# Operating Regime based Process Modeling and Identification

*Dr. Ing. Thesis*

**Tor Arne Johansen**

Department of Engineering Cybernetics
The Norwegian Institute of Technology - University of Trondheim

1994

Report 94-109-W
Department of Engineering Cybernetics
Norwegian Institute of Technology
N-7034 Trondheim, Norway

# Preface

This thesis is submitted in partial fulfilment of the requirements for the degree Dr. Ing. at the Norwegian Institute of Technology - NTH, Trondheim.

This work was supported by The Research Council of Norway under doctoral scholarship grant no. ST. 10.12.221718 and supervised by Prof. Bjarne Foss at the Department of Engineering Cybernetics, who has been of great inspiration and support. Thanks.

Moreover, I would like to thank Prof. Petros Ioannou at the University of Southern California for hosting my six month visit at USC. My interactions with him and his students improved my mathematical precision and resulted in some adaptive control results that are partially reported in this thesis. Two chapters in this thesis are based on manuscripts that are co-authored with Aage V. Sørensen at the Norwegian Institute of Technology, and Dr. Erik Weyer at The University of Queensland, Brisbane, Australia. Several parts of this thesis has benefited from numerous discussions with Dr. Rod Murray-Smith at Daimler-Benz AG, Berlin and Dr. Tom Kavli at SINTEF-SI. Among others, Mike Thompson at Massachusetts Institute of Technology, Dr. Magne Hillestad at STATOIL, Prof. Don Wiberg at the University of California in Los Angeles, and Olav Slupphaug at the Norwegian Institute of Technology have contributed with comments on parts of this work.

Finally, I would like to thank Trude for her support.

Trondheim, November 24, 1994

Tor Arne Johansen

# Summary


The development of mathematical models is a major bottleneck for the application of advanced model-based techniques for control, optimization, scheduling, automatic fault detection and diagnosis in the process industries. Hence, there is a potential for improved product quality and profitability, as well as improved production flexibility and safety if the cost of model development can be reduced.

In this thesis we address some aspects of non-linear modeling and identification using a combination of empirical process data and prior knowledge.

The major part of this thesis is concerned with a modeling framework based on an operating regime decomposition of the system's operating range. Within each operating regime, the system is modeled with a simple local model. The local models are patched together using a smooth interpolation technique. This framework supports the development of transparent semi-empirical or semi-mechanistic models, and is flexible with respect to the prior knowledge and empirical data required. We argue that the cost of model development can be low, yet the quality of the model can be high, through the application of this modeling framework, in some cases. These cases are characterized by a limited amount of process knowledge, and the availability of a reasonable amount of process data. Identification of model structure and parameters on the basis of process data in this framework is discussed in detail. The properties of the modeling framework is illustred with some semi-realistic examples, both simulated and experimental. In addition, the applicability of the modeling framework for model based control is investigated both through simulation examples and analysis.

A minor part of this thesis is an optimization formulation of the modeling and identification problem, where the idea is to minimize a criterion that penalizes mismatch between model behavior and empirical data, and inconsistency with the prior knowledge. The motivation is that in some cases, the application of prior knowledge may be simpler and more transparent than through the direct specification of a parameterized model structure, which is the traditional approach.

In addition, this thesis contains some theoretical results of more general interest. Some of these results are related to properties of some model structure identification criteria, while others are related to the stability and robustness of adaptive control loops based on non-linear models.

# Contents

# Chapter 1

# Introduction

In this thesis we develop a modeling framework that supports the development of non-linear dynamic models that are required to be valid under a wide range of operating conditions. Such models may be useful for e.g. analysis, optimization, control, or simulation of complex processes, e.g. chemical, biological, or metallurgical processes. The goal of this doctoral study has been a modeling framework that is flexible with respect to utilization of the prior knowledge and empirical data available.

## 1.1 Modeling Paradigms

In science and engineering, there are two fundamentally different philosophies that form the basis of modeling, namely the mechanistic and empirical approaches. A mechanistic model structure is developed on the basis of a detailed understanding of the generic underlying mechanisms, or laws, that governs the system behavior. Its parameters may be identified using empirical data. An empirical model, on the other hand, is derived on the basis of the specific observed behavior of the system. Its structure is often a generic black box that cannot be directly interpreted in terms of the system mechanisms. However, it may also be developed on the basis of empirical knowledge, including measured process data, and the experience of process operators and engineers. Often, neither of these two approaches are attractive. If the system mechanisms are only partially understood, the development of a mechanistic model structure may not be feasible. This may, however, also be the case for an empirical model, because of a lack of process data and difficulties to incorporate the available system knowledge in this approach. This is because engineering knowledge is incompatible with many empirical model representations, like black boxes. In practise, most models are based on an unbalanced combination of mechanistic and empirical knowledge. For example, in a mainly mechanistic approach certain aspects of the system that are not sufficiently well understood may be described by empirical correlations in the model. Also, in a dominantly empirical approach, some mechanistic understanding is often useful to make certain structural choices, like the model order, or non-linearities. Moreover, at early stages in the model development, an empirical model will often be useful as a starting point for gaining more knowledge and eventually designing a mechanistic model.

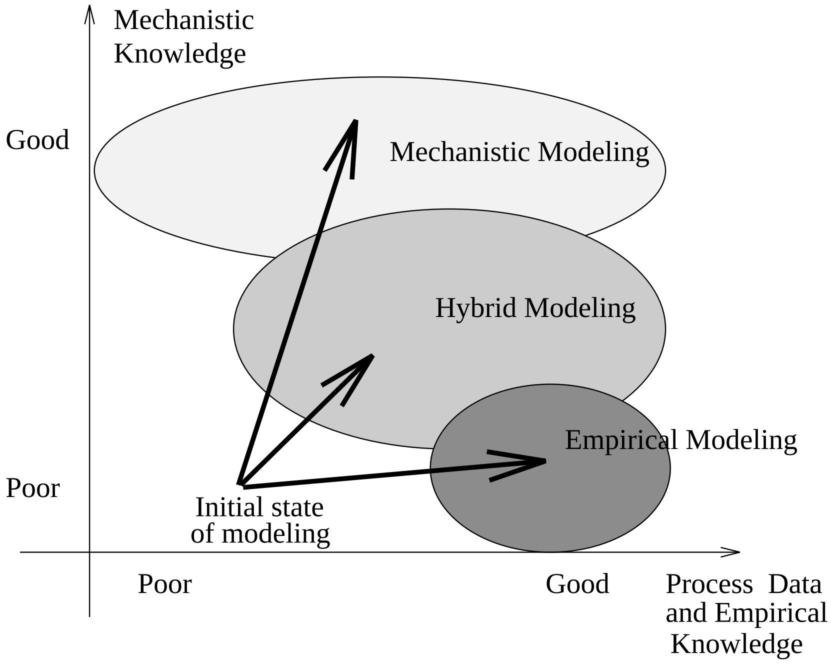


Figure 1.1: The figure illustrates different modeling paradigms.

Different modeling paradigms are visualized in Fig. 1.1, projected onto the "mechanistic" and "empirical" axes. A typical situation is illustrated, where the initial state of the model development is characterized by a lack of both empirical and mechanistic knowledge. With the empirical approach, one will collect more data, and end up in a state with more data and perhaps some improved mechanistic knowledge. With the mechanistic approach, on the other hand, one will end up with significantly improved mechanistic knowledge and perhaps some more data. The "in between" region and path corresponding to hybrid models (semi-empirical models, semi-mechanistic models, or models with a balanced utilization of empirical and mechanistic knowledge) is often approached in an ad. hoc. manner, or avoided. A major reason for this may be that powerful frameworks and software tools for such problems are lacking. The main objective of this thesis is to develop and study a framework that we believe may be useful for solving such hybrid modeling problems. It is our assumption that resources in many cases can be saved by choosing the hybrid approach. This will be justified in this thesis. We also think the framework is useful during a transition phase through the "in between" region towards a mechanistic model, if that is the ultimate goal.

In general, mechanistic models contribute to a larger extent to our scientific understanding of the system under study, and is the ultimate goal of all scientific studies. For engineering purposes, on the other hand, a deep understanding of the system mechanisms is not always required, and a model that contains a significant amount of empiricism can sometimes be accepted. Both the mechanistic and empirical modeling paradigms have serious drawbacks. The mechanistic approach is often resource demanding and sometimes infeasible if the system is not well understood. Empirical modeling, on the other hand, suffers from a perhaps even more serious drawback, namely that it is based on the principle of induction, which suggests that it is possible to generalize from a sufficiently large number of consistent observations. It is widely accepted that induction has serious limitations, since a finite amount of observations is not sufficient to cover the usually infinite number of conditions the model must be valid under. This, and other aspects, makes the empirical approach highly criticized as a scientific approach, e.g. (Chalmers 1982). However, for engineering purposes, the model is usually required to be valid only for a specific system that is operating in a significantly smaller operating range compared to scientific laws and models that are required to be valid for all possible systems and under a wide range of conditions. Often, one may be able to establish some continuity or smoothness properties of the system, so that limited interpolation and extrapolation of the empirical model can be allowed. It seems fair to state that mechanistic models are often reliable but expensive, while empirical models are often unreliable but inexpensive, where the reliability is with respect to extrapolation, and the expenses are development costs. Moreover, an empirical model is often more accurate than a mechanistic model when the system is operating under similar conditions as when the empirical data was collected. A hybrid model, based on a combination of both empirical and mechanistic knowledge and data, may be inexpensive, accurate, and reliable simultaneously, and should therefore be of potential interest for practical applications. It must be emphasized that these properties of hybrid models are by no means inherent, but require a minimum of prior knowledge, empirical data, and engineering skills.

In general, it is desirable that the model representation is transparent and compatible with the engineers understanding of the system, since this simplifies interpretation, analysis, and application of the model, and allows incorporation of prior knowledge.

The intended application of a model, together with the available knowledge and data, will restrict its required and attainable accuracy. For example, a model used for process design or optimization may need to be more accurate than a model used for control design. On the other hand, such a model may only need to be a steady-state model, while a control design will in general be based on a dynamic model that should be most accurate around the desired crossover frequency. Also, a model used for control during startup, shutdown or exceptional operating conditions may only need to roughly predict the major causal influences between variables, since this may be sufficient to bring the system quickly and reliably into the desired operating mode. There are often different requirements regarding accuracy of the model under different operating conditions.

The intended application of a model will also impose some restrictions on its structure and complexity. If a particular control algorithm is chosen a priori, it may be a restriction that the model must be a linear transfer-function model, for example. Furthermore, a model used for analysis need to be in a mathematically tractable form, contrary to a model used for numerical calculations only. Obviously, since physical states and parameters are not part of a black-box model, the use of a black-box model for estimation of such unmeasured physical variables is not straightforward. Mechanistic models are generally applicable for a wider class of problems than empirical models. The structure of hybrid models will in some cases restrict the applicability to the same as an empirical model, while in other cases, the form will be closer to a mechanistic model.

