x' & u_xd_n^{-2} \\ d_n^{-2}u_x' & d_n^{-2} \end{bmatrix} \\ &= tr(U_{n-1}D_{n-1}^{-2}U_{n-1}') + u_xd_n^{-2}u_x' + d_n^{-2} \\ &= tr(S_{n-1}'S_{n-1})^{-1} + \frac{\|u_n\|^2}{d_n^2} \end{aligned}$$

where $u_n' = \begin{bmatrix} -U_{n-1}\bar{r}_n \\ 1 \end{bmatrix}$. ■

The calculation of the variance contribution of each individual parameter can be used to find the cost of including one more parameter to an already ordered set. Calculating the additional variance may reveal leaps in the parameter variance which may provide crucial information for determining which are the most important parameters to include for estimation.

Calculating the minimum additional variance for the parameters can also be used as a way of ranking the parameters. The column vectors of S can

then be ordered in such a way as to give minimum additional variance. This requires comparing the variance associated with different subsets of column vectors, similar to ranking by condition number. A sensible starting point is to pick the column vector with the highest norm first, and then select a second depending on the smallest additional variance contribution and so on. It is also possible to check all possible combinations of $1, 2, ..., n_\theta$ sensitivity columns and find the combination of columns within each group which gives the smallest total variance, but this will not guarantee that the column with the largest norm is included.

5.5 Sensitivity analysis of Simod

This section can be considered an introduction to the remaining sections of this chapter which all deal with the sensitivity analysis of the Simod model.

The sensitivity analysis of Simod is a basis for:

- determining parameter ranking with respect to identifiability properties
- determining if there exist variations in parameter ranking depending on the operating point
- assessing the effect of the window length on the parameter ranking
- providing a basis for selecting an estimation method and criterion formulation

The sensitivity derivative has been obtained using a central difference approximation where each element s_{ij} in the sensitivity derivative $S \in \mathbb{R}^{(N*n_y)\times n_\theta}$ is

$$s_{ij} = \frac{1}{w_i^{\frac{1}{2}}} \cdot \frac{d\hat{y}_i}{d\theta_j} \approx \frac{1}{w_i^{\frac{1}{2}}} \cdot \frac{\hat{y}_i(\theta_j + \Delta\theta_j) - \hat{y}_i(\theta_j - \Delta\theta_j)}{2\Delta\theta_j}$$

$\Delta\theta_j$ is the perturbation interval for θ_j, which is the parameter value for element j in the parameter vector θ. In this case, $\frac{d\hat{y}}{d\theta}$ is scaled according to the natural variation range of the parameters and the outputs. This means that the sensitivity derivative is dimensionless, and the size of each s_{ij} represents a comparable "gain" from each parameters to each output. The parameter range, perturbation interval and output scaling are given in section 5.5.1.

The sensitivity derivative has been obtained using real input data from a silicon furnace.

Due to the nonlinear nature of the process, the sensitivity derivative has been examined in several operating points, see section 5.5.3.

An brief manual analysis of the sensitivity derivative has been performed in section 5.6 to obtain some initial knowledge about the parameter set and also to make it easier to interpret the results of the automatic ranking procedures.

In section 5.7, the parameters have been ranked using the mass flow measurements as outputs. The parameters have been ranked using successive orthogonalization and according to minimum additional variance. For comparison, the parameters have also been ranked according to the condition number of the parameter subsets, and the three methods have been compared to show to which extent they give different results, see section 5.7.4.

Since many of the parameters are constant or slowly varying, the output vector can be expanded using a sliding window of past data. As a rule, the window should be long enough to ensure good identifiability, but still minimal to ensure reasonable calculation times. The variance of an estimated parameter set is known to decrease with $\frac{1}{N}$ as the number of samples $N \to \infty$, (Ljung 1999). If the variance decreases more than proportionally to $\frac{1}{N}$, this is an indication that the window length should be increased. The development of the variance in the parameter set vs. window length has therefore been examined, see section 5.7.7.

Since many furnaces do not have individual weighing of the fuming dust production, a sensitivity analysis omitting this output has been performed in section 5.8. The input data given in 5.5.2 were used.

Sensitivity analysis is always dependent on scaling. In section 5.9, a scaling giving more weight to the fuming dust measurement has been used, and the results of the corresponding ranking have been compared with the ranking in 5.7.

5.5.1 The parameter set, perturbation interval and output scaling

The parameters in table 5.1 have been analyzed.

A description of the parameters and the inputs to Simod is given in sections 3.2 and 3.3.

The range for each parameter has been carefully chosen and is meant to express a physically sound range as well as a range where the model behavior

Table 5.1: Parameter names and range

	Parameter name	Symbol	Range	1% pert.
1	Specific electrode resistance	r_e	$0, .., 0.4$	0.004
2	Reactivity of carbon materials	r_1	$0.3, .., 0.7$	0.004
3	Carbon coverage loss	ΔC_C	$5, .., 20$	0.15
4	Direct fume rate fraction	$u_{\mathrm{Si\,O},tap}$	$0, .., 0.1$	0.001
5	Mean el. energy fraction to hearth	a_{mean}	$0.88, .., 0.92$	0.0004
6	Charge level set-point, shaft top	h_{top}	$0, .., 0.4$	0.004
7	Metal level set-point in hearth	h_{metal}	$0.1, .., 0.16$	0.0006

is sound. The perturbation interval has been chosen to be 1% of the total range for each parameter. The scaling of the parameters is therefore implicit in the perturbation interval. The response to each parameter on the different outputs can therefore be compared directly.

The scaling of the outputs has been made considering the standard deviation in the simulated outputs using the actual input series, and also considering the noise in the corresponding actual outputs and has been chosen to be

$$W^{-\frac{1}{2}} = diag\left[\begin{array}{ccc} \frac{1}{43} & \frac{1}{20} & \frac{1}{25} \end{array}\right] \tag{5.7}$$

5.5.2 Input data in the sensitivity analysis

The input data to the model used in the sensitivity analysis are the measurements of electric power, carbon coverage and electrode carbon from an actual silicon furnace. The scaled data are shown in figures 5.1, 5.2, and 5.3.

The input data correspond to approximately two months of normal operation. This also includes short periods with some relatively large excitations in the input data, see figure 5.1. It is of interest to see how this affects the sensitivity in the data, and the time these excitations occur should therefore be noted. The carbon coverage input in figure 5.2 shows a period with a relatively high carbon coverage level between samples $80-120$. The consumed electrode carbon is shown in figure 5.3. This carbon comes in addition to the carbon fed at the top.

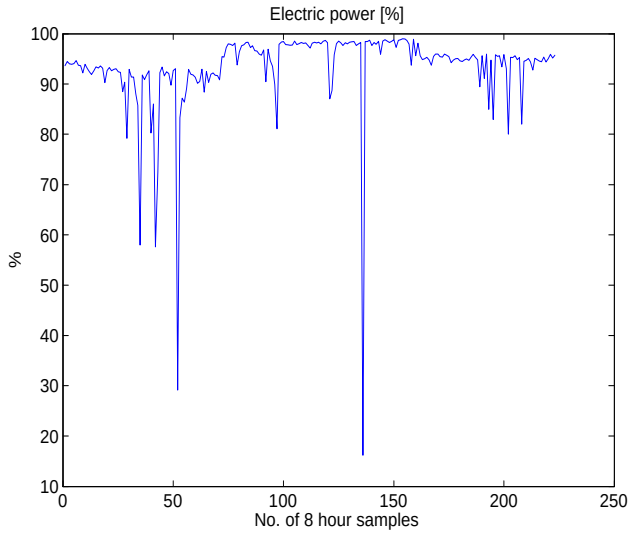

Figure 5.1: Electric power input used in sensitivity analysis.

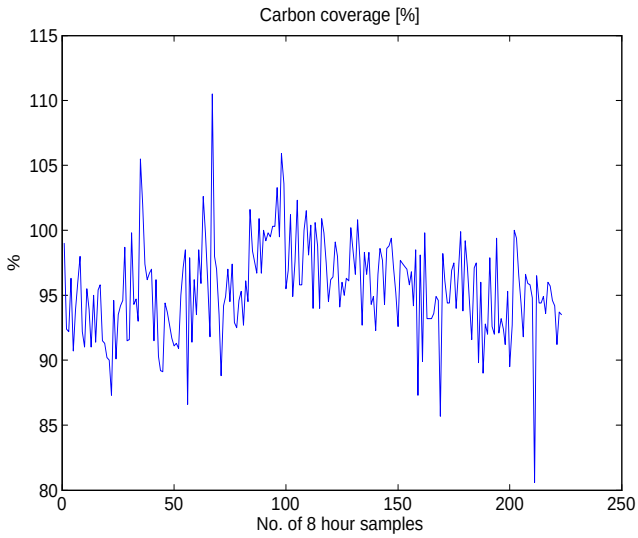

Figure 5.2: Carbon coverage input used in sensitivity analysis

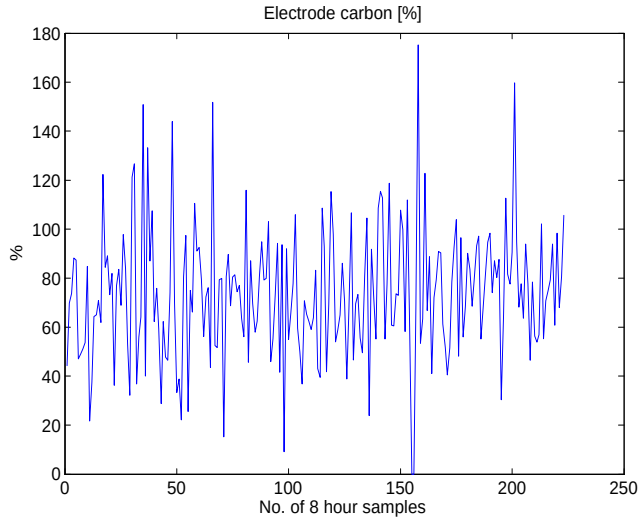

Figure 5.3: Electrode carbon used in sensitivity analysis.

5.5.3 Operating points for the sensitivity analysis

For a nonlinear model, global identifiability can not generally be proven. The best option is to analyze the sensitivity derivative at several and different operating points. Chapter 4 showed that some inputs caused a steady state response while other mainly caused transient responses. The analysis in chapter 4 also showed that the response to parameters also changed depending on the carbon coverage level. Since the loss of carbon coverage will represent an offset in the carbon coverage, it is of interest to check the sensitivity to the parameters for different values of this parameter. Since electrode carbon data from the process has been used as an input to the model in the sensitivity analysis and constitutes approximately 10% of the total carbon supply to the process, the carbon loss parameter should lie between 10 and 15%. The chosen values are therefore as given in table 5.2.

Some comments should be made regarding the choices of parameter vector values. θ_1 has low reactivity and a high electrode loss resistance. The low reactivity can give a high sensitivity to the reactivity parameter if the carbon coverage is high, according to the results in chapter 4. θ_2 contains

Table 5.2: Parameter vector values in sensitivity analysis

	Parameter	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
1	r_e	0.37	0.37	0.2	0.2	0.2	0.2
2	r_1	0.4	0.4	0.6	0.5	0.6	0.6
3	ΔC_C	13.7	13.7	13.7	15.0	10.0	13.7
4	$u_{\mathrm{Si}O,tap}$	0.05	0.001	0.03	0.05	0.03	0.09
5	a	0.9	0.9	0.92	0.92	0.88	0.9
6	h_{top}	0.2	0.2	0.1	0.1	0.1	0.1
7	h_{metal}	0.13	0.13	0.13	0.13	0.13	0.13

the same values as θ_1 except for direct fuming from hearth (parameter 4) which is nearly zero.

θ_3 represents a situation with high reactivity. θ_3 also contains a high average energy fraction to the hearth (element 5 in the parameter vector), a low electrode loss resistance (element 1) and a low charge level at the furnace top (element 6).

θ_4 has a high loss of carbon coverage, a high energy fraction to hearth and a low electrode loss resistance.

θ_5 has a low carbon coverage loss, high reactivity, low energy fraction to hearth and a low electrode loss resistance.

θ_6 has a high reactivity, low energy fraction to hearth, and a high direct fume loss.

One could argue that the choice of 6 different parameter vectors is a small number, even though the values have been chosen to represent quite different parts of the parameter space. However, since two of the parameters are losses in the inputs, the time varying input in effect gives a large variation range for the corresponding parameters since actual inputs are used. Also, the sensitivity to each parameter is calculated under quite different operational conditions along the time series. The sensitivity derivative calculations at a particular parameter vector takes a few hours calculation. This is another aspect, but is not limiting to the number of points, at least not at this level.

5.6 Manual inspection of the candidate set

In order to get an initial overview over key properties of the candidate parameter set, a manual inspection is made using the $n_y = 3$ outputs of a window of $N = 20$ time instants giving a $(N * n_y) \times n_\theta = 60 \times 6$ sen-

sitivity derivative. The sensitivity derivative has been obtained for $\theta_3 = \begin{bmatrix} 0.2 & 0.6 & 13.7 & 0.03 & 0.92 & 0.1 & 0.13 \end{bmatrix}$. A transformation of $S'S$ according to (5.2) gives

$$D = diag \begin{bmatrix} 8.0 & 2.52 & 32.4 & 1.38 & 1.55 & 3.49 & 1.33 \end{bmatrix} \tag{5.8}$$

and

$$C = \begin{bmatrix} 1.0 & 0.0996 & -0.172 & 0.0182 & 0.0764 & 0.848 & -0.021 \\ 0.0996 & 1.0 & 0.154 & 0.0984 & -0.135 & 0.321 & -0.544 \\ -0.172 & 0.154 & 1.0 & -0.023 & 0.118 & -0.295 & -0.195 \\ 0.0182 & 0.0984 & -0.023 & 1.0 & -0.616 & 0.189 & 0.219 \\ 0.0764 & -0.135 & 0.118 & -0.616 & 1.0 & 0.113 & -0.006 \\ 0.848 & 0.321 & -0.295 & 0.189 & 0.113 & 1.0 & 0.0342 \\ -0.021 & -0.544 & -0.195 & 0.219 & -0.0056 & 0.0342 & 1.0 \end{bmatrix}$$

By inspection of D in (5.8), the third parameter has the highest individual sensitivity. The low values in the off-diagonal elements of the third row and column of C indicate little linear dependence with the other parameters. The second choice can therefore be made independently of the first, and parameter 1 is a likely candidate since it has the second largest element in D. From this point on, manual selection of parameters is difficult. $c_{1,6}$ indicates that there is a relatively high linear dependence between parameter 1 and 6. Column 6 in S is longer than column 2, according to D, but with the linear dependence between columns 6 and 1, it is difficult to determine whether parameter 2 or parameter 6 should be the third choice. Also, element $c_{5,6}$ indicates that there is some linear dependence between the sensitivity derivative of these two parameters. From this the conclusion is that parameters 3 and 1 most likely are the first choices. As the third choice, parameter 2 or 6 are the most likely. Parameter 5 may be a choice before 4, but this is uncertain. Parameter 7 seems to end in the last place. This parameter has the lowest norm, and there is little linear dependence with other parameters.

The manual inspection has revealed important properties of the parameter set, and gives a basis to interpret the results of the automatic ranking. It is important to bear in mind that the analysis was made for one value of the parameter vector, based on a very short sequence of inputs in a process that is nonlinear. Manual ranking is therefore not sufficient, but useful as an additional tool.

5.7 Sensitivity analysis results, using 3 mass flow outputs

In this section, the parameters have been ranked according to both successive orthogonalization and minimum additional variance. All the three mass flow outputs, i.e. fed quartz, tapped silicon, and fuming dust production have been used, giving $n_y = 3$. All simulated outputs are accumulated over 8 hours.

The sensitivity derivative has been produced for the entire time series of 222 input samples, given in figures 5.1, 5.2, and 5.3. In the analysis, a sliding window subset of the total sensitivity derivative has been used, with time window lengths $N = 7$, $N = 20$, and $N = 50$ respectively. Since there are three outputs at each time instant, the sensitivity derivative of 20 time samples and 6 parameters will have a dimension $(N * n_y) \times n_\theta = (60 \times 6)$.

The sliding window approach on this set of data generates $(222 - N + 1)$ sensitivity derivative sub-matrices, and ranking is made for each of these. The analysis is made at the 6 different parameter value sets, see table 5.2. This generates a large number of ranking results, and mean value for the rank of each parameter along with a standard deviation has therefore been calculated.

5.7.1 Parameter ranking, $N = 7$

Ranking by successive orthogonalization

Table 5.3 shows the mean rank of each parameter at the six different parameter vector values in table 5.2, ranked by successive orthogonalization. The corresponding standard deviation of the ranking is shown in table 5.4.

Parameter 3, ΔC_C, is ranked first for all parameter vector values in table 5.3. The rank has zero standard deviation for parameter vector values θ_3 and θ_4, see table 5.4. This means that the parameter is ranked first in all of the sensitivity derivatives calculated in the windows along the whole time series. For $\theta_1$and θ_2, there is a higher standard deviation on the rank of parameter 3.

This means that other parameters have been ranked before this parameter in a few cases. Parameter 1, r_e, is ranked second with a close to zero standard deviation for θ_3 and θ_4. Parameter 2, the carbon reactivity, r_1, is ranked in third place. The average rank of r_1 is better when the reactivity is

Table 5.3: Average parameter ranking, orthogonalization, $n_y = 3, N = 7$

Par.no.	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
1	2.07	2.23	2.01	2.0	2.09	2.20
2	2.54	2.59	3.14	3.03	3.59	3.82
3	1.40	1.19	1.0	1.0	1.02	1.0
4	5.41	5.49	5.48	5.64	5.36	5.24
5	5.69	5.56	5.42	5.84	4.36	4.07
6	4.95	4.75	5.12	4.72	5.58	5.40
7	5.93	6.20	5.82	5.77	6.0	6.27

Table 5.4: Standard deviation in parameter ranking, orthogonalization, $n_y = 3, N = 7$

Par.no.	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
1	0.260	0.359	0.028	0.009	0.196	0.327
2	0.651	0.570	0.260	0.073	0.676	0.670
3	0.629	0.332	0	0	0.037	0.009
4	0.963	1.04	1.07	0.931	1.05	1.08
5	0.921	0.960	1.09	1.02	1.32	1.38
6	0.874	0.731	0.821	0.795	0.890	0.921
7	0.900	0.722	0.928	0.862	0.772	0.802

low, as in θ_1 and θ_2. This is consistent with the curves shown in figure 4.15 on page 60, which indicate that the sensitivity to the reactivity is higher around the carbon coverage optimum.

Examination of the ranking data directly shows that for some time windows, the reactivity is ranked before carbon coverage loss. This also explains the standard deviation of the carbon coverage loss found in θ_1 and θ_2.

Inspection of the mean values corresponding to parameters 4, 5, 6, and 7 give less conclusive results than for parameters 1, 2, and 3. Still, parameter 6, the height of the top shaft element may seem to be the parameter that could be ranked fourth. The parameter is ranked fourth 4 times. Parameters 4 and 5 are mainly ranked in fifth and sixth place, whereas parameter 7 seems to be the last choice. An observation is that parameter 5, the mean energy fraction to the hearth, ranks better for θ_5 and θ_6, for which the reactivity, parameter 2, has a less good rank. These values of θ gave a high reactivity and a low electrode resistance.

Ranking by minimum additional variance

Ranking by minimum additional variance and successive orthogonalization gave different ranking in 388 out of 1290 rankings (30%). To see if this causes a totally different overall ranking, the mean and standard deviation of the rankings for θ_1 to θ_6 has been examined, see tables 5.5 and 5.6.

Table 5.5 shows that parameters 3, 1, and 2 have the best rank. For $\theta_1 - \theta_4$, parameter 6 is ranked fourth, whereas for θ_5 and θ_6, parameter 5 is ranked fourth.

Table 5.5: Average parameter ranking, min. add. var., $n_y = 3, N = 7$

Par.no.	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
1	2.07	2.22	2.02	2.0	2.12	2.21
2	2.56	2.60	3.18	3.04	3.85	3.90
3	1.41	1.18	1.0	1.0	1.02	1.0
4	5.41	5.48	5.45	5.67	5.40	5.29
5	5.63	5.51	5.35	5.72	4.28	4.01
6	5.02	4.87	5.26	4.91	5.56	5.51
7	5.89	6.14	5.74	5.66	5.77	6.08

Table 5.6: Std. in parameter ranking, min. add. var., $n_y = 3, N = 7$

Par.no.	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
1	0.267	0.354	0.0365	0.009	0.241	0.338
2	0.665	0.562	0.326	0.082	0.807	0.718
3	0.636	0.316	0	0	0.037	0.009
4	0.991	1.11	1.07	0.929	1.12	1.08
5	0.930	0.956	1.15	1.10	1.35	1.35
6	0.851	0.732	0.830	0.826	0.975	0.965
7	0.937	0.716	0.996	0.906	0.923	0.840

5.7.2 Parameter ranking, $N = 20$

Inclusion of 20 time samples in the evaluation of each sensitivity derivative gives a $(N * n_y) \times n_\theta = (60 \times 6)$ sensitivity derivative and the following ranking of the parameters. The parameters have been ranked using orthogonalization and according to minimum additional variance. The mean

position of each parameter ranked according to orthogonalization is given in table 5.7 and the standard deviation of the position in table 5.8. The corresponding results for ranking according to minimum additional variance are given in tables 5.9 and 5.10.

Table 5.7: Average parameter ranking. Orthogonalization, $n_y = 3, N = 20$

Par.no.	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
1	2.16	2.38	2.0	2.0	2.01	2.06
2	2.60	2.57	3.16	3.0	3.47	3.74
3	1.24	1.05	1.0	1.0	1.0	1.0
4	4.89	5.28	5.35	5.43	5.51	5.65
5	6.06	5.55	5.11	6.06	3.83	3.45
6	4.80	4.58	5.04	4.61	5.66	5.47
7	6.25	6.59	6.34	5.90	6.51	6.63

Table 5.8: Average ranking std. Orthogonalization, $n_y = 3, N = 20$

Par.no.	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
1	0.300	0.472	0	0	0.029	0.120
2	0.572	0.525	0.268	0.010	0.517	0.456
3	0.420	0.095	0	0	0	0
4	0.719	0.748	0.951	0.935	0.822	0.719
5	0.769	0.897	1.10	0.829	0.753	0.742
6	0.813	0.711	0.765	0.697	0.664	0.763
7	0.725	0.537	0.742	0.814	0.621	0.493

The methods gave different ranking in 471 out of 1212 rankings (39%). The ranking order is again the same for $\theta_1, \theta_2, \theta_5$ and θ_6, while the two methods give different ranking orders for θ_3 and θ_4.

In tables 5.7 and 5.9, parameter 3 is always ranked first and parameter 1 second. For low reactivities, in θ_1 and θ_2, the average ranking of the reactivity, parameter 2, has a lower value than for high reactivity values. For θ_1 and θ_2, the average rank indicates that the reactivity in some cases has been ranked first. The ranking of parameter 5, average energy to hearth fraction, has a lower value for θ_5 and θ_6 where the reactivity is high and the electrode loss resistance is low. For θ_6, the average energy to hearth is ranked third, before the reactivity.

Table 5.9: Average parameter ranking, minimum variance, $n_y = 3, N = 20$

Par.no.	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
1	2.08	2.35	2.0	2.0	2.07	2.09
2	2.68	2.61	3.15	3.0	3.62	3.84
3	1.24	1.04	1.0	1.0	1.0	1.0
4	4.96	5.30	5.33	5.46	5.93	5.70
5	5.97	5.50	4.95	5.96	3.61	3.38
6	4.97	4.66	5.41	4.86	5.74	5.52
7	6.11	6.54	6.16	5.72	6.03	6.47

Table 5.10: Std. in parameter ranking, variance method $n_y = 3, N = 20$

Par.no.	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
1	0.182	0.453	0	0	0.129	0.162
2	0.510	0.505	0.260	0	0.525	0.398
3	0.420	0.076	0	0	0	0
4	0.814	0.804	0.910	0.986	0.867	0.837
5	0.810	0.916	1.08	0.891	0.803	0.737
6	0.889	0.775	0.842	0.768	0.750	0.747
7	0.803	0.550	0.942	0.942	0.688	0.603

5.7.3 Parameter ranking, $N = 50$

Inclusion of 50 time samples gives a $(N * n_y) \times n_\theta = (150 \times 7)$ sensitivity derivative. The average ranking results using successive orthogonalization and minimum additional variance are given in tables 5.11 and 5.12. The standard deviation is not shown.

The methods gave different ranking in 388 cases out of a total 1032 (37%). The average ranking order is largely the same. In θ_2 the reactivity ranks in second place in table 5.11. For θ_5 and θ_6, the average energy fraction to the hearth, parameter 5, ranks before the reactivity.

5.7.4 Comparison with ranking according to condition number

For comparison, the parameters were also ranked according to condition number, giving a third method. The selection according to condition num-

Table 5.11: Avg. ranking. Orthogonalization, $n_y = 3, N = 50$

Par.no.	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
1	2.16	2.60	2.0	2.0	2.0	2.0
2	2.84	2.40	3.0	3.0	3.56	3.86
3	1.0	1.0	1.0	1.0	1.0	1.0
4	4.67	5.24	5.07	5.24	5.40	5.59
5	6.37	5.43	4.87	6.34	3.52	3.14
6	4.45	4.39	5.12	4.22	5.76	5.52
7	6.51	6.94	6.94	6.20	6.76	6.88

Table 5.12: Average parameter ranking, minimum variance, $n_y = 3, N = 50$

Par.no.	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
1	2.06	2.42	2.0	2.0	2.0	2.0
2	2.94	2.58	3.0	3.0	3.72	3.87
3	1.0	1.0	1.0	1.0	1.0	1.0
4	4.79	5.25	5.16	5.17	5.95	5.45
5	6.31	5.37	4.66	6.30	3.30	3.13
6	4.51	4.52	5.46	4.40	5.70	5.82
7	6.38	6.86	6.72	6.14	6.33	6.73

ber was made successively. The overall conclusion using this method was also that parameters 3, 1, and 2 were the most important.

Table 5.13, 5.14, and 5.15 express to which extent three methods, successive orthogonalization (SO), minimum variance (MV), and minimum condition number (CN) have ranked different parameters in first place, second place, etc. using window lengths of 7, 20, and 50 samples respectively.

Table 5.13: Percentage of different ranking results for three different methods. $n_y = 3, N = 7$

	1.	**2.**	**3.**	**4.**	**5.**	**6.**	**7.**
SO vs. MV(%)	0	2.0	3.6	4.8	12.1	15.8	10.3
SO vs. CN(%)	0	3.8	6.9	7.4	15.4	19.2	12.2
MV vs. CN(%)	0	1.8	3.5	3.0	4.3	4.73	3.02

The tables show that all methods always agree on which parameter to select first. This is not surprising since all methods in practice select the column with the largest norm first. There is very little disagreement with

Table 5.14: Percentage of different ranking results for three different methods. $n_y = 3, N = 20$

	1.	2.	3.	4.	5.	6.	7.
SO vs. MV(%)	0	1.5	4.1	4.4	15.6	24.9	14.2
SO vs. CN(%)	0	3.3	7.9	7.3	20.0	31.6	18.7
MV vs. CN(%)	0	1.8	4.0	3.4	5.0	8.7	5.5

Table 5.15: Percentage of different ranking results for three different methods. $n_y = 3, N = 50$

	1.	2.	3.	4.	5.	6.	7.
SO vs. MV(%)	0	1.3	10.1	10.0	13.2	20.6	10.3
SO vs. CN(%)	0	7.6	17.1	12.1	17.9	31.7	19.6
MV vs. CN(%)	0	6.3	7.7	2.1	4.8	13.5	10.3

regards to the most dominant parameters. Selection according to minimum additonal variance and condition number seem to agree the more for the less dominant parameters, with different choices in only 3-5% of the cases in table 5.13. This is the tendency also in table 5.14. In table 5.15, this tendency is less pronounced.

5.7.5 Conclusions on parameter ranking using 3 mass flow outputs

From the analysis using 3 outputs, the conclusion is that loss of carbon coverage will be the most important parameter in an estimation scheme. This is not surprising since it was shown in the chapter 4 that the carbon coverage to the process has a significant effect on both the steady state and dynamic behavior of the model.

The parameter which is the second choice, and which also shows a low variance all over, is the electric loss resistance. This parameter directly affects the mass conversion rate through the furnace and has a significant steady state effect.

If the parameter set is to be expanded with a third parameter, the analysis shows that this should be the reactivity of carbon. The reactivity of carbon moves the point of optimal carbon coverage, as observed in chapter 4. It also has an effect on the dynamic response to carbon coverage changes.

If a fourth parameter were to be included, this could be the average energy to hearth parameter or the height of the top shaft element. Further

expansion of the data set should include the direct fuming parameter, the height of the furnace, and last the metal bath height.

The inclusion of a fourth parameter to the set causes a significant increase in the variance. Also, one also has to bear in mind that only a small number of parameter values have been examined, and the internal ranking of parameters 4, 5, 6, and 7 therefore remains somewhat uncertain.

5.7.6 Variance in the parameter set

In the following, the variance within the parameter set is examined more closely.

Based on the mean rank for the parameters presented for different window lengths in the tables presented previously, the parameters have been included in the order $3, 1, 2, 5, 6, 4, 7$.

Tables 5.16, 5.17 and 5.18 show the average additional variance for the parameters for $N = 7, N = 20$ and $N = 50$ respectively.

Table 5.16: Average variance, $n_y = 3, N = 7$

Par.no.	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
3	0.0047	0.0048	0.0037	0.0036	0.0026	0.0042
1	0.0260	0.0265	0.0249	0.0261	0.0230	0.0242
2	0.0552	0.0514	0.219	0.136	0.0470	0.136
5	3.29	1.59	0.814	1.19	0.597	0.930
6	0.346	0.289	0.872	0.921	0.636	0.727
4	1.86	1.51	2.14	3.06	8.98	0.602
7	10.2	7.26	1.67	7.87	1.37	4.32

Table 5.17: Average variance, $n_y = 3, N = 20$

Par.no.	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
3	0.0011	0.0013	0.0011	0.0011	0.0008	0.0012
1	0.0085	0.0092	0.0088	0.0091	0.0076	0.0086
2	0.0141	0.0144	0.0547	0.0327	0.0141	0.0227
5	0.254	0.550	0.134	0.350	0.0677	0.0437
6	0.138	0.111	0.103	0.124	0.0996	0.122
4	0.619	0.315	0.242	0.363	0.204	0.236
7	0.382	0.763	0.317	0.406	0.398	0.516

Table 5.18: Average variance, $n_y = 3, N = 50$

Par.no.	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
3	0.0003	0.0004	0.0004	0.0004	0.0001	0.0003
1	0.0033	0.0036	0.0035	0.0035	0.0026	0.0034
2	0.0064	0.0053	0.0184	0.0108	0.0066	0.0122
5	0.112	0.134	0.0518	0.0945	0.0196	0.0119
6	0.0685	0.0511	0.0586	0.0565	0.0339	0.0550
4	0.111	0.0895	0.0731	0.121	0.0503	0.0766
7	0.142	0.166	0.110	0.131	0.0896	0.155

Table 5.16 shows the average additional variance for each parameter within the set does not vary much for the different parameter values (θ_1, θ_2 etc.) The exception is however parameter members 2 (reactivity) and 5 (average energy fraction to hearth), ranked as third and fourth in the list. The variance for the average energy to hearth is lower for θ_5 and θ_6 than for θ_1 to θ_4. The average additional variance of the reactivity (ranked third) is higher in θ_5 and θ_6 than in θ_2. Using a window of $N = 20$ samples gives the same tendency as for $N = 7$, namely that the average variance in the reactivity and average energy to hearth changes depending on parameter values. Also, $N = 50$ confirms the same tendency.

The plots of additional variance for each parameter in the set for $N = 20$ for θ_4 and θ_6 are therefore included in figure 5.4 and 5.5. Figure 5.4 shows that there is a leap in additional variance between parameters 1, 2, 3 all having a low variance, and parameters 4 to 7 which add a significantly higher individual variance. From figure 5.5 it can be observed that the variance in parameters 4, the energy fraction to the hearth, is generally lower. This parameter sometimes gives a lower variance than the reactivity.

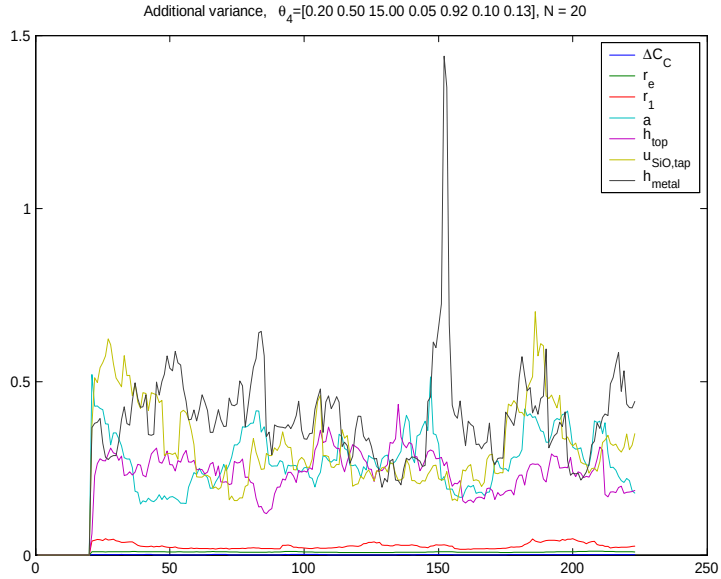

Figure 5.4: Additional variance $\theta_4, N = 20$.