Knowledge is not static, but will usually improve during plant operation, both as historic process data, and improved understanding of the plant during operation. It is therefore important that the model representation supports incremental modeling in the sense that falsified or incomplete knowledge and models can be substituted with improved knowledge and models in a simple way.

## 1.2 Modeling and Identification

Modeling and identification deals with the mapping of prior knowledge and empirical data to a model. More precisely, we formulate the modeling and identification problems as follows:

**The modeling problem:** Given a set of empirical data and some a priori knowledge about the system, find a set of models that is likely to contain a model that describes the desired aspects of the system adequately.

**The identification problem:** Given a model set, a set of empirical data, and an objective criterion that measures the mismatch between the prediction of a given model and the empirical data, find a model within the model set that minimizes the criterion. It is common practise to decompose the identification problem into at least two subproblems: Model structure identification, and parameter identification.

The modeling process is typically iterative and can be formalized as in Fig. 1.2:

1. The first step is the formalization and analysis of the prior knowledge, design of experiments, collecting and analyzing data. Sometimes the data or prior knowledge are given, and parts of this task may be omitted. This step is perhaps the most important in the whole modeling process, since the success of the modeling and identification will strongly depend on the information in the data, and the amount and reliability of the prior knowledge.

2. The choice of the model set is what we have defined as *modeling*, and is usually based on priori knowledge and/or some preliminary analysis of the data. With good a priori knowledge, the model set will typically contain mechanistic models. On the other hand, black box model sets are useful since they are generic, i.e. there is a reasonable chance that there is a model in the black-box model set that can describe any system within a large class of systems. Black-box model sets are typically very large, but their size can

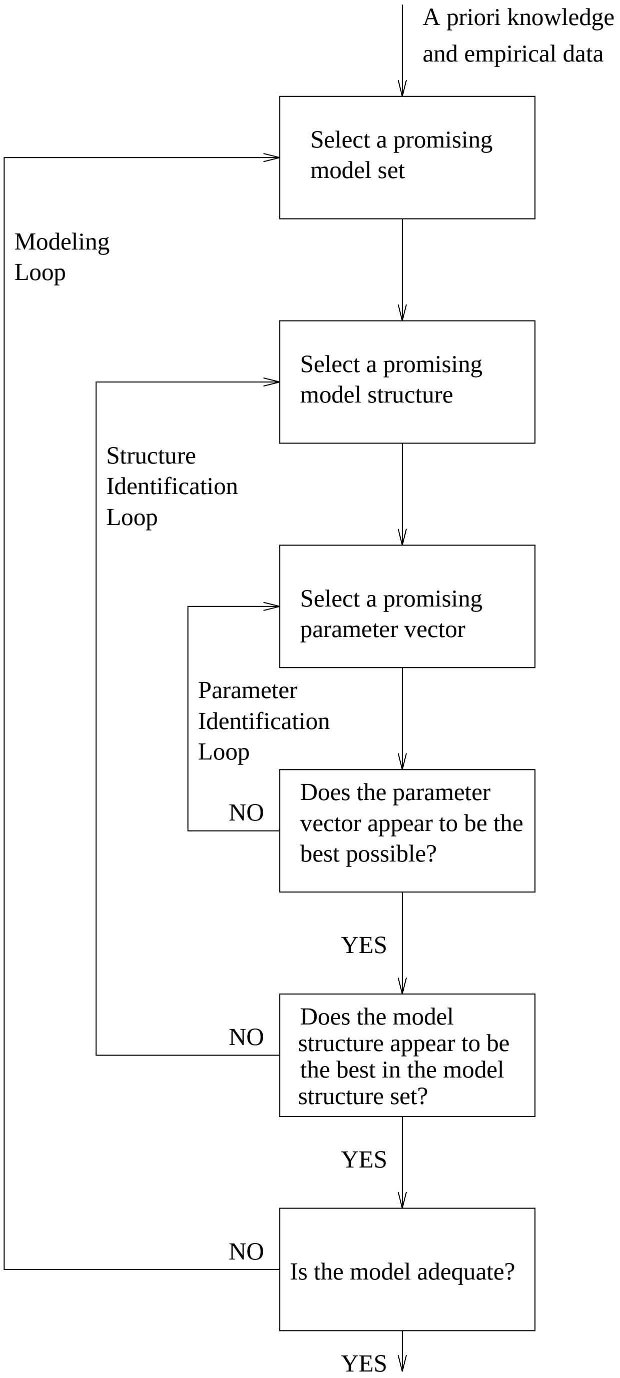


Figure 1.2: The iterative process of modeling and identification.

be reduced considerably by the application of some prior knowledge. This is important, since while it is obviously important that the model set is rich, it should at the same time be as restricted as possible to reduce the cost of searching through the model set. Clearly, one may put a lot of resources into the modeling part to reduce the complexity of the identification, and vice versa. In practice, time, knowledge, operational, and economic constraints will influence the choice of model set.

3. The model set is typically large, and parameterized by elements of function spaces, discrete spaces, and Euclidean spaces. *Model structure identification* deals with search and optimization in function spaces and discrete spaces. It is typically an iterative process, and the model structure from the previous step in the iteration is usually used as a basis for the selection of a more promising model structure, either refined or simplified. Often, the structure identification problem is a combinatorial one. Hence, good heuristics are important to avoid too many iterations in this loop.

4. In this thesis, *parameter identification* is defined as estimation of parameters belonging to an Euclidean space. This problem is much studied in the literature.

5. After the apparently best model in the model set has been found, it must be *validated*. A validation test may include prediction tests using a separate validation data set. If the model does not pass the validation test, a more general model set may be tried. Alternatively, the prior knowledge must be revised and improved, or more informative data must be collected.

Observe that each steps in this procedure can be done either manually, or automatically by a computer. The inner loops performs well defined tasks based on quantitative information and data, and are most suited for computer implementation, while the outer loops are based on qualitative knowledge, and therefore more difficult to formalize and implement on a computer.

Notice that the procedure outlined above is not intended to capture all facets of practical modeling and identification work. It is only meant to serve as a framework for later use in this thesis.

## 1.3 An Operating Regime based Modeling Framework

Any model will have a limited range of validity. This may be restricted by the modeling assumptions for a mechanistic model, or by the experimental conditions under which the data was logged for an empirical model. To emphasize this, a model that has a range of validity less than the desired range of validity will be called a *local model*, as opposed to a *global model* that is valid in the full range of operation. In this thesis we will be concerned with a modeling framework that is based on combining a number of local models, where it is of particular importance to describe the region in which each local model is valid. We call such a region an *operating regime*.

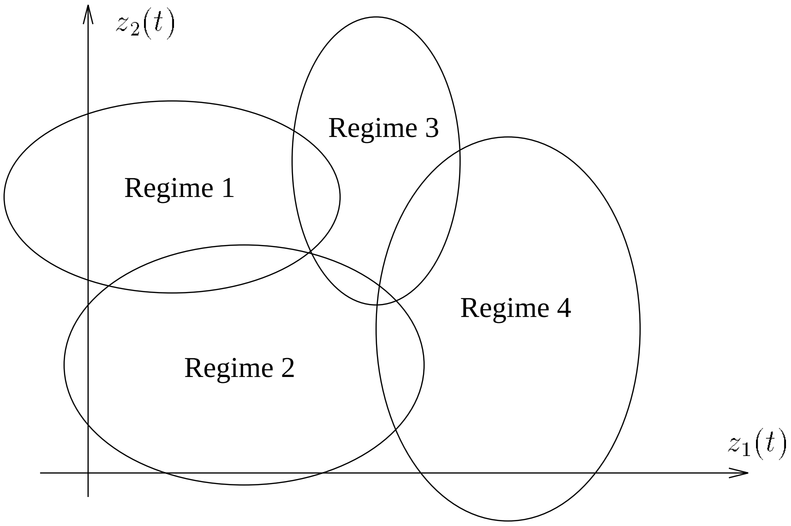


Figure 1.3: The set of two-dimensional operating points is decomposed into four regimes. The vector $z(t) = (z_1(t), z_2(t))$ is the operating point.

The framework can be conceptually illustrated as in Fig. 1.3. The system's full range of operation is completely covered by a number of possibly overlapping operating regimes. In each operating regime the system is modeled by a local model, and the local models can be combined into a global model using an interpolation technique. One motivation behind this framework is that global modeling is complicated because one will need to describe the interactions between a large number of phenomena that appear globally. Local modeling, on the other hand, may be considerably simpler because locally there may be a smaller number of phenomena that are relevant, and their interactions are simpler.

Already at this stage, we can to some extent justify the following attractive features of the framework:

- The decomposition of the modeling problem using operating regimes is a direct application of the divide-and-conquer strategy, which is a general approach to problem solving.
- Both the concept of operating regimes, and the model structure, are easy to understand. This is important, since the model structure can be interpreted both qualitatively in terms of the operating regimes, but also quantitatively in terms of the individual local models. Moreover, the operating regime concept offers a framework for interdisciplinary work because of its close relation to common engineering thinking.
- The framework is independent of the underlying model representation, in the sense that it works for input/output or state-space models, distributed or lumped models, and discrete-time, continuous-time, static, or discrete-event models.

- An immediate implication of the points above is that the framework forms a basis for hybrid modeling. For example, in some regimes the system can be described by an empirical input/output model, while it in other regimes can be described by a mechanistic state-space model. In other words, the framework is flexible with respect to the type of prior knowledge that can be applied. Moreover, the framework supports incremental modeling and simple model maintenance to some extent, because the modification of a single local model or operating regime has a predictable effect on the global model. In addition, we may use local models with different levels of accuracy according to what is required in different operating regimes.
- The local nature of the model representation has advantages when the model is applied for model based control, and in particular adaptive control. One reason is that only a few model parameters will be relevant at a given time instant. Hence, one will reduce drift phenomena in an on-line parameter estimator caused by lack of global persistence of excitation.

These properties will be further discussed and supported in the remaining of this thesis.

## 1.4 Outline of Thesis

First a few words on style. This thesis is based on a number of manuscripts that are published separately, or submitted for publication. We have attempted to reduce the redundancy somewhat, and more or less rewritten each manuscript. The notation within each chapter will be consistent, but there may be some minor inconsistencies in notation between the various chapters. Instead of providing the reader with a complete nomenclature list, each symbol is defined when first applied in each chapter. With a few standard expections, abbreviations are avoided. If not explicitly stated, the mathematical notation follows the textbook of Kreyszig (1978). The mathematical precision varies somewhat throughout this thesis. We have tried to apply the most natural precision level in the given context. For example, in theorems, no compromise on the mathematical precision is made, while in the definitions of some mathematical models it may be quite sloppy. In particular, we do not always distinguish between a function and its value, or a random variable and its realization.

The 2nd and 3rd chapters describe the core of the operating regime based modeling framework, and illustrate the basic approach with some examples.