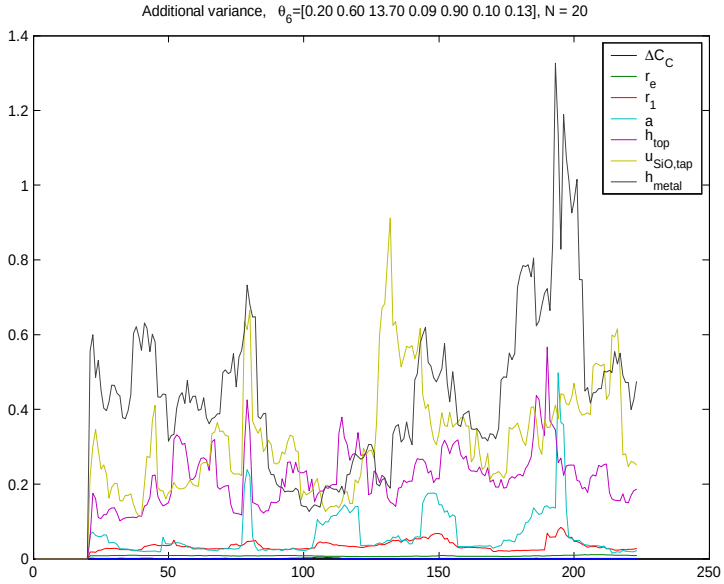

Figure 5.5: Additional variance, $\theta_6, N = 20$.

5.7.7 Additional variance vs. window length, using 3 mass flow outputs

In the following, the effect of the window length on the variance is examined. The aim of the analysis is to find a lower limit for the window length.

Previously, the ranking has been made for window lengths of 7, 20 and 50. Here the additional variance had been found as a "continuous" function of the window length. Since the variance normally decreases proportionally with the number of samples included in the estimation, the additional variance values have been multiplied with the number of samples included in each window, $(N * n_y)$. In this way it is possible to find the minimal window length by finding the number of samples below which the variance decreases with more than $1/(N * n_y)$.

In figure 5.6 the additional variance of ΔC_C and r_e lie the lowest.

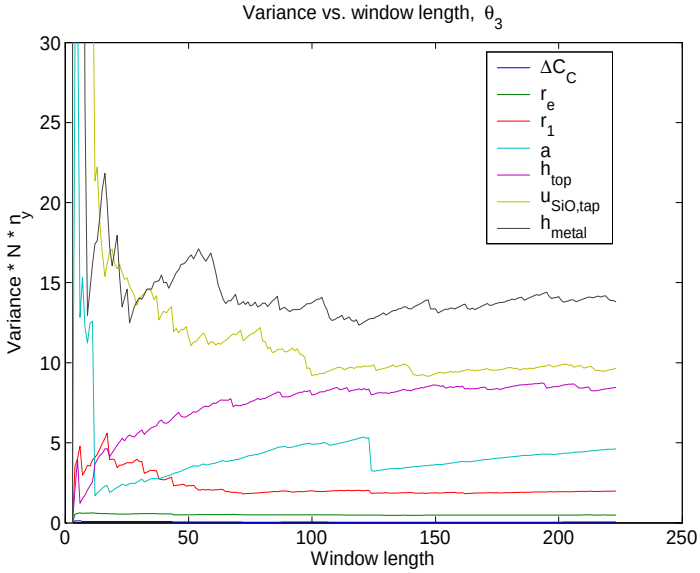

Figure 5.6: Additional variance multiplied with $n_y * N$.

There is a leap in variance between the variance contribution of these two parameters and the third and fourth, r_1 and a. The additional variance of r_1 decreases up until a window length of 50 samples.

The curve corresponding to a follows a different course than the other curves in that if first decreases rapidly, then increases slowly. This is an indication that new data actually does not decrease the variance of this

parameter. Then, at approximately 120, the variance drops. At this point, there is an excitation in the load which affects the identifiability of this parameter.

The variance contribution of h_{top} also increases slightly, but seems to reach a steady state value.

5.8 Sensitivity analysis using 2 mass flow outputs

In some furnaces, the fuming dust production is not measured and can therefore not be used for estimation.

The mean ranking of the parameters, using a window length of $N = 20$, is given in the table 5.19. The table shows that there are only minor differences in the parameters that are ranked the lowest compared to table 5.7 which included the fuming dust measurement.

Table 5.19: Avg. ranking. Orthogonalization, $n_y = 2$, $N = 50$

Par.no.	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
1	2.15	2.37	2.0	2.0	2.01	2.06
2	2.60	2.57	3.16	3.0	3.48	3.75
3	1.24	1.05	1.0	1.0	1.0	1.0
4	4.80	5.23	5.20	5.39	5.50	5.62
5	5.94	5.50	5.04	5.98	3.83	3.44
6	5.10	4.72	5.40	4.93	5.70	5.50
7	6.16	6.54	6.20	5.71	6.48	6.63

To investigate the consequence of omitting the fuming dust measurement, the increase in net additional variance for the individual parameters going from $n_y = 3$ to $n_y = 2$ outputs has been calculated. Window lengths of $N = 7$, 20 and 50 have been used. The variance has been averaged for parameter vectors θ_1 to θ_6. The increase is given in per cent.

Table 5.20 gives the net increase per parameter for the output scaling given by (5.7). The increase is low.

Table 5.20: Average added variance (%) from $n_y = 3$ to $n_y = 2$

Par.no.	$N = 7$	$N = 20$	$N = 50$
3	0.367%	0.392%	0.303%
1	0.222%	0.191%	0.130%
2	0.230%	0.232%	0.297%
5	0.414%	0.146%	0.107%
6	10.4%	3.24%	4.04%
4	16.2%	3.11%	2.97%
7	24.2%	2.79%	2.07 %

5.9 The effect of output scaling

Since the difference between including and excluding the fuming dust measurement seems to give a very small effect on the ranking an variance, a different scaling,

$$W^{-\frac{1}{2}} = diag\left[\begin{array}{ccc} \frac{1}{43} & \frac{1}{20} & \frac{1}{3} \end{array}\right] \tag{5.9}$$

has been used for calculation. This scaling gives much more emphasis to the fuming dust output.

The effect on parameter ranking has been investigated for $N = 20$, $n_y = 3$, and the ranking is given by table 5.21. The table shows that the ranking order is the same as in table 5.7 for the three most dominant parameters, but the height of the top shaft element, parameter 6, is the fourth choice, before the energy fraction to hearth, parameter 4.

Table 5.21: Average ranking, orthogonalization with scaling given by (5.9)

Par.no.	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
1	2.41	2.45	2.0	2.0	2.02	2.04
2	2.47	2.53	3.81	3.35	3.23	3.94
3	1.12	1.02	1.0	1.0	1.0	1.0
4	5.26	5.45	5.66	5.50	5.85	5.91
5	6.26	5.89	5.58	6.33	4.27	3.79
6	4.0	4.0	3.42	3.65	4.88	4.49
7	6.49	6.66	6.53	6.17	6.75	6.83

The difference in individual variance contribution is shown in table 5.22 using scaling (5.9).

Table 5.22: Average added variance (%) from $n_y = 3$ to $n_y = 2$ with scaling given by (5.9)

Par.no.	$N = 7$	$N = 20$	$N = 50$
3	25.4%	27.2%	20.9%
1	12.5%	10.8%	7.27%
2	12.9%	13.4%	17.0%
5	20.4%	7.82%	5.78%
6	77.1%	52.3%	63.4%
4	161.0%	43.2%	36.4%
7	139.0%	32.4%	21.9%

The effect of removing the fuming dust output is much larger when using the scaling in (5.9) for making the calculations of the full output configuration than when using (5.7). Also, the effect of removing the fuming dust measurements is much larger when using the full parameter set for a short window, i.e. $N = 7$, than for the longer windows.

5.10 Summary and discussion of the results

This chapter has analyzed the sensitivity derivative of the Simod model using a normalized form of the sensitivity derivative cross product to perform a manual analysis. The inspection revealed which parameters gave the strongest sensitivity, and also showed between which parameters strong linear dependence existed.

For automatic ranking, successive orthogonalization has been used. Since the deviation between the model outputs and the actual outputs is assumed to be a stochastic process with Gaussian properties, the parameters could also be ranked according to minimum additional variance.

The methods were compared in section 5.7 using three mass flow measurements as outputs, and for window lengths of $N = 7, 20$, and 50 time instants. The average ranking of the dominant parameters for all methods is mainly the same.

The tendency shown by the sensitivity analysis is that the candidate parameter set can be grouped into a "good" group and a "less good" group of parameters. The "good" group consists of parameters which give a low individual additional variance. The group consists of loss of carbon coverage, electrode loss resistance and carbon reactivity, in that order. The

ranking order of these three parameters is the same even if the scaling of the outputs is changed or if the fuming dust measurement is excluded.

The remaining four parameters in the "less good" group give a much higher additional variance, and with an internal ranking that changes depending on the parameter vector value. Among these four, the average energy to hearth fraction and the height of the shaft top element ranked in first and second place. The metal level in the hearth is ranked last. The ranking order remains the same even if the fuming dust measurement is excluded from the set. A change in scaling, to increase the weight on the fuming dust contributions, brought the height of the shaft element into fourth place.

Six different parameter vector values, θ_1 to θ_6, in combination with a 222 samples long time series of actual input data have been used in the sensitivity analysis. This is sufficient to determine that the three dominant parameters are parameters 3, 1, and 2. The ranking of the last four parameters is less conclusive, and if more than three parameters are to be estimated, the sensitivity analysis should be made for more values of the parameter vector.

However, based on the parameter vector values θ_1 to θ_6, and using 3 mass flow outputs, the ranking of the parameters is:

3. Carbon coverage loss
1. Electrode loss resistance
2. Carbon reactivity
5. Mean energy fraction to hearth
6. Height of top shaft element
4. Direct Fume loss
7. Hearth metal level

Chapter 6

On-line estimation formulation

The purpose of this chapter is to develop a parameter estimation scheme for the Simod model. This corresponds to step 6 in the list on page 9.

The chapter contains the following. In section 6.1, some general considerations are made regarding the choice of estimation schemes for rigorous, mechanistic simulation models. A literature overview is presented in section 6.2. Section 6.3 reviews some of the Simod model properties. In section 6.4 the Simod model is represented on a state space form, and a discrete model representation suitable for parameter estimation is developed. The development of an estimation criterion is made in section 6.5. Based on this criterion, a nonlinear, sliding window parameter estimation algorithm is described in section 6.6.

6.1 On-line simulation, general considerations

As pointed out in the Introduction, chapter 1, the choice of estimation scheme for a large model must be made considering the properties of the model at hand. The architecture and software implementation of the model may limit the selection of possible estimation methods, as was mentioned in chapter 1 and 3.

The possibilities for communication and outside control provided by the model architecture must be taken into consideration when estimation methods are considered. If a gradient type algorithm is preferred, and the gradient does not exist on an analytical form, the gradient may have to be obtained through perturbation of the model parameters. This requires restarting of the model from an arbitrary state vector, setting of the internal time in the model, and setting of the model parameter vector. An iterative

search algorithm also requires setting of the state vector, the time, and the model parameter vector. The resulting output produced within the model must be made available to the estimation algorithm.

The amount and quality of on-line data also affects the choice of estimation method. If the process has fast dynamics, and on-line data are sampled at a high rate, this may rule out computationally heavy methods. If constraints are not essential, a Kalman filter can be considered. If the model is rather nonlinear and there exists sufficient processing time, iterative optimization methods should be considered.

One important aspect, regardless of the complexity of the estimation method, is whether the estimation is to be done in open or closed loop. In closed loop, the estimation algorithm and estimates affect the dynamic properties of the total system, and the estimation algorithm must be analyzed as part of the total system. For closed loop identification see for instance Forsell & Ljung (1999).

6.2 Literature overview

The challenges which may be faced when applying a large model in estimation, spans most of the estimation and identification domain, and this overview therefore primarily addresses some of the topics relevant in the present case, which is the estimation of time varying parameters in a nonlinear, mechanistic model.

Time varying parameters may change in jumps or drift more or less slowly, and it is crucial that an estimation algorithm is able to detect and follow these changes. In Simod, the candidate parameters are assumed to have a drifting nature. Within the adaptive control tradition, Graham & Sin (1984, chapter 6) presents the use of a finite sliding window, exponential forgetting and bounding the covariance away from zero as techniques to ensure that the parameter estimate is allowed to vary with time. The book states that in the presence of measurement noise, it can be argued that the mean-square parameter estimation error due to the effect of the measurement noise, decreases inversely with the window length because of averaging whereas the effect of parameter drift increases proportionally to the window length. Thus, in general the mean square error has the form

$$\bar{e}^2 = \frac{K_1}{N} + K_2 N \tag{6.1}$$

where the N is the window length, K_1 is proportional to the noise variance and K_2 is proportional to the parameter drift. The optimal window length

is the length $N^* = \sqrt{\frac{K_1}{K_2}}$ that minimizes the squared error in (6.1). The minimum mean error is then $\bar{e}^{2*} = 2\sqrt{K_1 K_2}$. The product of the measurement noise and the rate of parameter drift therefore sets a lower limit to the error.

Jiang & Zhang (2004) applies a sliding window least squares for parameter estimation with exponential weighing as well as a rectangular window. Choi & Bien (1989) applies a sliding window parameter estimation algorithm with exponential weighing of the data.

Bounding the parameter covariance away from zero is another technique to ensure that the parameters are allowed to vary with time. The addition of a "process noise" covariance to the parameters estimated in an extended Kalman filter bounds the parameter covariance away from zero and ensures that the parameters are allowed to vary with time. The text book Lewis (1986) focuses on stochastic systems and optimal estimation and covers Kalman filtering and linear quadratic techniques.

In recent years there has been extensive research activity within Moving Horizon Estimation (MHE). MHE is based on the use of a fixed length moving window, and estimates a sequence of states within this window using constrained optimization. A key issue in MHE is how to summarize old information into the estimation criterion through estimation of the initial state within the window.

This literature of MHE (see page 8) is relevant, and the problem has many similarities with the time varying parameter estimation problem in Simod. The similarities are that in both cases a finite window is applied to limit the calculation time of the constrained optimization problem. Also, there is a need to determine how older information is to be incorporated and weighed in the criterion in both cases. This is discussed in Rao et al. (2001) for MHE applied to linear systems. The most important difference is that in the Simod case, the parameter vector is assumed to be constant within the time window, whereas in the MHE case, a state vector is estimated for each time instant within the window. Increasing the window size with assumed constant parameters means averaging the parameter estimate over more data, and increasing the required processing time. In the MHE case, an increased window length also implies increasing the number of optimized variables.

6.3 Properties of Simod affecting the design of the estimation scheme

In the Simod case, a set of time varying parameters is to be estimated. Some of the parameters directly affect the mass and energy balances of the model, and can also be viewed as process disturbances. This allows for correcting the model's mass and energy balance vs. actual data, still letting the model make sure that its mass and energy balances are fulfilled.

The Simod model is nonlinear. The component based architecture of the simulator/model makes it possible to control the execution of the model from an outside program. The model "time" can be set from an outside algorithm, and the model can be restarted from an arbitrary state and parameter vector on-line. This allows for calculation of a gradient through perturbation of the parameter vector and the use of iterative methods.

The 8 hour sampling time allows for considering computationally heavy methods, if needed. The only limit to calculation time is how long a user is willing to wait after new data has arrived. Also, due to the limited amount of on-line data available, it is important to make sure that the data is utilized in a best possible way. Iterative methods should therefore be considered.

In chapter 2, the available measurements from the process were presented and evaluated in light of the Simod model. It was concluded that there are significant uncertainties in many of the inputs, represented as parameters in the model and a candidate set of parameters for estimation was identified. In chapter 4, the effect of the most important inputs on the outputs were analyzed, and it was shown that the carbon coverage, electric power supply and carbon reactivity determine the optimal operating point for the process. In chapter 5, the sensitivity derivative of the candidate parameter vector corresponding to different values of the parameter vector was analyzed. The conclusion was that within the candidate parameter set, the loss of carbon coverage, the loss of energy input and the carbon reactivity showed better identifiability properties than the rest. The effect of the varying the window length on the output variance was also studied in the sensitivity analysis chapter 5.

Computation time is a factor when determining the number of parameters to be estimated and window length. More parameters generally also require a longer window to ensure good identifiability.

To ensure good identifiability in general, and to limit the window to a reasonable length, a term penalizing deviation from previous parameter estimates should be included in the criterion. Also, constraints on the pa-

rameters should be applied to ensure sound model behavior and to avoid that the estimated parameters take values which are not realistic.