Structure identification is a major issue in this thesis. In Chapters 4 and 5, we describe some new results. These results are of general interest, and are in particular applied in Chapter 6, that describes an algorithm for structure identification in the operating regime based modeling framework. Furthermore, identification is also the topic of Chapter 7.

Model predictive control and adaptive control using operating regime based models is discussed in Chapters 8 and 9, respectively. A theoretical contribution in the field of robust adaptive control can be found in Appendix A.

This organization of the thesis is chosen because it allows linear reading without jumping back and forth.

**Chapter 2. Operating Regime based Input/Output Modeling.** Some basic properties of the model representation based on interpolating a number of local models, are elaborated upon. First, we prove that the applied interpolation procedure is optimal, in a certain sense. Next, some approximation theoretical bounds on the modeling error are developed. The developments are constructive and provides fundamental insight into the design trade-offs. Moreover, a procedure for the development of NARMAX models is described, and the procedure is illustrated with some experimental results from a laboratory heat transfer process. The optimality result can be found in

Tor A. Johansen, "On the Optimality of the Takagi-Sugeno-Kang Fuzzy Inference Mechanism", *4th IEEE Int. Conference on Fuzzy Systems, Yokohama, Japan*, 1995 (accepted).

The approximation theoretical results and the modeling procedure is published in

Tor A. Johansen and Bjarne A. Foss, "Constructing NARMAX Models using ARMAX Models", *Int. J. Control*, Vol. 38, pp. 1125-1153, 1993.

The experimental results are from

Tor A. Johansen and Bjarne A. Foss, "Empirical Modeling of a Heat Transfer Process using Local Models and Interpolation", *American Control Conference, Seattle, Wa.*, 1995 (submitted).

**Chapter 3. Operating Regime based State-space Modeling.** The modeling procedure in Chapter 2 is extended to state-space modeling, and the additional difficulties are discussed. Simulation examples from the areas of population dynamics and biochemical engineering are used to illustrate the ideas. The core of this chapter was presented in

Tor A. Johansen and Bjarne A. Foss, "State-space Modeling using Operating Regime Decomposition and Local Models", *Preprints 12th IFAC World Congress, Sydney, Australia*, Vol. 1, pp. 431-434, 1993.

**Chapter 4. Model Structure Identification using Empirical Data from a Limited Operating Range.** A new structure identification criterion based on bootstrap estimation is developed. The purpose is to give reliable identification of model structure (in the sense of a model with good extrapolation properties) in cases when the model is required to work under a wider range of operating conditions than reflected in the identification data sequence. This work can be found in

Tor A. Johansen and Bjarne A. Foss, "A Dynamic Modeling Framework based on Local Models and Interpolation - Combining Empirical and Mechanistic Knowledge and Data", *Computers and Chemical Engineering* (submitted).

**Chapter 5. Model Structure Identification using Separate Validation Data - Asymptotic Properties.** Here we prove that under reasonable assumptions, the identification of model structure on the basis of a separate validation data sequence will asymptotically give the best possible model in the model set. This result applies to the structure identification algorithm in Chapter 6. This chapter is contained in

Tor A. Johansen and Erik Weyer, "Model Structure Identification using Separate Validation Data - Asymptotic Properties", *European Control Conference, Rome*, 1995 (submitted).

**Chapter 6. Identification of Operating Regimes.** We propose an algorithm that will automatically find a decomposition into regimes and local models on the basis of empirical data. The input to the algorithm is in addition to a sequence of empirical data, a set of alternative local model structures. Some statistical properties of this algorithm are briefly discussed. The algorithm is applied to the same simulated biochemical modeling problem as in Chapter 3, a simulated pH neutralization tank, and an experimental hydraulic manipulator modeling problem. In this last example, the results are compared to three other state-of-the-art empirical modeling algorithms from the literature. A short version of this chapter is

Tor A. Johansen and Bjarne A. Foss, "Identification of Non-linear System Structure and Parameters using Regime Decomposition", *Preprints IFAC Symposium on System Identification, Copenhagen*, Vol 1, pp. 131-136, 1994. To appear in *Automatica*, March 1995 (accepted).

Some of the examples are taken from the following manuscript

Tor A. Johansen and Bjarne A. Foss, "A Dynamic Modeling Framework based on Local Models and Interpolation - Combining Empirical and Mechanistic Knowledge and Data", *Computers and Chemical Engineering* (submitted).

**Chapter 7. Identification of Non-linear Systems using Prior Knowledge and Empirical Data - An Optimization Approach.** In this chapter we illustrate how various kinds of prior knowledge can be formulated as constraints or penalties in a prediction error identification criterion. The purpose of this is to elevate the modeling problem from the specification of a model structure in terms of e.g. a set of basis-functions, to a higher level. It is shown how different kinds of prior knowledge leads to different optimal model structures in some simple cases, and a practical "almost optimal" numerical procedure is suggested for the general case. The choice of optimization criterion corresponds to modeling in this approach, and a procedure that support the choice of criterion is suggested. The method is illustrated with the same simulated pH neutralization tank as in Chapter 6. This chapter is based on

Tor A. Johansen, "Identification of Non-linear Systems using Prior Knowledge and Empirical Data - An Optimization Approach", *Automatica* (submitted).

**Chapter 8. Operating Regime based Model Predictive Control.** With a simulation example, we illustrate the application of the operating regime based modeling framework for model predictive control of a batch fermentation reactor. The material in this chapter is taken from

Bjarne A. Foss, Tor A. Johansen and Aage V. Sørensen, "Nonlinear Predictive Control using Local Models - Applied to a Batch Process", *Preprints IFAC Symposium on Advanced Control of Chemical Processes (ADCHEM), Kyoto, Japan*", pp. 225-230, 1994. Also in *Control Engineering Practise*, summer 1995 (accepted).

**Chapter 9. Operating Regime based Adaptive Control.** We illustrate how the operating regime based modeling framework may give model structures that are attractive for the control of non-linear systems. We restrict our attention to a discrete-time non-linear decoupling controller based on a MIMO NARX model representation. First, we prove closed loop stability and robustness of the non-linear control system. Next, we formulate a certainty equivalence adaptive controller based on the same model representation, and develop a modified parameter estimation algorithm that only updates the parameters of the most relevant local models at each time instant. We show that this algorithm has basically the same convergence properties as the standard projection algorithm. Finally, we prove robust stability of this adaptive control system. The method and design trade-offs are illustrated with a simulation example, where a $2 \times 2$ Continuous Stirred Tank Reactor (CSTR) is controlled. This work was presented in

Tor A. Johansen, "Adaptive Control of MIMO Non-linear Systems using Local ARX Models and Interpolation", *Preprints IFAC Symposium on Advanced Control of Chemical Processes (ADCHEM), Kyoto, Japan*", pp. 159-166, 1994.

and published in a fuzzy disguise in

Tor A. Johansen, "Fuzzy Model Based Control: Stability, Robustness, and Performance Issues", *IEEE Transactions on Fuzzy Systems*, Vol. 2, pp 221-234, 1994.

**Chapter 10. Conclusions.** This chapter summarizes our main conclusions and contains some suggestions for future work.

**Appendix A. Robust Adaptive Control of Slowly Time-varying Nonlinear Systems.** Here we prove global stability and robustness of a SISO discrete-time adaptive non-linear decoupling controller for a quite general class of slowly time-varying non-linear systems. This result is a generalization of the related results in Chapter 9, and based on

Tor A. Johansen, "Weighted $l_2$-norms for Analysis of an Adaptive Control Loop based on a Non-linear Model", *IEEE Transactions on Automatic Control* (submitted).

Some results in the same spirit are given in

Tor A. Johansen and Petros A. Ioannou, "Robust Adaptive Control of Minimum Phase Non-linear Systems", *Int. J. Adaptive Control and Signal Processing* (submitted).

## 1.5 Literature overview

First we will give a brief overview of some recent approaches to hybrid modeling from the chemical engineering literature. Then an overview of some local modeling approaches will be presented.

### 1.5.1 Hybrid Modeling – In Chemical Engineering

Recently, the hybrid modeling problem has attracted considerable attention. One reason for this may be the interest in non-linear empirical modeling and neural networks over the last decade, combined with the increasing demand for non-linear models to be applied in intelligent controllers, diagnosis systems, and optimization based systems for design, scheduling, and quality or profit maximization. A very brief overview of some of these approaches follows.

The interactions between empirical and mechanistic knowledge is widely discussed in the engineering and scientific literature. Through a series of work, e.g. (Box and Youle 1955, Box and Hunter 1965, Box and Draper 1987), the interplay between model purpose, experimentation, empirical models, and mechanistic models has been discussed in great detail. Referring to Fig. 1.1, the tools developed by Box and co-workers will support the transition from a state with little system knowledge, to either an empirical or mechanistic model, rather than to a hybrid model. A more direct approach to the hybrid modeling problem is the grey-box approach of Bohlin (1989), see also (Bohlin and Graebe 1994). Grey-box models are characterized by a structure that is mainly mechanistic, but augmented with stochastic elements, empirical relations and parameters that need not have a direct physical interpretation. Physically based constraints that ensure the model is consistent with certain aspects of the physical reality, and prior knowledge about the parameters, can also be utilized (Tulleken 1993). In another approach, Lindskog and Ljung (1994) illustrates how simple mechanistic knowledge can be used to find adequate non-linearities in a semi-empirical model. Kramer, Thompson and Phagat (1992) has proposed an approach where an a priori given mechanistic model is used as a default model, and an empirical model based on a radial-basis function series expansion compensates for the possible mismatch between the prediction of the more or less inaccurate mechanistic model and the observed data, see also (Thompson and Kramer 1994, Su, Bhat and McAvoy 1992). In addition, the parameters of the empirical model part may be constrained. A related, but somewhat different approach was taken by Jian (1993) in his model of a blast furnace. A linear empirical (ARMAX) model with an on-line estimator was used as a rough model of the system. The difference between the predictions given by the ARMAX model and the system output was modeled with a set of empirical qualitative rules from a knowledge-base that was designed with the help of experienced operators. This hybrid model was reported to give better predictions than either the ARMAX model or the predictions of the operators alone. Another approach to the modeling problem is described in (Psichogios and Ungar 1992) and (Aoyama and Venkatasubramanian 1993). A mechanistic model structure containing some unknown internal variables is supposed to be known. The unknown internal variables will typically be reaction rates or thermodynamic properties that depend in a complex and often poorly known manner on a number of other model variables. The idea is to use a neural network as an empirical model of these dependences. Yet another approach will be the main topic of this thesis.

### 1.5.2 Local Models

The basic idea is simple, and several scientists and engineers has coined similar ideas. It is particularly interesting to observe that when we have presented these ideas for engineers practising in industry, "everybody" seems to recognize the basic concept. However, the industrial applications have in general been ad. hoc., in the sense that no formalized framework or theoretical foundation has been applied. Unfortunately, the industrial results are typically not published, and the proceeding literature overview will therefore mainly consist of contributions from academia.