6.4 Definition of the system

The Simod model can be represented with the following equations

$$\dot{x} = f(x,u,\theta,\rho,\eta) \tag{6.2}$$

$$g(x,\theta,u) = 0 \tag{6.3}$$

where:

- x vector of state variables
- u inputs
- θ vector of parameters assumed to be slowly varying with time
- ρ vector of known parameters
- η on-off control, $\eta \in (0,1)$
- T integration period

The observations made on the actual process, y_t, can only be taken as aggregated values over a time period of T,

$$y_t = \int_{t-T}^{t} Dx \; dt + w_k \tag{6.4}$$

The inputs, u_t, to the process are registered as averaged input values over the time period T,

$$u_t = \frac{1}{T} \int_{t-T}^{t} u \; dt \tag{6.5}$$

The system is assumed to exhibit non-minimum phase behavior in some operating regions, such as time delays and positive zeros. The process itself is assumed to be unstable in the sense that it has integrating modes in some operating regions, see figure 4.1 on page 46. These modes are stabilized by the on-off controllers which are considered part of the total system, and by nonlinear stabilizing mechanisms in the process itself. The total system (process and controllers) is therefore assumed to be stable.

6.5 Criterion design

To ensure a formulation which gives good identifiability and a reasonable calculation time, a number of rather brutal assumptions are made in order to formulate an estimation criterion based on the Simod model.

The Simod model is a continuous state space representation on a differential algebraic equation form, according to (6.2) and (6.3). The inputs and outputs are however only available as aggregated or averaged values at discrete time intervals, according to (6.5) and (6.4). Some of the parameters of the process are assumed to be slowly varying with time, also relative to the 8 hour sampling interval.

We assume that the uncertainties in the model vs. the real process lie in the time varying parameters, and that there are no other uncertainties to the states. The time varying parameters in θ_k are mainly properties of the inputs to the process, see chapter 3. An alternative could therefore be to view the parameters as time varying process disturbances, and the estimation problem as a process disturbance estimation problem. The parameters, or process disturbances are however not additive in nature.

Since the outputs and inputs are measured at discrete time intervals, a discrete model is assumed for (6.2), shown in (6.6) to (6.8). The vector of time varying parameters is also included as a time varying state, and the augmented state vector therefore consists of $\{x, \theta\}$ giving the model

$$x_{k+1} = f(x_{k,}\theta_k, u_{k+1}) \tag{6.6}$$

$$\theta_{k+1} = \theta_k + v_k \tag{6.7}$$

$$y_k = Dx_k + w_k \tag{6.8}$$

The inputs v_k and w_k are independent disturbance vectors. u_{k+1} is an input vector aggregated from k to $k+1$, and registered at $k+1$.

Due to the assumptions about θ holding the model uncertainties, x_{k+1} will be completely defined by x_0, θ_k and the measured inputs u_k.(6.6) to (6.7) can therefore be rewritten in a more compact form where a nonlinear measurement function includes the state transition equation. The following stochastic model is now the basis for estimation.

$$\theta_{k+1} = \theta_k + v_k \tag{6.9}$$

$$y_k = h(x_0, u_1, ..., u_k, \theta_1, ..., \theta_k) + w_k \tag{6.10}$$

6.5.1 Preliminaries. Bayes' rule.

The conditional probability of event a, given that we know the outcome of event b, is defined through Bayes' rule:

$$p(a|b) = \frac{p(ab)}{p(b)}$$

where $p(ab)$ is the joint probability of event a and b. $p(b)$ is the probability of event b. For three events Bayes' rule can be used in two steps giving

$$\begin{aligned} p(abc) &= p(ab)p(c|ab) \\ &= p(a)p(b|a)p(c|ab) \end{aligned}$$

This can be developed into the general multiplication sentence for conditional probabilities

$$p(a_1, ..., a_j) = p(a_1)p(a_2|a_1)p(a_3|a_1, a_2)...p(a_j|a_1, a_2..., a_{j-1}) \tag{6.11}$$

If the events $b_1, ..., b_j$ is a sequence of events of a Markov process, then the event b_j is only dependent on the most recent event, b_{j-1}, and independent of previous events. For a Markov process the multiplication sentence (6.11) therefore becomes

$$p(b_1, ..., b_j) = p(b_1)p(b_2|b_1)p(b_3|b_2)...p(b_j|b_{j-1}) \tag{6.12}$$

6.5.2 Development of a criterion from conditional probabilities

A fundamental assumption in this parameter estimation case is that the parameter vector can be assumed to be relatively constant within a certain time window. Assuming that the simulation is started time 0, the first output vector is y_1. A window of outputs from 1 to N is used to find the first parameter estimate.

Since it has already been assumed that the model errors lie in the parameter vector, and that there are no additional uncertainties exist in the state representation, only the initial state, x_0, and the parameter vector, θ, are left as the unknowns in (6.9) and (6.10). If the error in the initial guess of x_0 can be assumed to be small and die out quickly, x_0 is disregarded as an unknown variable, and θ_0 is the only unknown left in the first window. The solution to the problem is described in terms of a prediction/correction scheme in the following.

The optimal, corrected estimate of θ at time N, and given x_0, $\{u_k\}_1^N$ and $\{y_k\}_1^N$ is the estimate that maximizes the conditional probability density p

$$\{\hat{\theta}_{N|N}\} = \arg\max_{\{\theta\in\Theta\}} p(\theta|x_0, y_1, ..., y_N, u_1, ..., u_N) \tag{6.13}$$

Note that the input u_1 is an aggregated value from $T = 0$ to $T = 1$, registered at $T = 1$. u_1 is applied to the model from $T = 0$.

The conditional probability of the outputs given the parameter vector, the initial state and inputs is

$$\begin{aligned} p(y_1, ..., y_N|\theta, x_0, u_1, ..., u_N) &= \prod_{k=1}^{N} p(y_k|\theta, x_k, u_k) \\ &= \prod_{k=1}^{N} p(y_k - h(x_k, u_k, \theta)) \end{aligned}$$

An expression that maximizes (6.13) is developed in the following.

$$\begin{aligned} & p(\theta|x_0, y_1, ..., y_N, u_1, ..., u_N) \\ & = \frac{p(\theta, x_0, y_0, ..., y_N, u_1, ..., u_N)}{p(x_0, y_1, ..., y_N, u_1, ..., u_N)} \\ & = \frac{p(\theta)p(y_1|x_0, \theta)p(y_1|x_0, \theta)..p(y_N|x_0, \theta)}{p(x_0, y_1, ..., y_N, u_1, ..., u_N)} \\ & \propto p(\theta)\prod_{k=1}^{N} p(y_k - h(x_k, u_k, \theta)) \end{aligned} \tag{6.14}$$

Taking the negative logarithm of the expression in (6.14), gives

$$\{\hat{\theta}_{N|N}\}_{k=1}^{N} = \arg\min_{\{\theta\in\Theta\}} \sum_{k=1}^{N} L(w_k) + \Gamma(\theta) \tag{6.15}$$

where

$$L(w_k) = -\log p(y - h(x_0, u_1, ..., u_k, \theta))$$

and

$$\Gamma(\theta) = -\log p(\theta)$$

The model error $w_k = y_k - h(x_k, u_k, \theta)$ in (6.10) is assumed to be Gaussian $\mathcal{N}(0, W^{\frac{1}{2}})$, $W > 0$. giving

$$\{\hat{\theta}_{N|N}\}_{k=1}^{N} = \arg\min_{\{\theta\in\Theta\}} \Gamma(\theta) + \sum_{k=1}^{N} w_k' W^{-1} w_k$$

The term $\Gamma(\theta)$ makes it possible to incorporate a priori knowledge about the parameter vector and penalize the deviation from the a priori parameter estimate. It is difficult to argue from a probability distribution point of view with regards to the deviation from the a priori estimate. The reason for this is that there are upper and lower bounds on the parameter vector, and the deviation is therefore bounded. A quadratic penalty has still been chosen for the deviation between the a priori $\bar{\theta}_N$ and a posteriori estimate, $\hat{\theta}_{N|N}$, giving the complete criterion

$$\{\hat{\theta}_{N|N}\}_{k=1}^{N} = \arg\min_{\{\theta\in\Theta\}} \left[(\theta - \bar{\theta}_N)' \bar{P}_0^{-1} (\theta - \bar{\theta}_N) + \sum_{k=1}^{N} w_k' W^{-1} w_k \right] \quad (6.16)$$

When sample y_1 is discarded and y_{N+1} is included, then according to (6.9) the predicted value of θ at $N+1$ based on $y_1, ..., y_T$ is

$$\bar{\theta}_{N+1|N} = \hat{\theta}_{N|N}$$

For the next window, an initial state at $T = 1$ is needed. This is obtained by calculating the state at $T = 1$ from (6.6)

$$x_1 = f(x_0, \bar{\theta}_{N+1|N}, u_1, \eta)$$

At the next step, we therefore want to find the parameter estimate that maximizes

$$\{\hat{\theta}_{N+1|N+1}\} = \arg\max_{\{\theta\in\Theta\}} p(\theta|\bar{\theta}_{N+1|N}, x_1, y_2, ..., y_{N+1}, u_2, ..., u_{N+1})$$

which along similar arguments as for the first window yields the nonlinear constrained optimization problem

$$\{\hat{\theta}_{N+1|N+1}\}_{k=1}^{N+1}$$
$$= \arg\min_{\{\theta\in\Theta\}} (\theta - \bar{\theta}_{N+1|N})' \bar{P}_{N+1|N}^{-1} (\theta - \bar{\theta}_{N+1|N}) + \sum_{k=2}^{N+1} w_k' W^{-1} w_k \quad (6.17)$$

Using the notation $\bar{\theta}_{N+1} = \bar{\theta}_{N+1|N}$, $\hat{\theta}_{N+1} = \hat{\theta}_{N+1|N+1}$, $\bar{P}_{N+1}^{-1} = \bar{P}_{N+1|N}^{-1}$, and $T = N + 1$, this can be further developed to

$$\{\hat{\theta}_T\}_{k=1}^{T} = \arg\min_{\{\theta\in\Theta\}}(\theta - \bar{\theta}_T)'\bar{P}_T^{-1}(\theta - \bar{\theta}_T) + \sum_{k=T-N+1}^{T} w_k' W^{-1} w_k \qquad (6.18)$$

The inclusion of a term penalizing the deviation from the estimate of the previous window in the optimization criterion implies that older samples contribute. This part can therefore be regarded as a filtering term since it leads to further averaging of the estimate. The term will also contribute to the ensuring identifiability of the estimation problem, or convexity of the optimization problem.

Depending on the value of $\bar{P}_T^{-1}$, the term $(\theta - \bar{\theta}_T)'\bar{P}_T^{-1}(\theta - \bar{\theta}_T)$ will incorporate information about older samples to a higher or lower degree. The effect of emphasizing this term is therefore similar to increasing the window length. Since calculation time is an issue in the Simod case, the term can be used to shorten the window by increasing the emphasis to the filtering term. The window must however be long enough to ensure that a sensible gradient can be calculated. The balance between window length, $\bar{P}_T$ and W is therefore an important tuning issue. In the following an updating scheme for the time varying weight $\bar{P}_T^{-1}$ is developed.

6.5.3 Assessing the time varying weight $\bar{P}_T^{-1}$

Given that the weight $\bar{P}_T^{-1}$ is to be time varying, a sensible variation scheme would have $\bar{P}_T^{-1}$ small at startup since good a priori assumptions may be difficult to make. $\bar{P}_T^{-1}$ should become larger as time progresses since more and more samples have been involved in the estimation, and therefore should express a higher certainty that the obtained estimate is good. $\bar{P}_T^{-1}$ should however not be allowed to grow arbitrarily large since a too large $\bar{P}_T^{-1}$ would give a too high weight to old data, and let new observations contribute too little.

When assessing the time varying weight $\bar{P}_T^{-1}$, it is difficult to argue on a probabilistic basis since the parameter vector variation range is bounded, and the deviation from the true parameter value, θ_T^*, will therefore also be bounded.

The criterion in (6.18) does however resemble other well-known criteria within the parameter and state estimation domain. With a window length of 1, without constraints on the parameters, the criterion in (6.18) with the model equations (6.6) to (6.8), would be identical to the criterion of an

extended Kalman filter see for instance Bryson & Ho (1975). In an extended Kalman filter, the a posteriori parameter estimate found by solving the following linearized, unconstrained criterion

$$\hat{\theta}_T^k = \arg\min_{\theta}(\theta - \bar{\theta}_T^k)' \bar{P}_T^{k-1}(\theta - \bar{\theta}_T^k) + (y_T - \frac{dy(\bar{\theta}_T^k)}{d\theta}\theta)' W^{-1}(y_T - \frac{dy(\bar{\theta}_T^k)}{d\theta}\theta)] \tag{6.19}$$

having the solution

$$\hat{\theta}_T^k = \bar{\theta}_T^k + \hat{P}_T^k \frac{dy(\bar{\theta}_T^k)}{d\theta}' W^{-1}(y_T - h(\bar{\theta}_T^k)) \tag{6.20}$$

$$\hat{P}_T^k = \bar{P}_T^k - \bar{P}_T^k \frac{dy(\bar{\theta}_T^k)}{d\theta}' \left[\frac{dy(\bar{\theta}_T^k)}{d\theta} \bar{P}_T^k \frac{dy(\bar{\theta}_T^k)}{d\theta}' + W \right]^{-1} \frac{dy(\bar{\theta}_T^k)}{d\theta} \bar{P}_T^k \tag{6.21}$$

The a priori values at $T+1$ would be predicted based on

$$\begin{aligned} \bar{\theta}_{T+1}^k &= \hat{\theta}_T^k \\ \bar{P}_{T+1}^k &= \hat{P}_T^k + V \end{aligned}$$

Since the covariance updating scheme for the extended Kalman filter can give the properties of $\bar{P}_T^{-1}$ mentioned initially, this is a good candidate for use for the parameter estimates in Simod. The scheme could be utilized as follows

$$\{\hat{\theta}_T\}_{k=1}^T = \arg\min_{\{\theta \in \Theta\}}(\theta - \bar{\theta}_T)' \bar{P}_T^{-1}(\theta - \bar{\theta}_T) + \sum_{k=T-N+1}^{T} w_k' W^{-1} w_k \tag{6.22}$$

where $w_k = y_k - h(x_k, u_k, \theta)$, and

$$\hat{P}_T = \bar{P}_T - \bar{P}_T' \psi_T' \left[\psi_T \bar{P}_T' \psi_T' + W \right]^{-1} \psi_T \bar{P}_T \tag{6.23}$$

$$\begin{aligned} \bar{\theta}_{T+1} &= \hat{\theta}_T \\ \bar{P}_{T+1} &= \hat{P}_T + V \end{aligned} \tag{6.24}$$

where $\psi_T \in \mathbb{R}^{n_y \times n_\theta}$is the sensitivity derivative corresponding to the last, and only new sample within the window which has the total sensitivity

derivative Ψ_T,

$$\Psi_T = \begin{bmatrix} \psi_{T-N+1} \\ \psi_{T-N+2} \\ ... \\ \psi_{T-1} \\ \psi_T \end{bmatrix}_{|\hat{\theta}_T} = \begin{bmatrix} \frac{dy_{T-N+1}}{d\theta} \\ \frac{dy_{T-N+2}}{d\theta} \\ ... \\ \frac{dy_{T-1}}{d\theta} \\ \frac{dy_T}{d\theta} \end{bmatrix}_{|\hat{\theta}_T} \tag{6.25}$$

with $\Psi_T \in \mathbb{R}^{(N*n_y)\times n_\theta}$.

There are however several differences between the general nonlinear criterion in (6.22) and the extended Kalman filter:

- (6.22) is a constrained criterion, while the extended Kalman filter is unconstrained.
- $\hat{\theta}_T$ is found through iterative search.
- The corrected $\hat{P}_T^k$ in the Kalman filter is found using a gradient at $\bar{\theta}_T^k$, while $\hat{P}_T$ is found using gradient at $\hat{\theta}_T$.
- For the solution of (6.22) – (6.24), the output data are used several times due to the sliding window.

These differences are commented further in the following.

The effect of constraints

The primary effect of the constraints is that the size of the parameter error $\left\|\hat{\theta}_T - \theta_T^*\right\|$ is bounded, provided the correct parameter vector $\theta_T^* \in \Theta$. In the Kalman filter formulation, there is no guarantee that the a posteriori parameter estimate lies within Θ. Still, nothing general can be said about the size of the deviation in the unconstrained case compared to the constrained case, due to the nonlinear model equations.

With linear model equations, P_T , $W > 0$ for convexity and $\theta_T^* \in \Theta$, the unconstrained error would give an upper bound for the constrained error.

$$\left\|\hat{\theta}_T - \theta_T^*\right\| \leq \left\|\hat{\theta}_T^k - \theta_T^*\right\|$$

The effect of iteration

The primary effect of iteration is generally that the criterion function value is smaller. Other than this, it is impossible to say if using the a priori gradient as in (6.21) or the a posteriori gradient as in (6.23) will give a smaller or larger variance given that the model equations are nonlinear.