The earliest model we have found, that is directly based on local models, is from the field of mathematical biology, and dates back to Kolmogoroff (1936). The model is an alternative to the classical Lotka-Volterra model, and based on a decomposition of the system into different regimes (called zones) with qualitatively different behavior. This model will be discussed in some detail in section 3.3.1.

Piecewise linear models clearly fall into this framework. Early contribution describing modeling procedures based on piecewise linear models and an optimization formulation of the regime decomposition problem are found in (Bellman 1961*b*, Opoitsev 1970, Dorofeyuk, Kasavin and Torgovitsky 1970, Kasavin 1972, Rajbman, Dorofeyuk and Kasavin 1981). Various other procedures are described in (Breiman and Meisel 1976, Cyrot-Normand and Mien 1980, Tong and Lim 1980, Haber, Vajk and Keviczky 1982, Omohundro 1987, Farmer and Sidorowich 1987, Billings and Voon 1987, Billings and Chen 1989, Kavli 1990, Strömberg, Gustafsson and Ljung 1991, Skeppstedt, Ljung and Millnert 1992, Hilhorst 1992, Söderman, Top and Strömberg 1993).

A particularly interesting piecewise modeling approach from the biochemical engineering literature is described in (Zhang, Visala, Halme and Linko 1994), where the application of a prior knowledge for the decomposition into regimes (called functional states) is highlighted.

The use of more complex local models than linear, like a piecewise combination of several local neural network models (Sørheim 1990) and local polynomial models (Pottmann, Unbehauen and Seborg 1993) are also described in the literature. Clearly, piecewise polynomials are closely related to the spline-based modeling approach, e.g. (Wahba 1990). The main difference lies in the way splines handles multi-dimensionality. While piecewise polynomials are multi-dimensional by nature, splines are not, and multi-dimensionality is typically introduced by the means of tensor products.

In our view, the most interesting approach in the local modeling literature is the one of Takagi and Sugeno (1985), where each operating regime is represented as a fuzzy set (which may roughly be viewed as a set with soft boundaries). This representation is appealing, since many systems change behavior smoothly as a function of the operating point, and the soft transition between the regimes introduced by the fuzzy sets representations captures this feature in an elegant fashion. Algorithms for the identification of a decomposition into regimes on the basis of data are described by Sugeno and Kang (1988) and several experimental and simulated examples are given (Takagi and Sugeno 1985, Sugeno and Kang 1986, Sugeno and Kang 1988, Nakamori, Suzuki and Yamanaka 1992). Related identification algorithms based on cluster analysis are described in (Bezdek, Coray, Gunderson and Watson 1981*a*, Bezdek, Coray, Gunderson and Watson 1981*b*, Hathaway and Bezdek 1993, Yoshinari, Pedrycz and Hirota 1993, Nakamori and Ryoke 1994), and similar ideas are applied in chemometrics by Næs and Isaksson (1991) and Næs (1991).

Essentially the same ideas has evolved in the neural network community as well, where Stokbro, Hertz and Umberger (1990), Jones and co-workers (1991) describe approaches to modeling based on local linear models as generalizations of neural networks with localized receptive fields and radial basis-functions, e.g. (Moody and Darken 1989). Simultaneously, Jacobs, Jordan, Nowlan and Hinton (1991) developed a framework based on "local experts", and an effective identifiaction algorithm were later developed by Jordan and Jacobs (1993). Other identification algorithms were developed by Murray-Smith (1992), and Murray-Smith and Gollee (1994).

A standard approach to non-linear control design is gain scheduling, which is essentially a collection of local controllers. Different controllers are activated (scheduled) under different operating conditions. The operating regime based modeling approach has obvious similarities (it may be descibed as model scheduling). Notice that the design of a non-linear controller based on a combination of local models is more sophisticated than the use of gain scheduling, see also (Shamma and Athans 1990, Rugh 1991) that analyze the shortcomings of gain scheduling.

The works of Stokbro et al. (1990) and Jones and co-workers (1991) were the initial inspiration for the present thesis, and our initial ideas were published in (Johansen and Foss 1992*b*, Johansen and Foss 1992*a*, Johansen and Foss 1992*c*, Foss and Johansen 1993).

It is interesting to observe that most of this literature was found more or less by luck, i.e. browsing volumes of proceedings and journals. Essentially nothing was found by database searches. The reason for this is the interdisciplinary nature of these ideas, which have their origin in very different communities, including biology, fuzzy sets, physics, mathematics, statistics, neural nets, chemical engineering, control and identification theory, to mention some. The terminology varies considerably, which makes database search almost impossible. We expect there are significant contributions in perhaps completely different areas that we have failed to discover.

## 1.6 Contributions

Briefly, the main contributions in this thesis are

- The design procedure for operating regime based model structures based on a combination of qualitative, quantitative, empirical, and mechanistic system knowledge outlined in Chapters 2 and 3.
- The structure identification algorithm in Chapter 6.
- The bootstrap structure identification criteria described in Chapter 4.
- The asymptotic analysis of the structure identification criterion based on separate validation data, Chapter 5.
- The formulation of the modeling and identification problem as an optimization problem in Chapter 7.
- The proof of optimality of the local model interpolation method in section 2.1.
- The constructive approximation theoretical results in section 2.1.
- The procedure for combining local models with different state-spaces, section 3.1.1.
- The modified on-line identification algorithm in section 9.4.
- The proof of stability in Chapter 9.
- The application of weighted $l_2$-norms in the stability proof in Appendix A.

# Chapter 2

# Operating Regime based Input/Output Modeling

For some applications, one wants a model that only describes the input/output behavior of the system. The ARMAX model representation is a well known linear input/output model representation, while the NARMAX (Nonlinear ARMAX) model representation is an extension that represents the model as a nonlinear mapping of past inputs, outputs and noise terms to future outputs. In this paper we discuss how NARMAX models can be represented in the operating regime based modeling framework, and discuss in detail how a NARMAX model can be constructed from a set of local ARMAX models.

This chapter is organized as follows: First, in section 2.1, we present a model representation based on local models, and builds a theoretical foundation for this representation. The operating regime concept is introduced and we present a general result guiding the choice of variables used to characterize the operating regimes. Thereafter, we discuss some practical aspects of modeling using local ARMAX models in section 2.2, and some aspects of system identification in section 2.3. In section 2.4 the procedure is illustrated with some experimental results, and section 2.5 contains some discussion.

## 2.1 Model Representation

The NARMAX model representation

$$y(t) = f(y(t-1), ..., y(t-n_y), u(t-1), ..., u(t-n_u), e(t-1), ..., e(t-n_e)) + e(t) \quad (2.1)$$

is shown by Leontaritis and Billings (1985) to be able to represent the observable and controllable modes of a large class of discrete-time non-linear state-space models. Here $y(t) \in Y \subset R^m$ is the output vector, $u(t) \in U \subset R^r$ is the input vector, and $e(t) \in E \subset R^m$ is noise. Throughout this chapter we will assume $n_y$, $n_u$ and

$n_e$ are known, and focus on the problem of representing and constructing the nonlinear function $f : \Psi \rightarrow R^m$. We introduce the $(m(n_y + n_e) + rn_u)$-dimensional information vector

$$\psi(t-1) = \begin{pmatrix} y(t-1) \\ \vdots \\ y(t-n_y) \\ u(t-1) \\ \vdots \\ u(t-n_u) \\ e(t-1) \\ \vdots \\ e(t-n_e) \end{pmatrix}$$

belonging to the set $\Psi = Y^{n_y} \times U^{n_u} \times E^{n_e}$. This enables us to write (2.1) in the form

$$y(t) = f(\psi(t-1)) + e(t) \tag{2.2}$$

Provided necessary smoothness conditions on $f$ are satisfied, a general way of approximating $f$ is by series expansions. A first order Taylor-series expanded about an equilibrium point yields an ARMAX model. Second- and third-order Taylor-expansions are possible, while higher-order Taylor-expansions are not very useful in practice because the number of parameters in the model increases rapidly with the expansion order, and because of the poor extrapolation and interpolation capabilities of higher-order polynomials. Splines offers one possible solution to this problem, e.g. (Wahba 1990), where the idea is to patch together low-order polynomials. A representation that is closely related to splines in spirit, but still very different for multi-dimensional modeling problems, is based on patching together *local models*. In its basic form, it was introduced by Takagi and Sugeno (1985), and rediscovered independently by Stokbro et al. (1990), Jones and co-workers (1991), and (Jacobs et al. 1991). In its present form, it was introduced in (Johansen and Foss 1992*b*, Johansen and Foss 1992*a*). The material in this chapter is mainly based on (Johansen and Foss 1993*a*).

### 2.1.1 Optimal combination of Local Models

Suppose $N$ local models (indexed by $i \in \{1, 2, ..., N\}$)

$$y(t) = \hat{f}_i(\psi(t-1)) + e(t) \tag{2.3}$$

are available, and the different local models are accurate under different operating conditions. Hence, under some operating conditions there may be several local models that are accurate, while no local model may be accurate under other conditions. Suppose the relative validity (or relevance) of each local model is indicated by the functions $\tilde{\rho}_1, \tilde{\rho}_2, ..., \tilde{\rho}_N : \Psi \rightarrow [0, 1]$. If at a given $\psi \in \Psi$ the local model indexed with $i$ is accurate, then $\tilde{\rho}_i(\psi)$ will be close to one, while $\tilde{\rho}_i(\psi)$ is close to zero for all $\psi \in \Psi$ where local model $i$ is inaccurate.

This immediately raises the question about how to predict the system output for those $\psi$ where it is not exactly one local model is accurate. We essentially seek a global model

$$y(t) = \hat{f}(\psi(t-1)) + e(t)$$

based on a combination of the local models (2.3). From the definition of $\tilde{\rho}_i$, it is natural to require that $\hat{f}(\psi)$ should be close to $\hat{f}_i(\psi)$ at those $\psi \in \Psi$ where $\tilde{\rho}_i(\psi)$ is large. The subset of $\Psi$ where "$\tilde{\rho}_i(\psi)$ is large"[1] is denoted $\Psi_i$. This suggests that $\hat{f}$ should minimize a criterion such as

$$M(\hat{f}) = \sum_{i=1}^{N} \int_{\psi \in \Psi} \left|\left|\hat{f}(\psi) - \hat{f}_i(\psi)\right|\right|_2^2 \tilde{\rho}_i(\psi) d\psi$$

where $||\cdot||_2$ is the Euclidean norm. We now have the following result (Johansen 1995)

**Theorem 1** *Suppose*

1. *the functions* $\hat{f}_1, \hat{f}_2, ..., \hat{f}_N$ *belong to* $C^m(\Psi)$, *the set of all continuous m-dimensional functions defined on* $\Psi$,
2. $\sum_{i=1}^N \tilde{\rho}_i(\psi) > 0$, *for all* $\psi \in \Psi$.