The effect of a sliding window

The most important effect of a sliding data window is that the output data are used many times. This effect is important to take into account when establishing an update algorithm for the parameter covariance, so that one sample is not allowed to "improve" the covariance more than once. This problem is also addressed in MHE where an issue is to estimate the initial state within the window. If the initial state is at time $T-N$, then $\hat{x}_{T-N|T-N}$ or $\hat{x}_{T-N|T-1}$ can be used as initial states. As $\hat{x}_{T-N|T-N}$ has been estimated using only samples up to and including sample $T-N$, the estimate $\hat{x}_{T-N|T-1}$ has been estimated using both preceding and succeeding samples. Calculation of the covariance associated with $\hat{x}_{T-N|T-1}$ therefore has to be addressed differently than for $\hat{x}_{T-N|T-N}$ to avoid using the actual outputs more than once.

Estimation based on only preceding samples is called filtering, whereas the use of previous and succeeding samples is called "smoothing". Optimal smoothing is a separate section in the book Lewis (1986). In connection with MHE, Rao (2000) develops an expression for the smoothed covariance of $\hat{x}_{T-N|T-1}$ from probabilistic arguments. Rao shows how the smoothed covariance is calculated using a forward and backward "sweep".

In the Simod case, a simpler approach has been chosen, as only the sensitivity derivative corresponding to outputs which have not previously been used, are included in the covariance updating equation. In the first window, however, all output samples used in the optimization criterion are new, and therefore all are included in the correction of the parameter covariance.

It may seem like a contradiction to the previous arguments about the necessity of using a window to ensure identifiability, that only the last part of the sensitivity derivative is used for correction of the covariance. The strategy has however proven robust. One way of interpreting the approximation is to say that it is assumed that the gradients corresponding to $\hat{\theta}_{T-1}$ and $\hat{\theta}_T$ is fairly equal for the overlapping samples. This may or may not be the case. There is also a risk that the simple gradient updating scheme may give a bias in the variance. This does however have little consequence for

the algorithm since is added V each time the covariance is predicted, see (6.24).

6.6 Implemented estimation algorithm

On an overall level, the estimation algorithm which implements the nonlinear, constrained optimization using a sliding window of past data contains the following steps.

Algorithm 6.1 *Simod sliding window parameter estimation.*

Step 0: Initialization

Initialize model from a .simod file.
Set $\bar{\theta}_0$
Simulate model using m samples from input time series.
$x_{init} = x_m$
Set P_0
$T = m + N$

Step 1: Load $[y_{T-N+1}, ..., y_T]$, $[u_{T-N+1}, ..., u_T]$ *and other inputs to SQP algorithm.*

Step 2:Perform nonlinear constrained optimization using SQP algorithm
using $[y_{T-N+1}, ..., y_T]$, $[u_{T-N+1}, ..., u_T]$ *and iterate to find the* θ *that minimizes*

$$(\theta - \bar{\theta}_{T-N|T})' \bar{P}_T^{-1} (\theta - \bar{\theta}_{T-N|T}) + \sum_{k=T-N+1}^{T} w_k' W^{-1} w_k \tag{6.26}$$

subject to $\theta \in \Theta$.
Result: $\hat{\theta}_T$ *and* $\Psi_{T-N+1,T}$

Step 3: Correct the covariance

if (firsttime)
Correct the covariance using
$\Psi_{1,N} = [\psi_1, ..., \psi_N]'$
else
$\hat{P}_T = \bar{P}_T - \bar{P}_T \psi_T' (W + \psi_T \bar{P}_T \psi_T')^{-1} \psi_T \bar{P}_T$

Step 4: Update state vector for reinitialization at $T + 1$

$x_{T-N} = x_{init}$
Set $\hat{\theta}_T$ *in the model*
Simulate from $(T - N)$ *to* $(T - N + 1)$

$$x_{init} = \bar{x}_{T-N+1|\hat{\theta}_T}$$

Step 5: Predict θ *and* P

$$\bar{\theta}_{T+1} = \hat{\theta}_T$$
$$\bar{P}_{T+1} = \hat{P}_T + V$$

Step 6: $T = T + 1$

Repeat from Step 1.

6.6.1 Remarks and practical implementation

The loop illustrated in algorithm 6.1 is implemented as several executable programs which have been programmed in C and C++. The looping is provided by a main program which starts the other programs. The different programs communicate via files.

Choice of SQP algorithm

Step 2 in the Simod estimation algorithm is performed by an SQP (Nocedal & Wright 1999) based nonlinear optimization algorithm called Modelfit, developed by Cybernetica/Sintef. Modelfit uses a least squares criterion for the model error, and has been designed for off-line, batch applications. The term penalizing deviation from the previous estimate in (6.18) has been implemented through defining the parameter vector as part of the output vector.

The nonlinear SQP algorithm obtains the gradient through perturbation and calculation of a central difference approximation. Modelfit requires initialization files necessary to define input data, covariance matrices etc. These initialization files are generated by MATLAB scripts. In addition, a model interface must be programmed. This interface exists as a C language template, and must filled in by the application developer.

The results produced by Modelfit, the parameter vector and gradient among them are stored.

There are numerous software packages for nonlinear optimization, but the possibilities for interfacing to the Simod model limits the actual selection. The MEX interface developed early in the doctoral project would in principle permit the use of optimization packages running under Matlab. Two such packages were tested with limited success. One problem was memory leakage and error messages which were difficult to interpret since debugging was not possible across the interfaces. The memory leakage problems could probably have been overcome with some more effort. The lack of transparency in many of the optimization packages and also the

lack of debugging possibilities, favored Modelfit over Matlab-based tools. Modelfit supports the interface of the Simod model, and makes it possible to debug the total solution. Also, much of the key source code of Modelfit was available to the author. This transparency made it easier to interpret the results of the optimization, and also find the causes of problems along the way.

Handling of actual process data

Other than Step 2 in the algorithm 6.1 on page 130, the software has been developed by the author. Step 1 supplies the input and output data from an Excel file. The data from the process have been inspected manually, mostly to ensure that readable values are placed in each cell in the excel file.

An important topic when using logged process data, is wild point editing. The logged values used here are however shift values, and therefore some quality control of the data had been made before the data have been entered into the process data base.

Modelfit also provides the possibility to "flag out" the use of values in the output vector. The corresponding estimated and measured output will then be excluded from the calculation of the criterion. This mechanism is used if there have been periods of abnormal operation. The model may not be able to follow the actual data too well during such periods. If such data are included in the quadratic criterion, these outputs may dominate over the more normal data in the optimization. In principle there are several ways of reducing the impact of such data. For instance, model errors over a certain limit could be weighed linearly instead of quadratically. Yet, to avoid having to change the source code of Modelfit itself, a lower limit has been set to the taprate and dust measurement data. If the actual data lie below these limits, all outputs at this time instant are omitted from the criterion through setting a control flag. If this output vector lies as the last sample of the window, the covariance of the estimate is not updated.

Covariance updating

Step 3 performs covariance updating. Cholesky factorization (Bierman 1977) has been implemented in order to ensure symmetry in $\hat{P}_T$. The same software also stores and time tags the result from Modelfit. The parameter estimate and predicted covariance matrix are also stored in the Modelfit input files for use in the next loop.

Initialization and initial state vector

The model is initialized from a steady state situation found through simulation for a particular furnace. This steady state situation, along with the initial parameter values, represent a well under-coked situation. The reason for doing this is to ensure that the furnace model is in a region where the dominant time constants are small, and a give a quick response to the inputs, see chapter 4. Approximately 50 samples from the actual data is used initialization.

Ideally, the initial state of the model should have been estimated to ensure that the process and the model starts from the same state. The failure to do this will cause bias in the estimated parameters.

There are important practical aspects regarding this. One is that the Simod model is not observable. However, one could argue that the states that are actually observable states should have been estimated. Another important aspect is calculation time, and obtaining the sensitivity derivative for a large number of states is time consuming. The calculation time for the gradient is as follows

$$t_c = 2 * n_s * 2.5s * N$$

where t_c is the total time, n_s is the number of states, N is the window length. The Simod model requires $2.5s$ for simulation of 8 hours on a $2.52GHz$ Pentium 4 computer. Positive and negative perturbation is necessary. With a window length of for instance 40, and if 50 of the states where to be estimated, the calculation time would be almost three hours to calculate the gradient once. This number has to be multiplied with a likely number of iterations, and some time also needs to be added for simulations for search for the optimal step length in the gradient direction. As computer processing capacity increases, this topic could however be reconsidered.

Another aspect, which has not been examined, is that the initial carbon content in the furnace may give a similar sensitivity in the outputs as the loss of carbon coverage parameter, although with opposite sign. Therefore if the estimates corresponding to the first windows should be interpreted in light of the initial state. If for instance the estimated carbon coverage loss is very much lower than expected, it may be that this is due to that carbon content in the furnace was much lower in the initial state vector than in the real data.

The reason for giving a high initial reactivity value is also to make sure that the model is in the under-coked region.

Chapter 7

Estimation results

The estimation scheme presented on page 6.1 in chapter 6 has been applied for estimation of parameters using actual process data from two different furnaces, called furnace 1, cf. section 7.1 and furnace 2, cf. section 7.2. Furnace 1 is the primary case. This chapter corresponds to step 7 in the list on page 9.

The implementation of algorithm 6.1 has been made to ensure that the user has flexibility with regards to the number of included parameters and outputs, the window length and other tuning parameters to the algorithm. Parametrization of the algorithm is therefore necessary. These issues are addressed in detail for furnace 1 in section 7.1.1.

To ensure a reasonable computation time, the estimated parameter set in this chapter consists of the three parameters. These are the parameters which ranked the highest in the sensitivity analysis of chapter 5, i.e. the specific electrode resistance, r_e, the carbon reactivity, r_1, and the carbon coverage loss, ΔC_C.

Section 7.1.2 presents the estimation results when all 3 mass flow measurements are included in the estimation criterion. The results are discussed in section 7.1.3.

Since not all furnaces have an individual fuming dust measurement, it is of interest to analyze the effect of removing this measurement on the parameter estimates, cf. section 7.1.4.

In section 7.2, input and output data from furnace 2 have been used. These data contain a bias in the mass balance, and the estimation has been made after corrections has been made to the fuming dust measurement.

7.1 Estimation results, furnace 1

The input data used for furnace 1 correspond to a three month operation period and are the same input data used for the sensitivity analysis in chapter 5. The 50 first samples in figures 5.1-5.3 have been used for initialization of the model, and the succeeding samples for estimation.

7.1.1 Parametrization, furnace 1

The user can make several choices regarding parametrization of the algorithm

- Number of parameters to be estimated
- The taprate of silicon can in principle be used both as an input and an output
- Determine the minimum window length
- Window length vs. weighing of a priori parameter estimate and noise covariance

The algorithm can estimate the whole candidate parameter set, or only a subset. Here, only 3 parameters are estimated, as previously mentioned.

The way the optimization criterion is formulated in (6.22) – (6.24) it can be seen that increasing V or decreasing W gives less weight to previous parameter estimates. Also, increasing N gives also less weight to previous parameter estimates. From a pragmatic point of view, these three quantities can be viewed as tuning variables to the estimation algorithm. Therefore, it may not be easy to know where to start when assigning values to W, V and N.

In the following, sensible values are first assigned to W. The reason for starting with W is that the covariance in the measurement error in the actual outputs can be used as a lower limit to W. However, the measurement error is not the only cause of deviation between acutal and estimated outputs, as will be discussed in the following.

The contents of W

The causes of the deviation between the actual measurement vector and simulated outputs are

- Measurement noise (and other measurement errors)
- Unmodelled behavior in the process
- Effects of averaging of particularly the inputs and also outputs
- Bias in the parameter estimates and model states

The individual contribution of the three last bullets is difficult to assess. The contribution of the measurement noise is however independent of the estimated outputs and can therefore be addressed individually.

Measurement noise In the Simod case, the outputs are weighed mass flows. The shift values used in the estimation are a sum of several, individually weighed batches. If each of measurement has a variance of σ_1^2, then M batches give a variance

$$\sigma_M^2 = M\sigma_1^2$$

In silicon production, quartz is added in several small, individually weighed batches. There may be around one hundred such small batches each shift. Silicon is weighed by weighing the ladle, but there are relatively few such ladles during one shift. To compare the effect of few large batches or many small batches on the covariance, the following hypothetical situation is analyzed. Assume that 10 Tons of material is weighed in 100 batches with standard deviation 0.5% of $100kg$, then the total batch of $10Tons$ will have the standard deviation

$$\sigma_{100} = \sqrt{100} \cdot 0.5kg = 5kg$$

If the total batch was split in 2 and weighed with a standard deviation of 0.5% of $5Tons$ the following measurement standard deviation would result

$$\sigma_{100} = \sqrt{2} \cdot 25kg = 35kg$$

In connection with the data used for estimation in this case, and without going into detail about the underlying figures, the variance due to the measurement noise contributions, σ_m^2 is set to

$$\sigma_m^2 = diag \left[\begin{array}{ccc} 17.75^2 & 43.3^2 & 15.5^2 \end{array} \right]$$

The diagonal elements correspond to the shift value of quartz, silicon weight and fuming dust, respectively. The measurement errors are assumed to be independent. The relatively large figure for the silicon measurement is mainly due to the fact that silicon is weighed in large batches.

Unmodelled behaviour. Any model will be a simplification, and a number of assumptions and idealizations have been made in the Simod modelling process. Some important examples are mentioned here.

One important idealization in the Simod case is the discretization of the model in finite volumes and the assumption about homogenous conditions within the volume. This eliminates gradients in the horizontal direction which are always present in the real furnace. Some of these effects have been taken care of through parameters in the model, for instance through quantifying the chemically active cross section of the furnace.

Also, there are idealizations and uncertainties concerning the reactions and reaction rates. There is for instance some uncertainty to how well the model represents the furnace close to over-coking which was discussed in chapters 3 and 4.

Effects of averaging in inputs and outputs. The inputs used are 8 hour averages. This evidently averages out variations in the inputs, and subsequently variations in the model outputs.

Fixed parameters. Only a small set of parameters are estimated. This does in principle lead to a bias in both the estimated parameters as well as in those left fixed. By altering only a small subset of parameters through estimation, there are limitations to how well the behavior of the process can be mimiced.

Some of the properties defined as fixed parameters in the model may also very well be time varying. This also leads to a "calmer" behavior of the model than of the actual process.

Possible choices of W

The impact of unmodelled behavior, effects of averaging and fixed parameters is difficult to assess. There are several possible approaches. One is to use the measurement error covariance which is estimated to

$$W_1 = diag \begin{bmatrix} 17.75^2 & 43.3^2 & 15.5^2 \end{bmatrix}$$

Another possibility is to use the scaling matrices used in the sensitivity analysis

$$W_2 = diag \begin{bmatrix} 42.9^2 & 20.1^2 & 105.4^2 \end{bmatrix}$$

In the comparison of W_1 and W_2, the relative weight given to the outputs are of interest. W_1 would give more weight to the quartz and fuming dust

measurements, while less weight would be given to the silicon measurement. In W_2 the deviation in the fuming dust output will be emphasized less than quartz and silicon.

Both W_1 and W_2 were used for estimation, along with other values for W. Based on estimation results with these rather different values of W, this matrix was finally chosen as the covariance matrix for the model error

$$W_3 = diag \begin{bmatrix} 7284800 & 5505537 & 303772 \end{bmatrix} \tag{7.1}$$

The matrix also acts as a scaling matrix since it contains the natural variation range of the outputs. From a pragmatic point of view, it is the relative size of the elements of W that matters.

Choice of N and V

The window length N should be long enough to ensure that the problem is identifiable, but not longer than necessary due to the significant processing time of the model. A window length of 12 samples has been chosen. This means in effect that simulated and actual output values from 10 time samples are included in the criterion. This is regarded as a short window.

The choice of V in (6.24) determines the degree of filtering of the parameter estimates. The size of the diagonal elements of V has been chosen so that the contribution of the filtering term of the optimization criterion is approximately 10%

$$V = diag \begin{bmatrix} 0.0099^2 & 0.01^2 & 0.375^2 \end{bmatrix} \tag{7.2}$$

The values of V in (7.2) therefore represent a relatively low degree of filtering of the parameter estimates.

7.1.2 Estimation results, 3 parameters using 3 outputs, furnace 1

Figures 7.1-7.3 show the parameter estimates for specific electrode resistance, carbon reactivity and loss of carbon coverage using the quartz feed, tapped silicon, and fuming dust production measurements in the estimation criterion. The energy fraction to the hearth has not been estimated, but a mean value has been specified, and a time varying parameter to the model based on measurements from the process according to (3.17) on page 33.

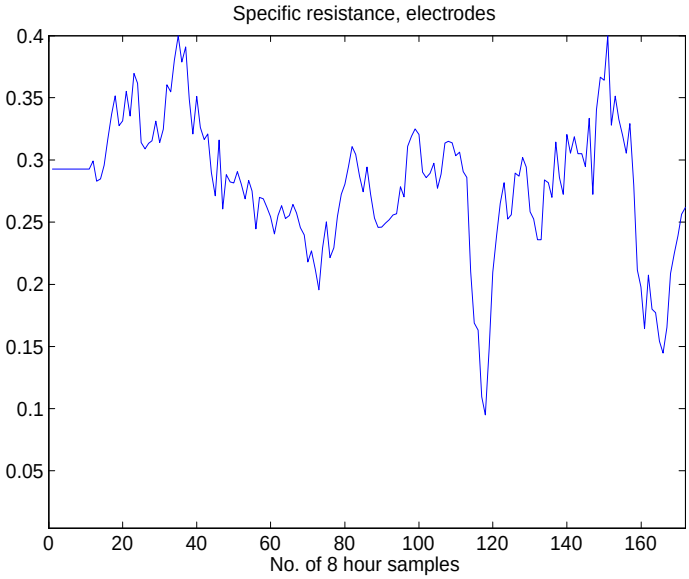

Figure 7.1: Estimated electrode specific resistance, r_e.

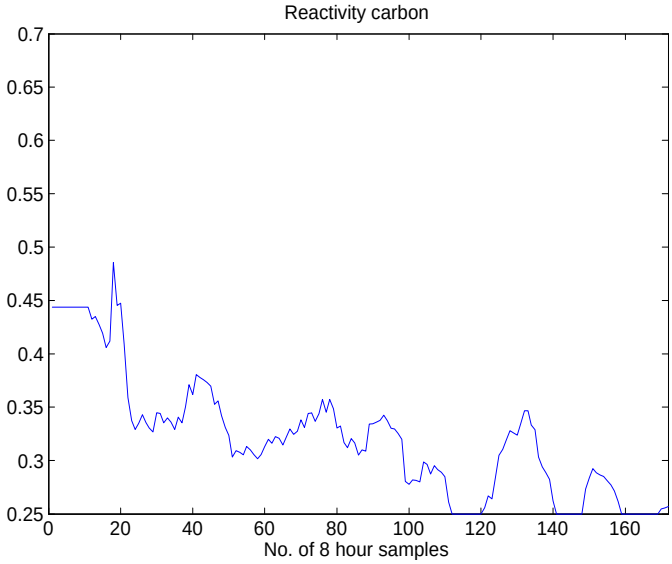

Figure 7.2: Estimated carbon reactivity, r_1.

The parameter estimates shown in figures 7.1-7.3 are the a posteriori corrected values using the actual and simulated output vectors within a window. The scaling of the y-axis in the figures is equal to the permissible variation range for each parameter.

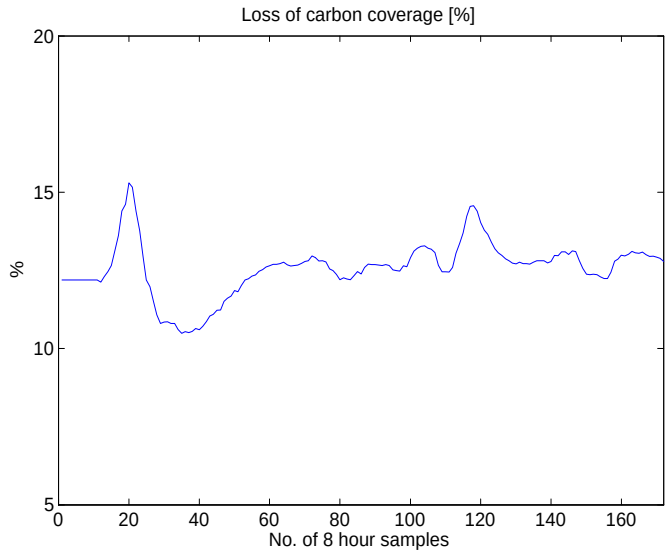

Figure 7.3: Estimated loss of carbon coverage, ΔC_C.

The first 10 samples of the time series in figures 7.1-7.3 show a flat region which corresponds to the first window. In this window there are also several discarded samples. From the second window and succeeding windows, the parameter estimates are registered at the time instant corresponding to the last sample in each window.

Simulated output vector

The sequence of parameter estimates can be reflected in a time series of estimated outputs in many different ways since the parameter values have been obtained as an averages value within a window. Also, the parameter estimate is filtered against previous parameter values incorporating information preceding the window as well. Possible choices of presentation are:

1. Use the whole output vector of the first window, and for the succeeding windows, pick the last, estimated output vector within each window.

2. Apply $\hat{\theta}_N$ to the first window, and then $\hat{\theta}_T$ at $T-1$ to T and simulate.
3. Apply $\hat{\theta}_N$ (estimate of the first window) to $T = 0$ to 1, and $\hat{\theta}_{N+1}$ to $T = 1$ to 2 etc., and simulate. The last parameter estimate in the sequence will then be applied to the N last samples.
4. Apply $\hat{\theta}_T$ at $T - N/2$, i.e. in the middle of a window and simulate
5. Apply $\hat{\theta}_T$ at $T - M$ within the window.

The first option implies collecting the last output vector resulting from the optimization within a window. The estimate $\hat{\theta}_T$ results from applying $\hat{\theta}_T$ within the whole window. The second option implies that the parameter estimate obtained at T is applied to future values. This obviously introduces a phase lag. The third option means that the parameter vector is applied at the beginning of each window. Depending on the window length and the degree of filtering of the estimate, this means that we apply an estimate which is also or mainly (if the degree of filtering is small) based on future data at this point. The fourth option implies applying the parameter estimate in the middle of the window. Without the filtering term in the optimization criterion, this option would probably give the smallest phase shift. However, if the filtering is relatively strong, then $T-N/2$, the "middle point" of the data sequence will lie further back. A pragmatic approach is to test different offsets within the window, and find a fixed offset that gives as little phase shift as possible.

In an on-line setting, the use of the parameter estimate will maybe be to simulate into the future to study the effect of different input values. However, the choices regarding where to apply the estimated parameter values also apply in the on-line setting.

Resulting output vector. Last output vector within each window

Here, the output data are presented according to choice 1 in the list on page 141, and shown in figures 7.4-7.6. The simulated outputs are marked as red curves and the actual data as blue curves. For the first window, the whole simulated output vector has been used. For succeeding windows, the last output vector in the window selected.

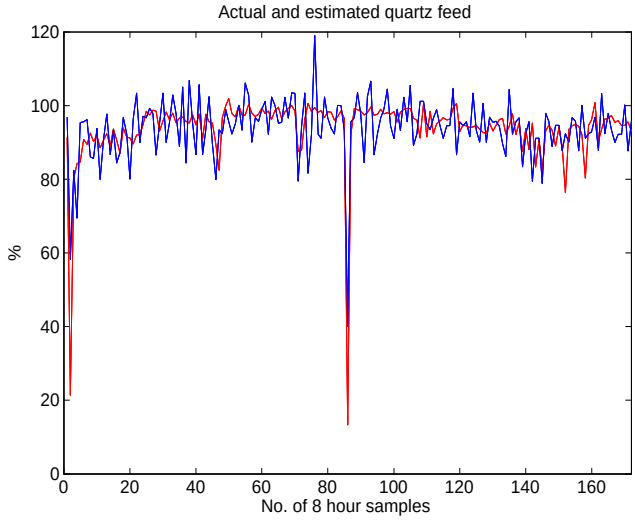

Figure 7.4: Scaled actual (blue) and estimated (red) quartz feed. The simulated outputs is the sequence of the last estimated outputs within each window.

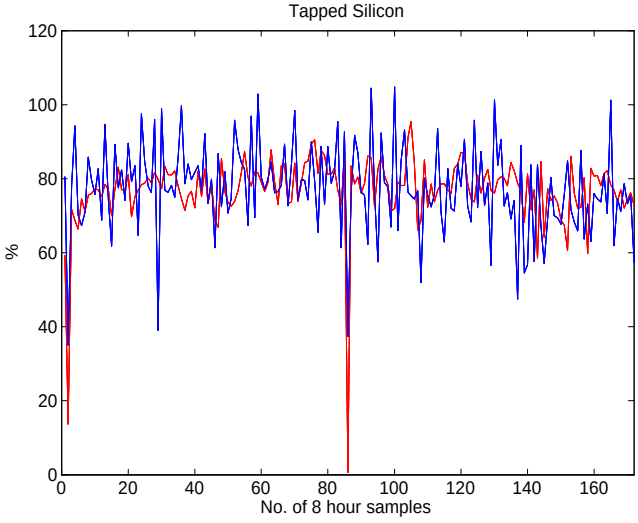

Figure 7.5: Scaled actual (blue) and estimated (red) taprate of silicon. The simulated outputs is the sequence of the last estimated outputs within each window.

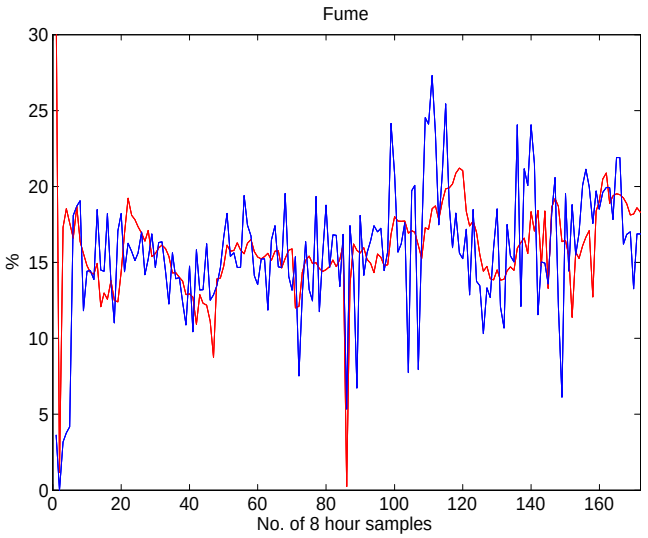

Figure 7.6: Scaled actual (blue) and estimated (red) fuming dust values. The simulated outputs is the sequence of the last estimated outputs within each window.

Resulting output vector by application of $\hat{\theta}_T$ at $T-3$

Since the parameter values have been obtained over a window, and the filtering term gives a small contribution to the total optimization criterion, the parameter values shown in figures 7.1-7.3 have been applied at $T-3$ and simulated. This corresponds to the fifth option in the list of choices on page 141. The resulting outputs are shown in figures 7.7-7.9.

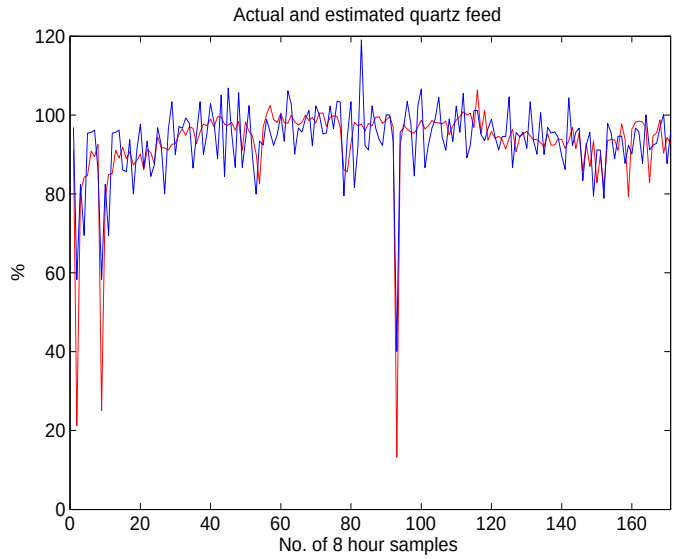

Figure 7.7: Actual and estimate quartz feed with parameter estimates applied at $T-3$.

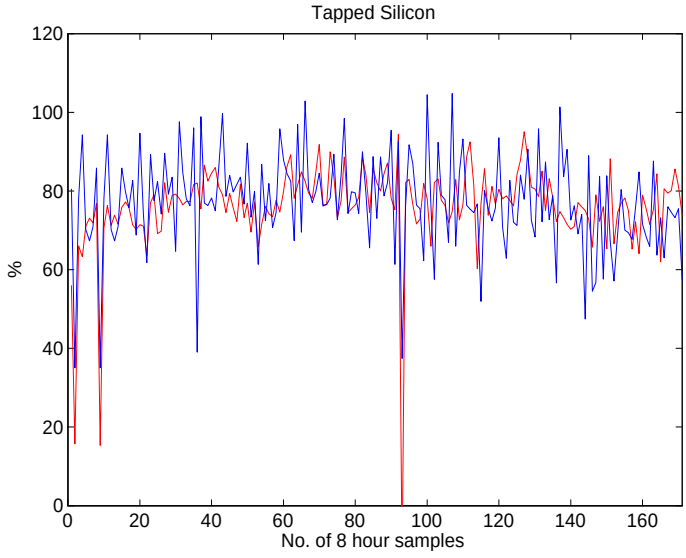

Figure 7.8: Actual and estimated tapped silicon with parameter estimates applied at $T-3$.

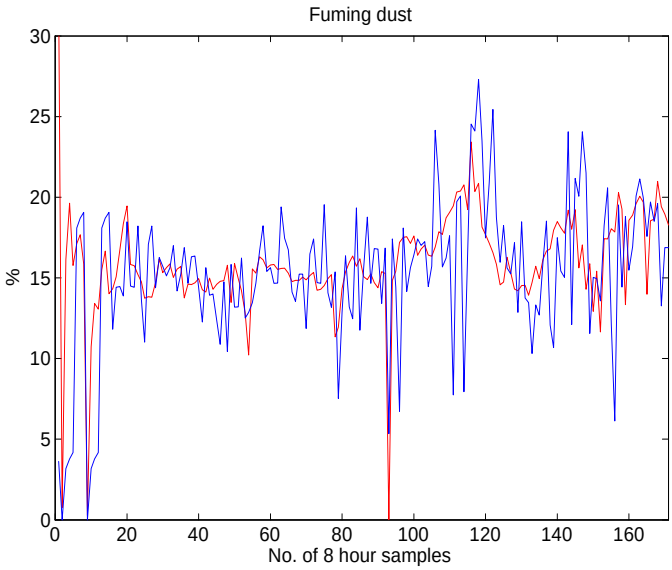

Figure 7.9: Actual and estimate fuming dust with parameter estimates applied at $T-3$.

7.1.3 Analysis and discussion of the results for 3 outputs

At the very start of the time series, there is a dip in the actual mass flow data, cf. figures 7.7-7.9 which indicates operational problems in the furnace. Since the estimation algorithm omits time samples where the fuming dust or tapped silicon is below a specified limit, the estimation is made on very few time instants in the beginning. One could claim that this is not a smart place to start the estimation. On the other hand, the fact that an estimate is found indicates that the estimation algorithm is relatively robust and can be started anywhere.

From figure 7.3 it can be observed that the loss of carbon coverage lies well within the upper and lower bounds set for this parameter. In figure 7.2, the estimate first lies within the limits, but from approximately sample 112 the estimate reaches the lower bound several times. The underlying carbon material composition data (not included here) show that the carbon composition has been relatively constant up until sample 85. At 85 a step change in the carbon composition occurs, and also at 109, 147, and 161. These changes may or may not have an effect on the resulting reactivity of

the carbon materials. However, the estimation results for the same period show a lowered estimated carbon reactivity, cf. figure 7.2.

The time period after approximately sample 110 also seems to have a rather large model errors, especially in the fuming dust production. The actual data show relatively large fluctuations which are only partly reflected by the model. This may be an indication of phenomena in the process which are not well reproduced by the model, for instance tapping problems. The model performs ideal tapping due to use of the tapping controller, while the process does not. Silicon can be "stored" in the model as SiC by lowering the reactivity. However, silicon carbide build-up in the furnace is often associated with tapping problems, therefore the the lowered carbon reactivity estimate may still be realistic. More fume can be generated in the model by increasing the loss of carbon coverage or adding more energy to the furnace. A lowered electrode resistance is another way to increase the fume.

7.1.4 Estimation results, 3 parameters, 2 outputs, furnace 1

Since individual measurements of fuming dust exist in some furnaces only, it is of interest to see how omitting the fuming dust output from the criterion affects the estimates.

Other than excluding the fuming dust outputs, the estimation has been made using the same input and output data as in section 7.1.2. The values of V are given in (7.2). The values $W(1,1)$ and $W(2,2)$ of W in (7.1) corresponding to the quartz feed rate and silicon taprate have been used. A larger window has been used to partly compensate for the lack of a fuming dust measurement, and the outputs from 17 time instants are included in the criterion.