*Then the function* $\hat{f}$ *defined by*

$$\hat{f}(\psi) = \sum_{i=1}^{N} \hat{f}_i(\psi)\tilde{w}_i(\psi) \tag{2.4}$$

$$\tilde{w}_i(\psi) = \frac{\tilde{\rho}_i(\psi)}{\sum_{j=1}^N \tilde{\rho}_j(\psi)} \tag{2.5}$$

*minimizes* $M$ *on* $C^m(\Psi)$.

**Proof:** The variation of $M$ with respect to any perturbation $\Delta f \in C^m(\Psi)$ is

$$\delta M(\hat{f}; \Delta f) = 2\sum_{i=1}^{N} \int_{\psi \in \Psi} (\hat{f}(\psi) - \hat{f}_i(\psi))\tilde{\rho}_i(\psi)\Delta f(\psi) d\psi$$

Notice that $M$ is strictly convex. A necessary and sufficient condition for global optimality of $\hat{f}$ is now (Luenberger 1969)

$$\sum_{i=1}^{N}(\hat{f}(\psi) - \hat{f}_i(\psi))\tilde{\rho}_i(\psi) = 0 \qquad \text{for all } \psi \in \Psi$$

From the second assumption, it is evident that $\hat{f}$ is well defined by this equation.

□

[1] We will later illustrate how the expression "is large" can be made mathematically precise by the introduction of fuzzy set theory. For the time being, let us just accept this unprecise definition of $\Psi_i$.

The first assumption is quite reasonable, since we are allowed to choose the local models ourselves in most applications. The second assumption simply means that at least one local model should be relevant under every possible operating condition. This is also reasonable, since we do not need an *accurate* local model, just a *relevant* one.

We call the functions in the set $\{\tilde{w}_i\}_{i=1}^N$ *interpolation functions* because they are used to interpolate[2] between *local models* $\hat{f}_i$, which are supposed to be accurate descriptions of "the true $f$" locally. The relative weight given to each local model in the interpolation is determined by the relative validity of the local models. We can interpret $\tilde{w}_i$ as a function that has largest value in the parts of $\Psi$ where the function $\hat{f}_i$ is the *best* approximation to $f$, and close to zero elsewhere (notice that *best* does not imply *good*). By the definition of $\tilde{w}_i$ we know that $\sum_{i=1}^N \tilde{w}_i(\psi) = 1$ for all $\psi \in \Psi$. Hence, the resulting global model is a convex combination of the local models. In the remaining of this thesis we will consistently apply this method for combining local models.

### 2.1.2 The Model Set

The set of all functions of the form (2.4) with local models of polynomial order $p$ is denoted

$$\tilde{\mathcal{F}}_p(\tilde{\mathcal{K}}) = \left\{ \hat{f} : \Psi \rightarrow R^m \;\middle|\; \hat{f}(\psi) = \sum_{i=1}^N \hat{f}_i(\psi)\tilde{w}_i(\psi) \right\}$$

It is assumed that the function $\tilde{\rho}_i$ belong to a given set $\tilde{\mathcal{K}}$ of smooth kernel functions that go sufficiently fast to zero, for example the set of all Gaussians:

$$\tilde{\mathcal{K}} = \left\{ k : \Psi \rightarrow [0,1] \;\middle|\; k(\psi) = \exp\left(-\frac{1}{2}(\psi-\psi_i)^T\Sigma_i^{-1}(\psi-\psi_i)\right) \right\} \tag{2.6}$$

where $\Sigma_i$ is a positive definite scaling matrix. The choice of $\tilde{\mathcal{K}}$ will be elaborated upon later.

Next, let us briefly discuss the choice of local model structures. At the extreme, a zeroth order Taylor-expansion of $f$ about a point $\psi_i \in \Psi_i$ may be used to define the local model

$$\hat{f}_i(\psi) = f(\psi_i) \tag{2.7}$$

or the parameterized local model structure

$$\hat{f}_i(\psi) = \theta_i \tag{2.8}$$

where $\theta_i$ is an unknown parameter vector.

[2] Strictly speaking, it is not an interpolation, rather a combination. However, we will still use the word interpolation throughout this thesis.

A first order Taylor-expansion of $f$ about $\psi_i$ provides better extrapolation and interpolation than the zeroth order expansion (2.7). Assuming the first derivative of $f$ in $\psi_i$ exists, the local models are given by

$$\hat{f}_i(\psi) = f(\psi_i) + \nabla f(\psi_i)(\psi - \psi_i) \tag{2.9}$$

or parameterized local model structure may be defined as

$$\hat{f}_i(\psi) = \theta_i + \Theta_i(\psi - \psi_i) \tag{2.10}$$

where $\theta_i$ is an unknown parameter vector and $\Theta_i$ is an unknown parameter matrix. Observe that (2.9) is actually an ARMAX model resulting from a linearization of (2.2) about $\psi_i$. Both the Weighted Linear Maps of Stokbro et al. (1990), Stokbro and Umberger (1990) and the Connectionist Normalized Linear Spline Networks of Jones and co-workers (1991), and Jones, Lee, Barnes, Flake, Lee, Lewis and Qian (1989) use a first order expansion locally. This representation makes it possible to construct a NARMAX model by interpolating a number of ARMAX models.

Higher order and non-polynomial local models can also be used. Furthermore, there is no requirement that all the local models must have the same structure. In particular, the local models may have different dynamic or polynomial order.

### 2.1.3 Approximation Properties

It is intuitive that any smooth function can be approximated arbitrarily well with this representation by making the decomposition of $\Psi$ into subsets $\Psi_i$ sufficiently fine. This is indeed the case, as illustrated in the following. We use the following norm to measure the approximation accuracy

$$||f - \hat{f}||_\infty = \sup_{\psi \in \Psi} ||f(\psi) - \hat{f}(\psi)||_2$$

The $(p+1)$th derivative of the vector function $f$ at the point $\psi$ is denoted $\nabla^{p+1} f(\psi)$. Assume $f$ is continuously differentiable $p+1$ times, and $\hat{f}_i$ is a local model equal to the first $p+1$ terms of the Taylor-series expansion of $f$ about $\psi_i$. For any $\psi \in \Psi$, we have

$$f(\psi) - \hat{f}(\psi) = \sum_{i=1}^{N} (f(\psi) - \hat{f}_i(\psi))\tilde{w}_i(\psi)$$

If we assume $||\nabla^{p+1} f(\psi)|| < M$ for all $\psi \in \Psi$, where $||\cdot||$ denotes the induced norm, we obtain by Taylor's theorem

$$||f(\psi) - \hat{f}(\psi)||_2 < \sum_{i=1}^{N} \frac{M}{(p+1)!} ||\psi - \psi_i||_2^{p+1} \ \tilde{w}_i(\psi)$$

In order to ensure that the right-hand side is no larger than an arbitrary $\epsilon > 0$, we must ensure that for any $\psi \in \Psi$ the following condition holds

$$\sum_{i=1}^{N} ||\psi - \psi_i||_2^{p+1} \tilde{\rho}_i(\psi) \leq \epsilon \frac{(p+1)!}{M} \sum_{i=1}^{N} \tilde{\rho}_i(\psi) \tag{2.11}$$

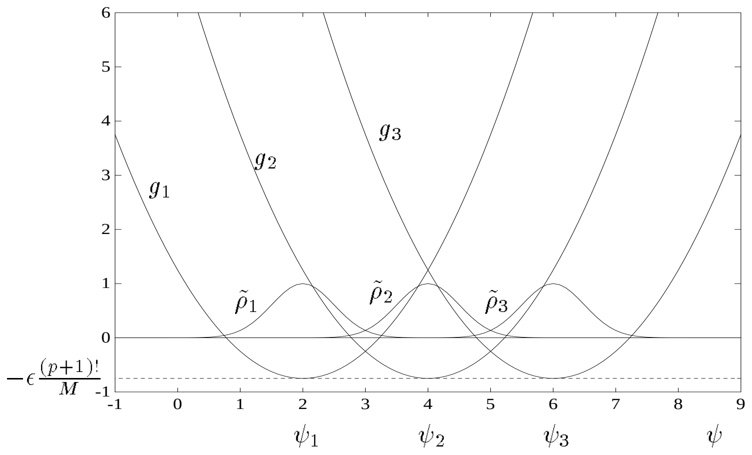


Figure 2.1: A geometric interpretation of the constraints on $\tilde{\rho}_i$

Defining the function $g_i : \Psi \to R$ by

$$g_i(\psi) = ||\psi - \psi_i||_2^{p+1} - \epsilon\frac{(p+1)!}{M}$$

and rewriting (2.11) gives the following condition that must hold for any $\psi \in \Psi$

$$\sum_{i=1}^{N} g_i(\psi)\tilde{\rho}_i(\psi) \leq 0 \tag{2.12}$$

or equivalently

$$\sum_{i=1}^{N} g_i(\psi)\tilde{w}_i(\psi) \leq 0 \tag{2.13}$$

The problem is now to find conditions on $N$ and the set of functions $\{\tilde{\rho}_i\}_{i=1}^N \subset \tilde{\mathcal{K}}$ to ensure that equation (2.12) holds for any given $\epsilon > 0$. A geometric interpretation of (2.12) is given in Figure 2.1. Certainly, this equation holds if the negative contribution of one term $g_i(\psi)\tilde{\rho}_i(\psi)$ in (2.12) dominates the (possibly positive) contribution of all other terms.

Notice that the *shape* of the $g_i$-functions is fixed and given by the specifications. We are, however, free to choose the location and number of local models. Let us choose the set $\{\psi_i\}_{i=1}^N$ so large and "sufficiently dense in $\Psi$" that at least one of the functions $\{g_i\}_{i=1}^N$ will be negative at any $\psi \in \Psi$. Then the functions $\{g_i\}_{i=1}^N$ are fixed, and we must choose the $\tilde{\rho}_i$-functions such that (2.12) hold. This can be accomplished in several ways. In the limit when the width of the $\tilde{\rho}_i$-functions go to zero, the interpolation functions $\tilde{w}_i$ will approach step-functions as shown in Figure 2.2. The model will then approach a piecewise constant model if $p = 0$, a

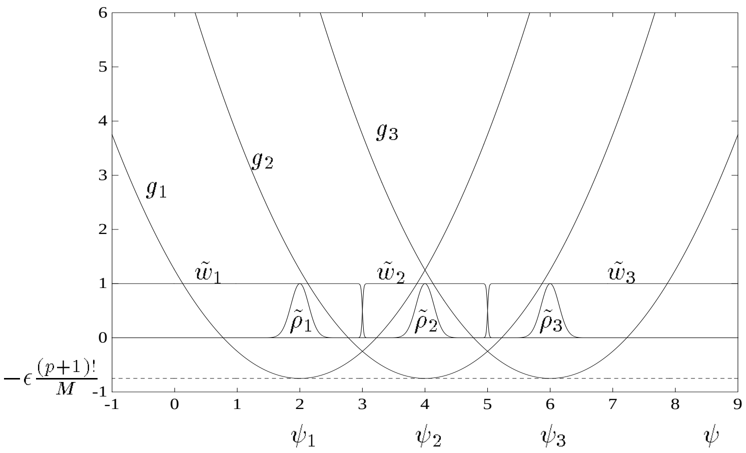


Figure 2.2: Situation when the width of $\tilde{\rho}_i$ goes to zero.

piecewise linear model if $p = 1$, etc. In this limit, at any $\psi \in \Psi$ there will exist a $j$ such that

$$\tilde{w}_i(\psi) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

By the choice of $\{\psi_i\}_{i=1}^N$ we know that $g_j(\psi) < 0$, and since $\tilde{w}_i(\psi) = 0$ for $i \neq j$, (2.13) will hold. We can now provide a result for the case when $\Psi$ is a compact set:

**Theorem 2** *Suppose any integer $p \geq 0$ is given, and $f$ has continuous $(p+1)$-th derivative. If $\Psi$ is compact, then for any $\epsilon > 0$ there is an $\hat{f} \in \tilde{\mathcal{F}}_p(\tilde{\mathcal{K}})$ such that $||f - \hat{f}||_\infty < \epsilon$*

**Proof:** Since $\nabla^{p+1} f(\psi)$ is continuous, its norm is bounded (by $M < \infty$) on the compact set $\Psi$. Moreover, since $\Psi$ is compact, a finite $N$ is sufficient to ensure that at least one $g_i$-function is negative at any point in $\Psi$. Since $N$ is finite, it is not necessary to go to the limit and make $\{\tilde{w}_i\}_{i=1}^N$ step-functions, but choose $\{\tilde{\rho}_i\}_{i=1}^N$ as smooth kernel functions that are sufficiently narrow for (2.13) to hold. This is possible because the kernel functions go to zero sufficiently fast.

□

This is an existence result. However, the proof is constructive and gives indications on how to construct the approximation. In order to use this proof to formulate an upper bound on the approximation error, we introduce the following definition of distance between sets, similar to the Haussdorf metric:

**Definition 1** *Assume $A$ and $B$ are non-empty subsets of $R^n$. Then the distance between the sets is defined as*

$$\mathcal{D}(A, B) = \inf_{a \in A} \sup_{b \in B} ||a - b||_2$$

□

The crux in the proof of Theorem 2 is that at any point $\psi \in \Psi$ one of the $g_i$-functions is negative and that the $\tilde{\rho}_i$-functions are chosen such that the negative term $g_i(\psi)\tilde{\rho}_i(\psi)$ will dominate the sum (2.12). At least one $g_i$-function will be negative at any $\psi \in \Psi$ if the following condition holds

$$\mathcal{D}(\{\psi_i\}, \Psi) \leq \left( \epsilon \frac{(p+1)!}{M} \right)^{\frac{1}{p+1}} \tag{2.14}$$

If the set $\{\psi_i\}_{i=1}^N$ is dense in $\Psi$, this distance will be zero. The term "sufficiently dense" used informally above, means that the set $\{\psi_i\}_{i=1}^N$ should be chosen such that (2.14) holds for the given $\epsilon$.

**Theorem 3** *Suppose an integer $p \geq 0$ is given. If $\Psi$ is compact, and $f$ has bounded $(p+1)$-th derivative, i.e. $||\nabla^{p+1} f(\psi)|| \leq M$ for all $\psi \in \Psi$, then there exists polynomials $\hat{f}_1, ..., \hat{f}_N$ of order $p$ and smooth kernel functions $\tilde{\rho}_1, ..., \tilde{\rho}_N \in \tilde{\mathcal{K}}$ such that the approximation error is bounded by*

$$||f - \hat{f}||_\infty < \frac{M}{(p+1)!} \left(\mathcal{D}(\{\psi_i\}, \Psi)\right)^{p+1} \tag{2.15}$$

**Proof:** Eq. (2.14) will hold for $\epsilon$ equal to the right-hand side of (2.15). From the previous discussion, it is evident that (2.13) holds for any $\psi \in \Psi$. Hence, $||f - \hat{f}||_\infty$ will be bounded by $\epsilon$, and the result follows.

□

The result above holds under the condition that the $\tilde{\rho}_i$-functions are chosen sufficiently narrow, and the local models are truncated Taylor-series expansions of order $p$. Hence, the bound (2.15) may be conservative, since better local models with different parameters than truncated Taylor-series may exist. Moreover, for $\tilde{\rho}_i$-functions that are not too narrow and not too wide, one may expect better accuracy. From the deviations above it is evident that choosing $\tilde{\rho}_i$ from the set of Gaussians, cf. (2.6), is only one possibility. The Gaussian goes to zero faster than $g_i$ goes to infinity, which is clearly a desirable property. In general, a smooth kernel function set $\tilde{\mathcal{K}}$ can be generated by the application of wavelet functions (Benveniste, Juditsky, Delyon, Zhang and Glorennec 1994). Then $\tilde{\mathcal{K}}$ consists of all possible translations and scalings of a given wavelet function that go fast to zero. $\tilde{\mathcal{K}}$ defined by (2.6) is such a set, where translation is parameterized by $\psi_i$, and scaling by $\Sigma_i$.

Obviously, if $f$ does not satisfy the smoothness conditions in Theorem 2, the proof does not hold. However, if $f$ is such that it can be approximated arbitrarily well by a smooth function, then it is trivial to show that $f$ can be approximated arbitrarily well by smoothly interpolating local models. In patricular we have:

**Corollary 1** *The conclusion in Theorem 2 also holds if the smoothness assumption on $f$ is relaxed to assuming only continuity. In other words, the set $\tilde{\mathcal{F}}_p(\tilde{\mathcal{K}})$ is dense in $C^m(\Psi)$.*

**Proof:** By the Weierstrass approximation theorem, e.g. (Kreyszig 1978), for any $\epsilon > 0$ there exists a polynomial $\tilde{f}$ such that $||f - \tilde{f}||_\infty \leq \epsilon/2$. By Theorem 2, there exists an $\hat{f} \in \tilde{\mathcal{F}}_p(\tilde{\mathcal{K}})$ such that $||\tilde{f} - \hat{f}||_\infty < \epsilon/2$.

□

### Example 1

Assume $p = 1$, i.e. the local models are ARMAX models. Then (2.15) can be written

$$||f - \hat{f}||_\infty \le \frac{M}{2}(\mathcal{D}(\{\psi_i\}, \Psi))^2 = \epsilon \tag{2.16}$$

where $M$ is a bound on the Hessian of $f$. If $f$ is linear, then $M = 0$ and one local linear model is sufficient to make an arbitrary good global model (of course). $M$ is an indication of the nonlinearity or complexity of the function, and we expect $\epsilon$ to increase with increasing $M$, i.e. increasing nonlinearity, which is indeed the case, cf. (2.16). However, using the upper bound $M$ gives a conservative bound since $f$ may be more complex (less linear) in some regions than others. Hence, we do not need high density of local models in parts of $\Psi$ where the system does not have strongly nonlinear behavior.

□

### Example 2

With a simple example we illustrate an application of Theorem 3. Consider the function $f : [0, 2] \to R$ given by $f(\psi) = \psi^2 + 1$. Assume that we have two local linear models (truncated Taylor series expansions) located at $\psi_0 = 0.5$ and $\psi_1 = 1.5$. Then $\mathcal{D}(\{0.5, 1.5\}, [0, 2]) = 0.5$, $p = 1$ and $M = 2$. Theorem 3 predicts the bound $\epsilon = 0.25$ on the approximation accuracy. As shown by Figure 2.3, this bound is not conservative when using infinitely narrow functions $\tilde{\rho}_i$, i.e. a piecewise linear approximation. The reason for this is that $M = \frac{d^2 f}{d\psi^2}(\psi) = 2$ for all $\psi \in [0, 2]$, hence there are no parts of $\Psi$ where $f$ is "less nonlinear" or "less complex". As we shall see in Example 4, better approximations can be achieved using well-chosen $\tilde{\rho}_i$-functions. From Figure 2.3 it is also evident that if the local linear models are not chosen as a first order Taylor-expansion, but chosen on the basis of e.g. a least squares regression, better accuracy can be achieved.

□

Since the system function $f$ can be approximated arbitrarily well, we are able to make arbitrarily good predictions on a finite horizon provided there is no noise, the initial values are correct, and the model predicts outputs that give vectors $\psi$ that remain in $\Psi$ (Polycarpou, Ioannou and Ahmed-Zaid 1992). However, it is well known that the solution to some difference equations are sensitive with respect to incorrect initial values or noise. Examples of such systems are chaotic and unstable systems.

### 2.1.4 Operating Regimes

In the representation (2.4) the interpolation functions $\tilde{w}_i$ are defined on the set $\Psi$. This is a subset of the information-space. If the information-space has high dimension, the *curse of dimensionality* (Bellman 1961*a*) will restricts the applicability of all local modeling approaches. The core of this problem is that the

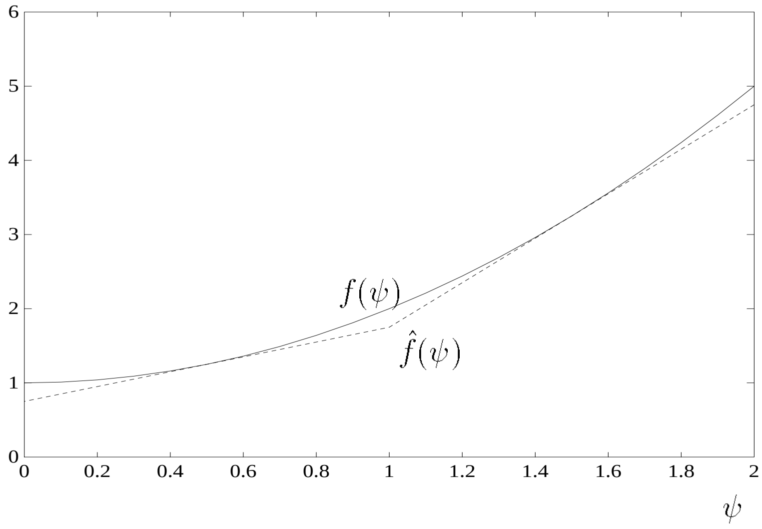


Figure 2.3: Approximation of $f(\psi) = \psi^2 + 1$ using two local linear models.

number of regions needed to uniformly partition a subset of a vector space increases exponentially with the dimension of the space. Uniform partitioning is often not necessary, but the problem is still significant. However, in some cases the interpolation functions may be defined on a space of lower dimension, which will reduce the curse of dimensionality considerably. This is our motivation for introducing the terms *operating regime* and *operating point*. First, we define $Z$ to be the *set of operating points*. An operating point $z \in Z$ is a vector of variables that characterizes the system's different modes of behavior. It should be mentioned that this definition of operating point differs from its definition in system theory, where it is an equilibrium point in the model's state-space.