The estimated parameters are shown in figures 7.10-7.12. Figure 7.12 shows very little variation in the carbon coverage loss, whereas figure 7.11 shows large variations in the carbon reactivity. Still, the parameters largely follow the same course as in figures 7.1-7.3. The actual and estimated outputs are shown in figures 7.13-7.15, according to method 1 in the list on page 141. The adaptation between actual and estimated quartz and tapped silicon data seems to be good.

The explanation of the "lack" of dynamic behavior in the carbon coverage loss and apparent increased sensitivity in the reactivity may of course be that the estimation criterion should have been retuned by assigning other values to V when the fuming dust measurement was removed.

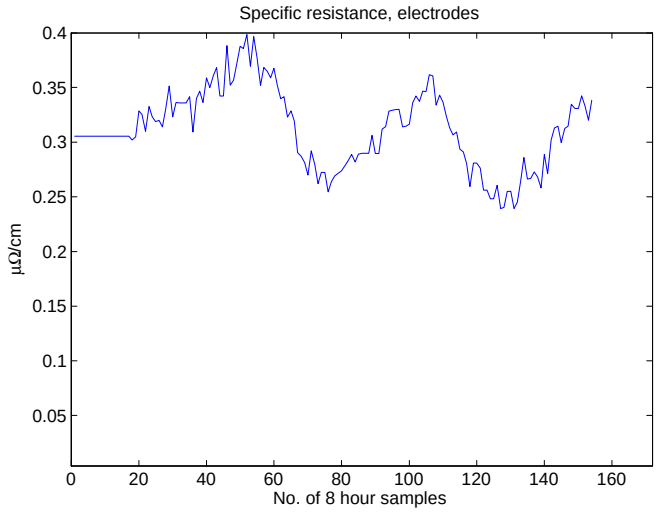

Figure 7.10: Specific electrode resistance, r_e, without fuming dust measurement.

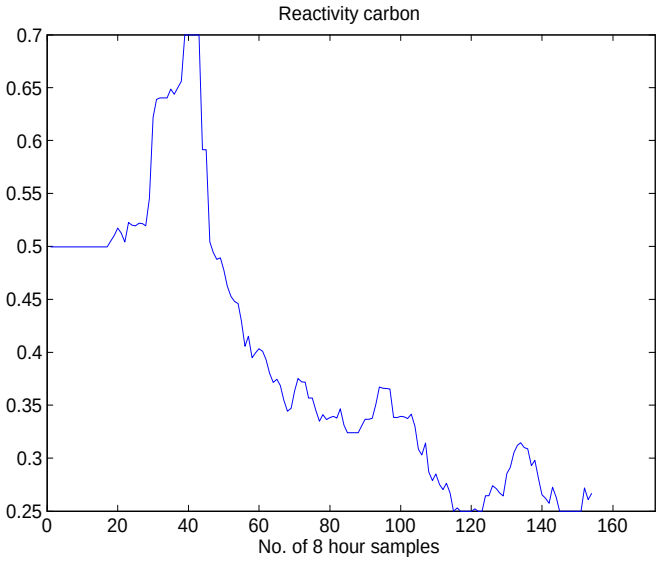

Figure 7.11: Carbon reactivity estimate, r_1, without fuming dust output.

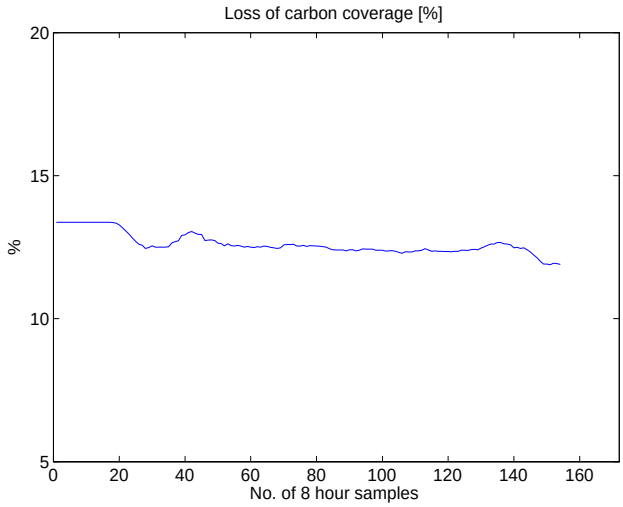

Figure 7.12: Carbon coverage loss estimate, ΔC_C, without fuming dust measurement in estimation criterion.

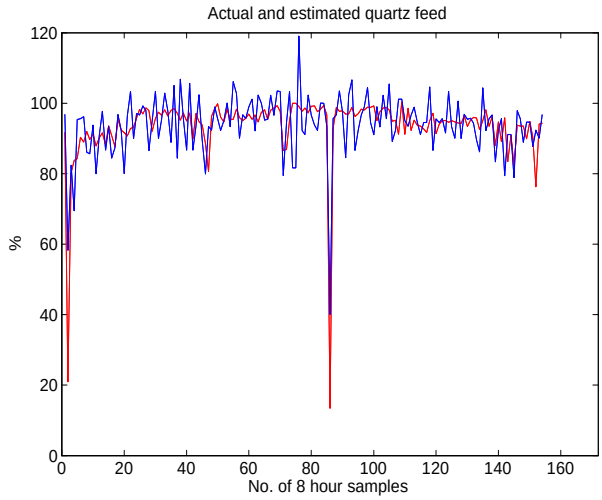

Figure 7.13: Actual (blue) and estimated (red) quartz fed to the furnace, without fuming dust in the estimation criterion.

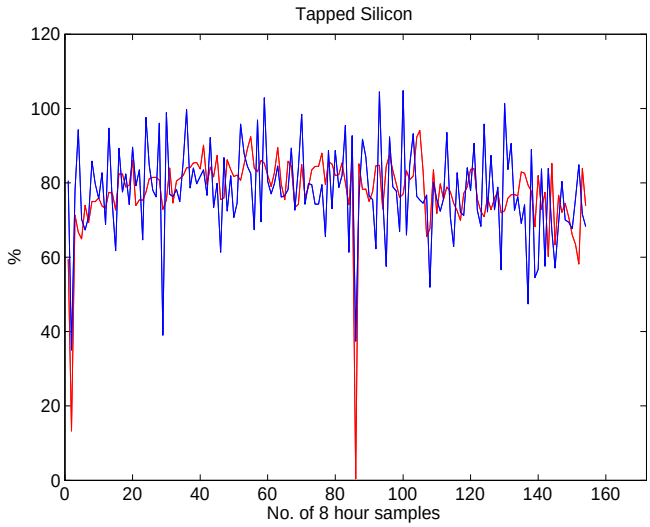

Figure 7.14: Actual (blue) and estimated (red) tapped silicon without the fuming dust measurement.

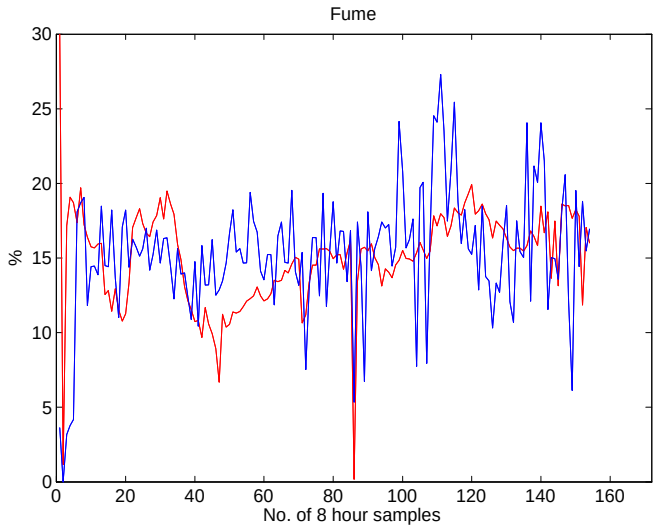

Figure 7.15: Actual (blue) and estimated (red) fuming dust measurement without fuming dust in the estimation criterion.

Another part of the explanation may be connected to the fact that the reactivity has a great impact on the dynamic behavior of the model taprate. Since the silicon taprate is the only mass flow output in the criterion, the estimation algorithm may attempt to "follow" any variations in the taprate due to for instance tapping problems. If there are tapping problems, which are not represented in the model, then the reactivity may be turned down to "store" silicon in the model. When the silicon again is tapped according to the actual data, the reactivity parameter is turned up again to make the model generate more silicon.

The exclusion of the fuming dust and increase of the time frame generally seem to have the effect that the sensitivity to the reactivity parameter is increased whereas the sensitivity to the carbon coverage loss is decreased. Some of the variation patterns can be recognized but generally it seems necessary to retune the estimation algorithm in order to increase the sensitivity in the carbon coverage loss estimate, and possibly decrease the sensitivity in the reactivity somewhat.

7.2 Estimation with biased mass "balance" data, furnace 2

Biased data give biased estimates, and the inclusion of this section may therefore seem puzzling. When mass balance data are biased, i.e. they do not balance, it may not be easy to determine which measurement or measurements are to blame. Analysis of the biased estimates and the corresponding output data may however give useful information when looking for the source(s) of mismatch.

In the silicon furnace, the input quartz feed should balance out with the output tapped silicon and fuming dust measurements, at least over a long period of time. A small bias may be due to some silicon lost as silicon oxide through the tap-hole, but larger offsets must have other explanations.

The data used here are taken from a furnace where there is a bias in the measured silicon balance. The bias is positive, meaning that more silicon in the form of quartz is put into the furnace than what is taken out in the form of silicon and fuming dust, according to the actual data. If the bias were to be blamed on the quartz weight, the values should be reduced with 2.7%. If the silicon weight were to be blamed for the bias, the silicon values should be increased with 3.2%. Correction of the fuming dust weight alone would require an increase of values with 21.2%.

Since it is suspected that the fuming dust weight can be the source

of the error, the entire time series of fuming dust measurements has been multiplied with the factor 1.212.

7.2.1 Parametrization and input data

A relatively short window of 12 samples has been used. The model error covariance has been set to

$$W = diag \left[\begin{array}{ccc} 2700^2 & 2350^2 & 551^2 \end{array} \right]$$

and the process noise covariance has been set to

$$V = diag \left[\begin{array}{ccc} 0.00990^2 & 0.0112^2 & 0.375^2 \end{array} \right]$$

The input data from this furnace are shown in figures 7.16, 7.17, and 7.18.

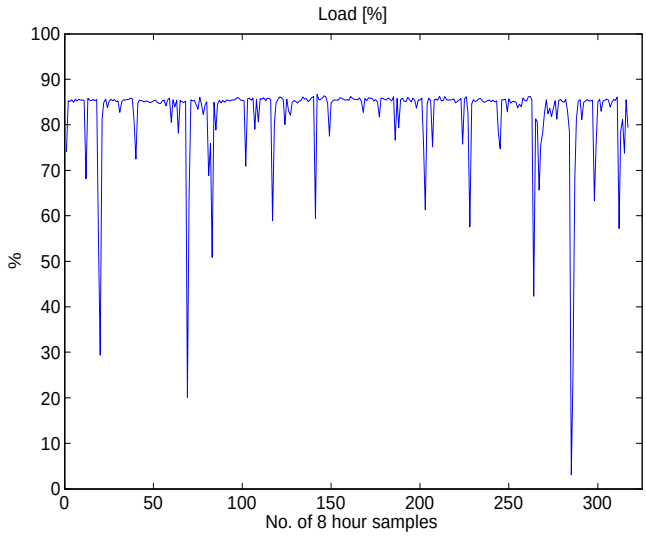

Figure 7.16: Supplied electric power (load), in percent of a nominal level.

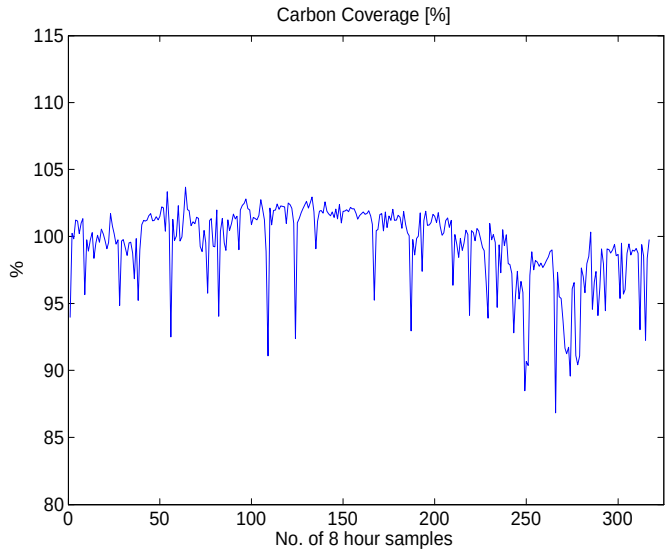

Figure 7.17: Carbon coverage.

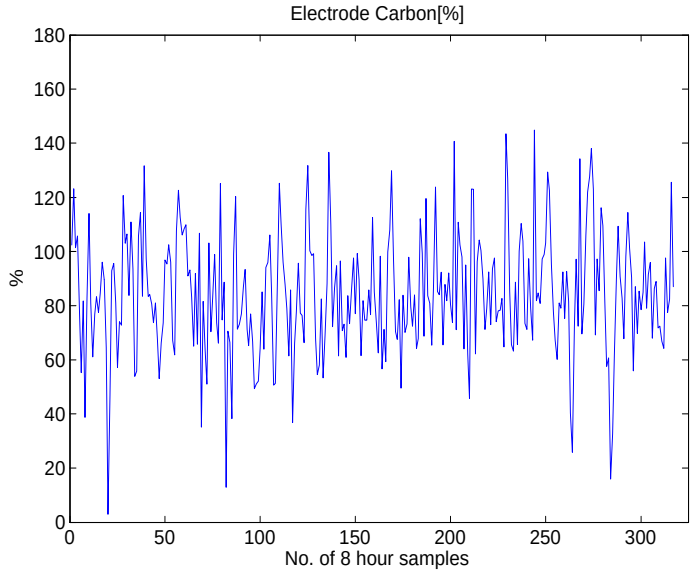

Figure 7.18: Electrode carbon in percent of a nominal level.

7.2.2 Estimation results, 3 parameters, 3 outputs, furnace 2

The resulting parameter estimates show a very low specific resistance in figure 7.19, a fairly high carbon coverage loss in figure 7.20 and a fairly normal carbon reactivity value in figure 7.21.

The estimated specific electrode resistance lies at its lower boundary value, especially in the first part of the time series. It is very unlikely that there is no resistance in the carbon electrodes, and the estimated value must therefore be biased. The carbon coverage loss is a parameter which can be used to alter the silicon yield in the model. The model seems to turn up the carbon coverage loss in order to be able to follow the measured fuming dust data. A loss of carbon coverage at almost 20% is definitely on the high side, even if this is within the permissible range shown by the range on the ordinate.

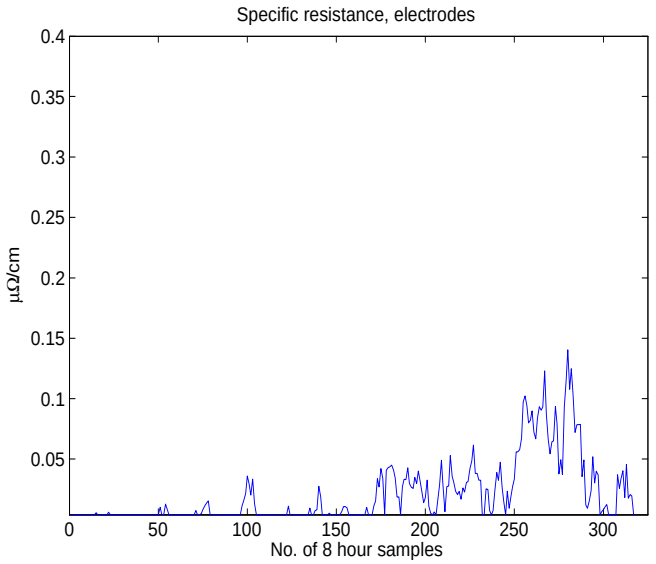

Figure 7.19: Estimated electrode specific resistance, r_e.

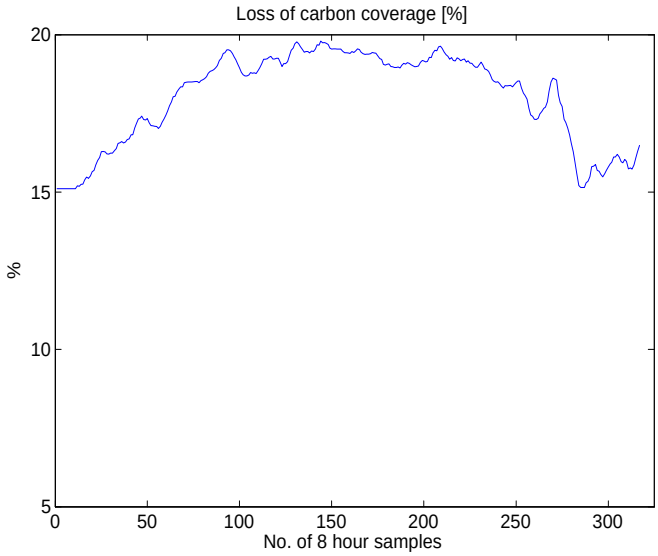

Figure 7.20: Estimated loss of carbon coverage, ΔC_C.