It is convenient to define an operating regime $Z_i$ as a subset of $Z$ where the system behavior can be adequately described with a given local model structure. Hence, in the case of a local linear model structure, within an operating regime the system behavior is "approximately linear". A *local model validity function* $\rho_i : Z \longrightarrow [0,1]$ belongs to a set of smooth kernel functions $\mathcal{K}$, and satisfies $\rho_i(z) \approx 1$ for $z \in Z_i$, and goes to quickly zero outside $Z_i$. The interpolation functions $w_i : Z \longrightarrow [0,1]$ are now defined as

$$w_i(z) = \frac{\rho_i(z)}{\sum_{j=1}^{N} \rho_j(z)}$$

assuming that at every operating point $z \in Z$, not all local model validity functions vanish.

In many cases there will exist a function $H : \Psi \longrightarrow Z$ such that $z(t) = H(\psi(t))$ for all $t$. $H$ will typically be a projection, i.e. $Z$ will be in a sub-space or submanifold of the information space. When the operating point is calculated on the basis of filtered or estimated quantities, the relationship between $\psi(t)$ and $z(t)$ is more complex, and must be described by an operator $\mathcal{H}$ mapping past data to the operating point. Although of some importance, this complicates the analysis considerably, and we will not consider this case in this section.

To summarize, the model representation we investigate at this stage is

$$y(t) = \sum_{i=1}^{N} \hat{f}_i(\psi(t-1))w_i(z(t-1)) + e(t) \tag{2.17}$$

Next, we define the set of functions

$$\mathcal{F}_p(\mathcal{K}) = \left\{ \hat{f} : \Psi \rightarrow R^m \;\middle|\; \hat{f}(\psi) = \sum_{i=1}^{N} \hat{f}_i(\psi)w_i(H(\psi)) \right\}$$

where $p$ is the polynomial order of $\hat{f}_i$, and $\rho_i \in \mathcal{K}$. Now we want to state some general results regarding the function $H$ from the information space to the operating space. In general, $f$ can always be written as an affine function of the form

$$f(\psi) = f(\psi_L, \psi_N) = f_1(\psi_N) + f_2(\psi_N)\psi_L \tag{2.18}$$

where $\psi_L \in \Psi_L$, $\psi_N \in \Psi_N$, and $\Psi_L$ and $\Psi_N$ are sub-sets of $\Psi$ with the property $\Psi_L \times \Psi_N = \Psi$. Furthermore, $f_1 : \Psi_N \rightarrow R^m$ and $f_2 : \Psi_N \rightarrow R^{m\times \dim(\psi_L)}$ are non-linear vector- and matrix-valued functions, respectively. A trivial proof of (2.18) is the not very useful choice $\Psi_N = \Psi$ and $\Psi_L = \emptyset$. Our principal result guiding the choice of $Z$ and $H$ is the following, which indicates that with local linear models, the operating point must be chosen such that it captures the system's non-linearities:

**Theorem 4** *Assume $f$ given by (2.18) is continuous, and $\Psi$ is compact. Then for any $\epsilon > 0$ there exists an $\hat{f} \in \mathcal{F}_1(\mathcal{K})$ with $z = H(\psi) = \psi_N$ such that $||f - \hat{f}||_\infty < \epsilon$.*

**Proof:** Fix an arbitrary $\psi \in \Psi$ and let $\psi_L$ and $\psi_N$ be the corresponding elements in $\Psi_L$ and $\Psi_N$, respectively. Then

$$\begin{aligned} f(\psi) - \hat{f}(\psi) &= f(\psi_N, \psi_L) - \sum_{i=1}^{N} \hat{f}_i(\psi_N, \psi_L)w_i(\psi_N) \\ &= \sum_{i=1}^{N} (f_1(\psi_N) + f_2(\psi_N)\psi_L - \hat{f}_i(\psi_N, \psi_L))w_i(\psi_N) \\ &= \sum_{i=1}^{N} (f_1(\psi_N) - \hat{f}_{Ni}(\psi_N) + f_2(\psi_N)\psi_L - \hat{f}_{Li}(\psi_L))w_i(\psi_N) \end{aligned}$$

In the last line we split the linear function $\hat{f}_i : \Psi \rightarrow R^m$ into two linear functions $\hat{f}_{Ni} : \Psi_N \rightarrow R^m$ and $\hat{f}_{Li} : \Psi_L \rightarrow R^m$. We represent $\hat{f}_{Li}$ as $\hat{f}_{Li}(\psi_L) = \beta_i\psi_L$, where $\beta_i$ is a not yet specified constant parameter matrix. Then we have

$$||f(\psi) - \hat{f}(\psi)||_2 = \left|\left| \sum_{i=1}^{N} (f_1(\psi_N) - \hat{f}_{Ni}(\psi_N) + (f_2(\psi_N) - \beta_i)\psi_L)w_i(\psi_N) \right|\right|_2$$

$$
\begin{aligned}
&\leq \left\| \sum_{i=1}^{N} (f_1(\psi_N) - \hat{f}_{Ni}(\psi_N)) w_i(\psi_N) \right\|_2 \\
&\qquad + \left\| \sum_{i=1}^{N} (f_2(\psi_N) - \beta_i) \psi_L w_i(\psi_N) \right\|_2 \\
&\leq \left\| f_1(\psi_N) - \sum_{i=1}^{N} \hat{f}_{Ni}(\psi_N) w_i(\psi_N) \right\|_2 \\
&\qquad + ||\psi_L||_2 \cdot \left\| f_2(\psi_N) - \sum_{i=1}^{N} \beta_i w_i(\psi_N) \right\|_2
\end{aligned} \tag{2.19}
$$

The first term in this equation can be made arbitrarily small, cf. Corollary 1 with $p = 1$ since $\hat{f}_{Ni}$ is linear and $\Psi_N$ is compact. The second term can be made arbitrarily small by the same corollary with $p = 0$ through the choice of $\beta_i$, since $\Psi_L$ is compact. Hence, for any $\epsilon > 0$ we can make $||f(\psi) - \hat{f}(\psi)||_2 < \epsilon$. Since $\psi$ is arbitrary, we get $||f - \hat{f}||_\infty < \epsilon$.

□

It is clear that the curse of dimensionality is not avoided, but the reduction in dimension reduces this problem considerably for a significant number of modeling problems, as we shall illustrate with some examples in later chapters.

## Example 3

If the system is of the affine form

$$
\begin{aligned}
y(t) &= f(y(t-1), u(t-1)) + e(t) \\
&= f_1(y(t-1)) + f_2(y(t-1))u(t-1) + e(t)
\end{aligned}
$$

the choice $z(t) = y(t)$ does not restrict the attainable prediction accuracy using local ARX models.

□

This result can be generalized to local expansions of (polynomial) order $p$. We can always decompose the set $\Psi$ into two subsets $\Psi_L$ and $\Psi_H$ such that $\Psi_L \times \Psi_H = \Psi$, and

$$
f(\psi) = f(\psi_L, \psi_H) = f_{H1}(\psi_H) + f_{H2}(\psi_H) f_L(\psi_L) \tag{2.20}
$$

where $f_L : \Psi_L \to R^n$ is of polynomial order no larger than $p$, and $f_{H1} : \Psi_H \to R^m$ and $f_{H2} : \Psi_H \to R^{m \times n}$ may be of higher order.

**Theorem 5** *Suppose $f$ given by (2.20) is continuous, and $\Psi$ is compact. Then for any $\epsilon > 0$ there exists an $\hat{f} \in \mathcal{F}_p(\mathcal{K})$ with $z = H(\psi) = \psi_H$ such that $||f - \hat{f}||_\infty < \epsilon$.*

**Proof:** The proof follows the same idea as the proof of Theorem 4, but requires some tedious notation, and is therefore omitted.

□

### 2.1.5 Some Comparisons

Using local linear model structures we can write the model representation (2.17) as

$$\begin{aligned}
\hat{f}(\psi(t-1)) &= \sum_{i=1}^{N}(\theta_i + \Theta_i \psi(t-1)) w_i(z(t-1)) \\
&= \sum_{i=1}^{N} \theta_i w_i(z(t-1)) + \left( \sum_{i=1}^{N} \Theta_i w_i(z(t-1)) \right) \psi(t-1) \\
&= \theta(z(t-1)) + \Theta(z(t-1))\psi(t-1)
\end{aligned}$$

This means that the non-linear model structure can be written as a quasi-linear model structure, where the parameters are dependent on the operating point. Priestley (1988) introduced *State-dependent models* which are of the form

$$y(t) = \theta(x(t-1)) + \Theta(x(t-1))\psi(t-1) + e(t) \tag{2.21}$$

where $x$ is the state-vector, $\theta$ is a state-dependent vector, and $\Theta$ is a state-dependent matrix. In general, the canonical state-space $x = \psi$ was suggested, but it was also observed that this might be redundant, so a simpler vector may be used to describe the parameter dependence. The present approach with $x = z$ has obvious similarities, and offers one particular representation for the state-dependent parameters. Billings and Voon (1987) discuss the use of models with *signal-dependent parameters*, which are similar to (2.21) this $x = \omega$, where $\omega(t)$ is the auxiliary signal. Polynomials are used to define the dependence of the parameters on the auxiliary signal, i.e. $\theta(\omega(t))$ and $\Theta(\omega(t))$ are polynomials in $\omega(t)$. Our approach is also similar, but prior knowledge is applied to design the $\rho_i$-functions, which again defines $\theta(z(t))$ and $\Theta(z(t))$. The *Threshold AR Model* (TAR) of Tong and Lim (1980) can also be written in the form (2.21) with $x(t-1) = y(t-1)$ and

$$\begin{aligned}
\theta(y(t-1)) &= \begin{cases} \theta_1, & \text{if } y(t-1) \in Y_1 \\ \theta_2, & \text{if } y(t-1) \in Y_2 \end{cases} \\
\Theta(y(t-1)) &= \begin{cases} \Theta_1, & \text{if } y(t-1) \in Y_1 \\ \Theta_2, & \text{if } y(t-1) \in Y_2 \end{cases}
\end{aligned}$$

where $Y = Y_1 \cup Y_2$ and $Y_1 \cap Y_2 = \emptyset$. The parameter values are chosen from a set of two possible parameter values, where the decision is based on the value of $y(t-1)$. The resulting model is a piecewise linear model and related to our approach if $z(t-1) = y(t-1)$ and the interpolation functions are step-functions. While the functional forms above may appear very similar, we would like to stress that the operating regime based design of the $\rho_i$-functions appears to be considerably more transparent and attractive from an engineering point of view than the other approaches.