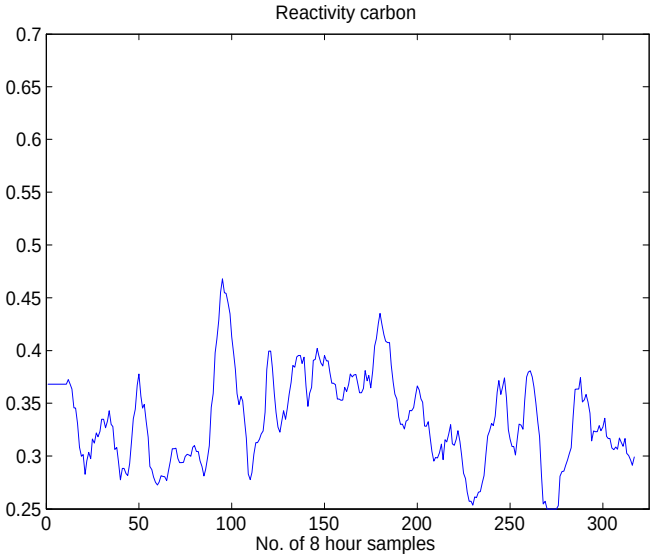

Figure 7.21: Estimated carbon reactivity, r_1

The very high loss of carbon coverage and the very low specific electrode resistance may be signs that the model is "struggling" to generate enough fuming dust, and may be a sign that the dust weight is not the cause of the entire bias.

Examination of the outputs (not included as figures) shows that there is a deviation between actual and estimated quartz with a mean value of 270kg/8 hours, between the actual fuming dust and the estimated fuming dust of 170kg/8 hours, and for silicon taprate 14 kg/8 hours. This means that the model is unable to keep up with the production level of the actual data, despite setting the electrode resistance to almost zero.

In chapter 4, figures 4.23 and 4.24 showed that the electric power level, or electric power largely determines the production level of the furnace. Since the production level in the actual outputs is higher than in the simulated outputs, and the estimated electrode resistance is set to almost zero by the model, the actual electric power level seems unable to match the actual quartz feed level. Therefore, either the electric power level or the quartz feed rate has a bias. Combining this with the knowledge about a positive bias in the actual material balance, i.e. more quartz is put in than what is taken out as silicon and fuming dust, one may suspect that some of the material balance bias is caused by the quartz weight, and not entirely by the fuming dust weight. In addition comes the abnormally high estimated loss of carbon coverage which indicates that it may not have been right to blame the whole mass balance bias on the fuming dust measurement.

The above is an example that the model and the parameter estimation can provide the redundancy in information needed to find the cause of a bias in mass balance data, in a situation when there is not sufficient redundancy in the measurements. The reasoning and conclusion is qualitative and not quantitative. The question whether this could have been done quantitatively will be left unanswered, but initially it seems difficult to estimate both the loss of electric power and an offset in the quartz weight. This is however not rejected as an idea, and could be a topic for further analysis. Tightening the permissible variation range for the parameters, and analyzing the resulting offset in the estimates could be one way to go. This would however seem just like a way of sharing the bias between the estimates and the measurements.

The work done in this section touches the field of data reconciliation. In the review article by Crowe (1996) data reconciliation is defined as the "...procedure of optimally adjusting measured data so that the adjusted values obey the conservation laws and other constraints". Models, estima-

tion and other statistical methods are important components in this field which deals with both systematic as well as spurious errors. Data reconciliation generally relies on some degree of redundancy in the process measurements and information provided by the model. Data reconciliation is a large field judging from the number of publications categorized under this term. Thomas F. Edgar and co-workers at the University of Texas have several contributions within this field. One example is the paper McBrayer, Soderstrom, Edgar & Young (1998).

The possible application of data reconciliation techniques to the silicon furnace data will not be pursued any further in this work.

Chapter 8

Conclusions

This dissertation has addressed the use of detailed simulation models for improved operation of process plants. Since there exists no general methodology, the problem area has been approached through a case. The main approaches have been

- analysis and extensive simulation studies to characterize and improve the understanding of process behavior.
- on-line estimation using the simulation model

Extensive simulation studies and analysis have been carried out in order to characterize the nonlinear and multivariate silicon furnace behavior. The work contributes to the understanding of silicon furnace behaviour with compact illustrations of some the relationships between the main inputs and outputs of the process. The analysis especially focuses on the furnace behavior around optimal production, and shows how changes in the dominant dynamics can be used to determine the margins to optimality. This result has been verified qualitatively by process experts.

The second approach has been to implement a scheme for on-line updating of the simulation model. The estimation scheme is based on a sliding window formulation. The criterion is a least squares criterion which also includes a term penalizing the deviation from the parameter estimate of the previous window. The weight of this term is updated using a Kalman filter covariance updating scheme. The estimate is found using a SQP search method. The gradient is obtained through perturbation of the parameter vector. Upper and lower bounds on the parameters are included. The implementation has been made ensuring flexibility with regards to window

length, covariance specifications, number of parameters to be estimated, and number of inputs and outputs to be used.

The estimation scheme has been applied to data sets from two different furnaces and performs well. Some of the estimated parameters express losses in the mass and energy inputs. The estimation therefore adapts the model's mass and energy balance to actual mass and energy balance data, but still lets the model fulfil its internal mass and energy balances.

When a general problem area is addressed through a case, it is always difficult to determine to which degree the properties of the case are general. Without being categorical, the following aspects identified in the present case are likely to be important when applying other existing, large, mechanistic simulation models for on-line estimation.

The model state vector is most likely not observable, and the model's parameter vector not identifiable. Process and model insight is required in order to determine how the model can be applied for on-line estimation, and for identifying candidates for estimation. The candidate set may be large, and a only a subset may be included in an estimation algorithm. Efficient methods are needed for ranking and reduction of a candidate variable set. The ranking method based on successive orthogonalization of the sensitivity derivative is an efficient method for analysis and reduction of the parameter set.

If the mass and energy balances of the model are to be corrected vs. actual process data, it is important to consider how the model is to fulfil its mass and energy balance equations during estimation to ensure sound model behavior.

The software architecture and interfaces of the model may have a significant effect on the selection of possible estimation methods, and the possibilities for calculating gradients, using iterative techniques should be assessed early in the design process.

Domain knowledge is also required in evaluating the outcome of the estimation. Since the estimated quantities in a mechanistic model can be interpreted physically, the estimation results will always be discussed and second-guessed by domain experts.

The possibilities for physical interpretation of the estimation results also provides a means for integration of individual process measurements and interpretation of the process state at a more complete level.

The design steps proposed on page 9 for bringing the Simod model on-line, are therefore likely to be relevant for bringing other large, mechanistic models on-line as well.

Bibliography

Allgöwer, F., Badgwell, T. A., Qin, J. S., Rawlings, J. B. & Wright, S. J. (1999), *Advances in Control. Highlights of ECC 99*, Springer Verlag, chapter 12. Nonlinear predictive control and moving horizon estimation-an introductory overview, pp. 391–449.

Belsley, D. A. (1991), *Conditioning Diagnostics. Collinearity and Weak Data in Regression*, Wiley.

Berntsen, H. E. (1977), Modelling, Parameter Estimation and Identifiability in Electrocadiology, PhD thesis, Norwegian Institute of Technology.

Bierman, G. J. (1977), *Factorization Methods for Discrete Sequential Estimation*, Acedemic Press, New York.

Brown, R. G. & Hwang, P. Y. C. (1997), *Introduction to Random Signals and Applied Kalman Filtering*, Wiley and Sons.

Brun, R., Reichert, P. & Kunsch, H. R. (2000), Practical identifiability analysis of large environmental simulation models, Technical report, EAWAG, ETH.

Bryson, A. E. J. & Ho, Y.-C. (1975), *Applied Optimal Control*, Hemisphere Publishing Corporation, Washington D.C.

Choi, B. Y. & Bien, Z. (1989), 'Sliding-window weighted recursive least-squares method for parameter estimation', *Electronic letters* **25**(20), 1381–1382.

Crowe, C. M. (1996), 'Data reconciliation - progress and challenges', *Journal of Process Control* **6**(2/3), 89–98.

Davis, W. J. (1998), *On-Line Simulation: Need and Evolving Research Requirements*, Wiley, chapter 13, pp. 465–516.

Diez, L. I., Cortes, C. & Campo, A. (2005), 'Modelling of pulverized coal boilers: Review and validation of on-line simulation techniques', *Applied Thermal Engineering* **25**, 1516–1533.

Elkem (2002), 'Elkem annual report 2002'.

Elkem (2004*a*), 'Elkem årsrapport 2003'.

Elkem (2004*b*), 'Http://www.silicon.elkem.com/'.

Forsell, U. & Ljung, L. (1999), 'Closed-loop identification revisited', *Automatica* **35**, 1215–1241.

Foss, B. A. & Wasbø, S. O. (2001), 'An integration scheme for stiff solid-gas reactor models', *Computer Methods in Applied Mechanics and Engineering* **190/45**, 6009–6021.

Foss, B., Halfdanarson, J. & Wasbø, S. (2000), Dynamisk si-modell - simon v.1.50, Technical Report Confidential report no. STF72 F00307, SINTEF.

Golub, G. H. & Van Loan, C. F. (1996), *Matrix Computations*, 3rd edn, John Hopkins University Press.

Graham, C. G. & Sin, K. S. (1984), *Adaptive Filtering, Prediction and Control*, Prentice-Hall.

Hauksdottir, A. S., Gestsson, A. & Vesteinsson, A. (2002), 'Current control of a three-phase submerged arc ferrosilicon furnace', *Control Engineering Practice* **10**, 457–463.

Hyllseth, M. & Cameron, D. (2003), Operator training and operator support using multiphase pipeline models and dynamic process simulation:subsea production and on-shore processing, *in* A. Kraslawski & I. Turunen, eds, 'European Symposium on Computer Aided Process Engineering - 13', Elsevier Sciences, pp. 425–430.

Jacobsen, E. W. (1994), 'Dynamics of systems with steady state input multiplicity', *Presented at AiChE Annual Meeting, San Fransisco* .

Jacobsen, E. W. (1999), 'On the dynamics of integrated plants - non-minimum phase behavior', *Journal of Process Control* **9**, 439–451.

Jiang, J. & Zhang, Y. (2004), 'A revisit to block and recursive least squares for parameter estimation', *Computers and Electrical Engineering* **30**, 403–416.

Johansen, T. A. (1997), 'On tikhonov regularization, bias and variance in nonlinear system identification', *Automatica* **33**(3), 441–446.

Johansen, T. A. (1998), 'Constrained and regularized system identification', *Modelling, Identification and Control* **19**(2), 109–116.

Jumar, U. & Tschepetzki, R. (2001), 'Implementation of a WWTP operation support tool based on on-line simulation', *Preprints of 1st IWA conference on instrumentation, control and automation* **2**, 711–718.

Kailath, T., Sayed, A. H. & Hassibi, B. (2000), *Linear Estimation*, Prentice Hall.

Kalman, R. E. (1960), 'A new approach to linear filtering and prediction problems', *Transactions of the ASME-Journal of Basic Engineering* **82**(Series D), 35–45.

Kanjilal, P. P., Ballav, P. & Saha, G. (1995), 'Fast successive selection of variables in linear models using modified QR factorization', *Electronics letters* **31**(14), 1204–1205.

Kanjilal, P. P., Saha, G. & Koickal, T. J. (1999), 'On robust nonlinear modeling of a complex process with large number of inputs using m-QRcp factorization and statistic', *IEEE Transactions on systems, man, and cybernetics* **29**, 1–12.

Kolbeinsen, L., Lindstad, T., Tveit, H., Bruno, M. & Nygaard, L. (1995), 'Energy recovery in the norwegian ferro alloy industry', *INFACON 7, published by the Norwegian Ferroalloy Reseach Organization* pp. 165–177.

Kuhlmann, A. & Bogle, D. (1997), 'Study on nonminimum phase behavious and optimal operation', *Computers Chem. Engineering* **21**, 397–402.

Levine, W. S. (1996), *The Control Handbook*, CRC Press and IEEE Press.

Lewis, F. L. (1986), *Optimal Estimation*, Wiley and Sons.

Li, R., Henson, M. A. & Kurtz, M. J. (2004), 'Selection of model parameters for off-line parameter estimation', *IEEE Trans. on Control Sys. Tech.* **12**(3), 402–412.

Ljung, L. (1999), *System Identification. Theory for the User*, Prentice-Hall Inc.

Ljungquist, D. (2002), Observability of simod variables, Technical report, Elkem.

Luenberger, D. G. (1966), 'Observers for multivariable systems', *IEEE Transactions on Automatic Control* **11**, 190–197.

Luenberger, D. G. (1971), 'An introduction to observers', *IEEE Transactions on Automatic Control* **16**(6), 596–602.

Lund, B. F., Berntsen, H. E. & Foss, B. A. (2005), Methods for parameter ranking in nonlinear, mechanistic models. IFAC World Congress, Prague, 2005.

Lund, B. F. & Foss, B. A. (2005), Parameter ranking by orthogonalization - applied to nonlinear mechanistic models. Submitted as Automatica brief paper.

Lund, B. F., Foss, B. A. & Løvåsen, K. R. (2005), Analysis and characterization of complex reactor behaviour. A case study. Submitted to Journal of Process Control.

Lund, B. F., Foss, B. A., Løvåsen, K. R. & Ydstie, B. E. (2004*a*), 'Sensitivity analysis of a dynamic model for submerged arc silicon furnaces', *Proceedings of the Infacon X Congress* .

Lund, B. F., Foss, B. A., Løvåsen, K. R. & Ydstie, B. E. (2004*b*), System analysis of complex reactor behavior - a case study, *in* 'Proceedings of DYCOPS 7', IFAC.

Marino, R. & Tomei, P. (1995), *Nonlinear Control Design*, Prentice Hall, Europe.

McBrayer, K. F., Soderstrom, T. A., Edgar, T. F. & Young, R. E. (1998), 'The application of nonlinear dynamic data reconsiliation to plant data', *Computers in Chemical Engineering* **22**, 1907–1911.

Morud, J. & Skogestad, S. (1996), 'Dynamic behaviour of integrated plants', *Journal of Process Control* **6**(2/3), 145–156.

Myrhaug, E. H. (2003), Non-Fossil Reduction Materials in the Silicon Process - Properties and Behaviour, PhD thesis, NTNU.

Nocedal, J. & Wright, S. J. (1999), *Numerical Optimization*, Springer.

Rao, C. V. (2000), Moving Horizon Strategies for the Constrained Monitoring and Control of Nonlinear Discrete Systems, PhD thesis, University of Wisconsin-Madison.

Rao, C. V., Rawlings, J. B. & Lee, J. (2001), 'Constrained linear state estimation - a moving horizon approach', *Automatica* **37**, 1619–1628.

Rao, C. V., Rawlings, J. B. & Mayne, D. Q. (2003), 'Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations', *IEEE Transactions on Automatic Control* **48**(2), 246–258.

Ruszkowski, M. & Ydstie, B. E. (2002), 'Stabilization of reaction-convective systems using inventory control', *Proc. of the American Control Conference 2002* pp. 1028–1033.

Schei, A., Tuset, J. K. & Tveit, H. (1998), *Production of High Silicon Alloys*, Tapir forlag, Trondheim, Norway.

Sjöberg, J., Ljung, L. & McKelvey, T. (1993), 'On the use of regularization in system identification', *Proc. 12th IFAC World Congress* pp. 381–386,.

Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Glorennec, P.-Y., Hjalmarsson, H. & Juditsky, A. (1995), 'Nonlinear black-box modeling in system identification : A unified overview', *Automatica* **31**(12), 1691–1724.

Söderström, T. & Stoica, P. (1989), *System Identification*, Prentice Hall.

Strang, G. (1988), *Linear Algebra and its Applications*, 3rd edn, Saunders HBJC Publishers.

Tikhonov, A. N. & Arsenin, V. Y. (1977), *Solutions to Ill-Posed Problems*, Wiley.

Valderhaug, A. M. (1992), Modelling and Control of Submerged-Arc Ferrosilicon Furnaces, PhD thesis, Norwegian Institute of Technology.

Weijers, S. R. & Vanrolleghem, P. A. (1997), 'Procedure for selecting best identifiable parameters in calibrating activated slugde model no. 1 to full-scale plant data', *Water Science and Technology* **36**(5), 69–79.