Takagi and Sugeno (1985) suggested a fuzzy set (Zadeh 1965) based technique for combining in a smooth fashion a set of linear models into a non-linear model. It turns out that if the operating regimes $Z_i$ are viewed as fuzzy sets with membership functions $\mu_{Z_i}(z)$ equal to the local model validity functions $\rho_i(z)$, then fuzzy inference on a rule-base of the form

$$\text{IF} \quad z(t-1) \text{ IS } Z_i \quad \text{THEN} \quad y(t) = \hat{f}_i(\psi(t-1)) + e(t)$$

results in a model of the form (2.4)-(2.5). This suggests the use of fuzzy sets and rules as a means of designing the operating regimes. This is appealing since it is natural to interpret operating regimes as sets with soft boundaries. The fuzzy set theory provides a powerful representation of operating regimes through the following definitions of union, intersection, and complement

$$\begin{aligned}
\mu_{Z_a \cup Z_b}(z) &= \max(\mu_{Z_a}(z), \mu_{Z_b}(z)) \\
\mu_{Z_a \cap Z_b}(z) &= \min(\mu_{Z_a}(z), \mu_{Z_b}(z)) \\
\mu_{Z_a^C}(z) &= 1 - \mu_{Z_a}(z)
\end{aligned}$$

where $Z_a$ and $Z_b$ are fuzzy sets. For example, a fuzzy set $Z_i$ defined by

$$z \text{ IS } Z_i = (z_1 \text{ IS LARGE}) \text{ AND } (z_3 \text{ IS NOT INCREASING})$$

has the membership function

$$\mu_{Z_i}(z) = \min\left(\mu_{z_{1,\text{large}}}(z_1), 1 - \mu_{\dot{z}_{3,\text{incr}}}(\dot{z}_3)\right)$$

where $\mu_{z_{1,\text{large}}}(z_1)$ is the mono-variable membership function for the fuzzy subset where "$z_1$ is large" and $\mu_{\dot{z}_{3,\text{incr}}}(\dot{z}_3)$ is the mono-variable membership function that represent the fuzzy subset where "$z_3$ is increasing". In particularly for complex high-dimensional modeling problems, this representation may be useful for structuring the operating regime knowledge since multi-variable membership-functions are constructed from mono-variable ones. However, in order to keep the presentation as simple as possible, we have chosen not to focus on the fuzzy set interpretation in this thesis.

We have also chosen to suppress the neural network interpretation of the model representation (Stokbro et al. 1990, Jones and co-workers 1991, Jacobs et al. 1991). The reason is simply that such an interpretation does not appear to contribute with any advantages that counterweight the possible confusion it may introduce, in our view.

A constant local model structure is closely related to an interpolating memory (Albus 1972, Lane, Handelman and Gelfand 1992, Cybenko, Saarinen, Gray, Wu and Khrabrov 1994). Since just the constant parameter value $\theta_i$ is extrapolated locally, each local model will have a small range of validity, and a large number of operating regimes is needed. Moreover, the curse of dimensionality will be more apparent, since the dimension of the operating point cannot be reduced, and must equal the dimension of the information vector. The constant local model case is functionally identical to neural networks with localized and normalized receptive fields, e.g. (Moody and Darken 1989), and some fuzzy models, e.g. (Wang and

Mendel 1992). Considering $\{\tilde{w}_i\}_{i=1}^N$ as a set of basis-functions, the method is also similar to radial basis-function (RBF) expansions (Powell 1987, Broomhead and Lowe 1988, Dyn 1989). Using RBF's, a non-linear function is modeled with the series expansion

$$\hat{f}(\psi) = \sum_{i=1}^{N} \theta_i r_i(||\psi - \psi_i||_2)$$

where $r_i : R^+ \rightarrow R$ is typically chosen as a thin plate spline, multi-quadric, or Gaussian function. If the functions $\tilde{\rho}_i$ are chosen as local radial functions, the normalized function $\tilde{w}_i$ defined by (2.5) will not be radial in general, but it will qualitatively have much the same shape and features as $\tilde{\rho}_i$, except near the boundary of $\Psi$. The effect of normalization in discussed in (Shorten and Murray-Smith 1994). Another class of related model representations can be found in the statistical smoothing literature, e.g. (Hastie and Tibshirani 1990).

## Example 4

We consider again the very simple function in Example 2, and the following four modeling approaches:

1. Two first order truncated Taylor-expansions at $\psi_1 = 0.5$ and $\psi_2 = 1.5$. These are combined into a piecewise linear one, as in Example 2. This may also be interpreted as a Thresholded AR model.
2. We choose Gaussian model validity functions
$$\rho_i(z) = \exp\left(-\frac{1}{2}\left(\frac{z - z_i}{\sigma_i}\right)^2\right) \tag{2.22}$$
with $z_1 = \psi_1 = 0.5, z_2 = \psi_2 = 1.5, \sigma_i = 0.5$, and use two linear local model structures.
3. Five constant local model structures centered at $\psi_1 = 0, \psi_2 = 0.5, \psi_3 = 1, \psi_4 = 1.5, \psi_5 = 2$, and Gaussian model validity functions with $\sigma_i = 0.5$,
4. A radial basis-function expansion with five Gaussian basis-functions centered at $\psi_1 = 0, \psi_2 = 0.5, \psi_3 = 1, \psi_4 = 1.5, \psi_5 = 2$, and $\sigma_i = 0.5$.

Linear regression is used to estimate the model parameters on the basis of a uniformly distributed data set containing 100 samples. The results are shown in Figure 2.4. Comparing Figure 2.4a with 2.4b, it appears that interpolating local linear models with estimated parameters and well chosen model validity functions can improve the accuracy compared to the piecewise linear model with truncated Taylor-expansions locally.

Notice that $f$ in this example is defined on the interval $[-1, 3]$, while data from the interval $[0, 2]$ is used for parameter estimation. The extrapolation capabilities can thus be evaluated, and we see that the model based on local linear models gives first order extrapolation, as would be expected, while the constant local models give zeroth order extrapolation. The RBF approach does not given any sensible extrapolation at all, since all basis-functions go to zero. It is interesting to observe that while all methods give roughly comparable results in the region where there are data available, there are fundamental differences concerning the extrapolation capabilities.

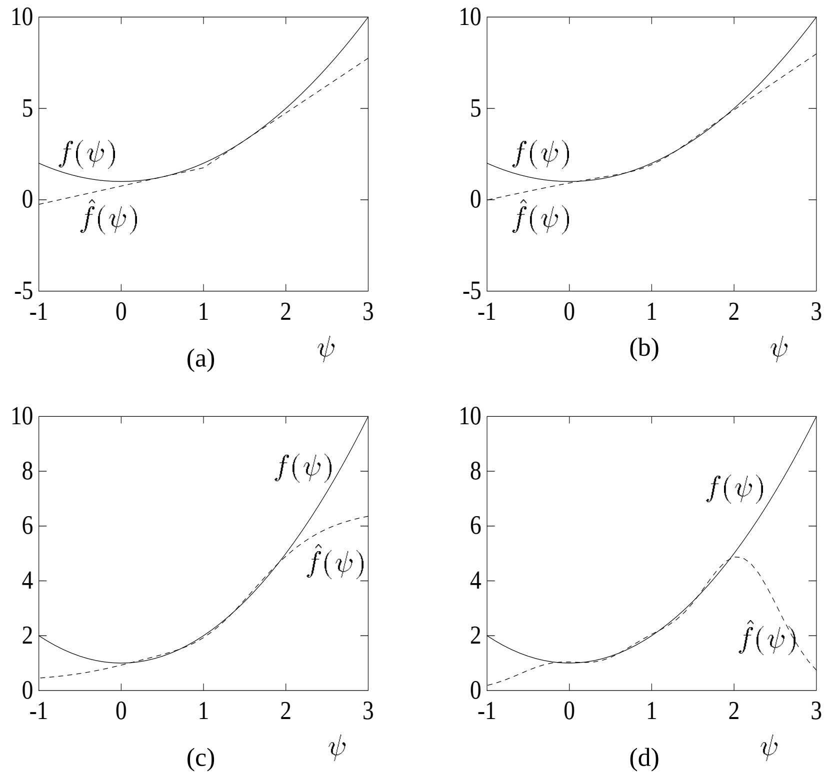


Figure 2.4: a) Approximation of $f(\psi) = \psi^2 + 1$ using a piecewise linear model with two local linear models. b) Model based on two local linear models and Gaussian model validity functions. c) Model based on five constant local models. d) Model based on a radial basis-function expansion with five Gaussian basis-functions.

□

## 2.2 Modeling

Within the framework described here, the development of a model typically consists of the following steps

1. Design experiments and collect data.
2. Choose an operating point set $Z$, including a choice of which variables to characterize the system's operating conditions with.

3. Decompose $Z$ into regimes and find an adequate local model structure and a local model validity function for each regime.
4. Identify the unknown local model parameters (and any unknown local model validity function parameters).

Prior knowledge can be applied in all these steps, and is particularly important in the second and third steps. Hence, this procedure should by no means be looked upon as a black-box approach, because the model representation is transparent in the sense that it allows easy interpretation and incorporation of prior knowledge. The transparency is related to the possibility to relate the operating regime decomposition to the system mechanisms or high-level behavior. Moreover, it is related to the transparency of each local model, which is typically good, because the local model structures are chosen to be simple.

Selection of the operating point set includes a specification of which variables the operating regimes are characterized with. As we have seen in Theorem 4, with local linear model structures it is important that the operating point contains variables that captures that non-linearities. To get an accurate and transparent model, it is important that prior knowledge is applied to define the operating point using as few and interpretable variables as possible.

The decomposition into regimes is typically accomplished with the help of empirical or mechanistic prior knowledge about the behavior of the system under different operating conditions.

The choice of local model structures and local model validity functions is closely related to the decomposition into operating regimes, as the operating regimes must be designed with a particular local model structure in mind, or vice versa. It is evident that there is a trade-off between low dimension of the operating point vector and the simplicity of the local models. At one extreme, one may have only one complex local model which will also be the global model. At the other extreme, one may need a large number of very simple local models, like constant ones, and a high-dimensional operating point vector. This gives a significant amount of flexibility with this approach. Like with ARMAX models, one must specify the local model structure order parameters, and analyze the structure of the noise and disturbances in order to select the MA part of the local model structures.

It must be emphasized that there does not exist a unique best model structure with this approach, and that there is significant flexibility regarding the different structural choices, like decomposition into regimes, which variables to characterize the operating regimes with, and which local model structures to choose. Together with its simplicity and transparency, this flexibility is a major reason for the power of this procedure.

## 2.3 Identification

First we consider identification of the local model parameters based on local criteria for each local model. Then we discuss identification on the basis of a global criterion, before we briefly discuss identification of local model validity function parameters.

### 2.3.1 Identifying Local Model Parameters using Local Criteria

The prediction error at time $t$ using the local model $\hat{f}_i$ is

$$\begin{aligned}\varepsilon_i(t) &= y(t) - \hat{f}_i(y(t-1), ..., y(t-n_y), \\ &\qquad u(t-1), ..., u(t-n_u), \varepsilon_i(t-1), ..., \varepsilon_i(t-n_e))\end{aligned} \tag{2.23}$$

The rationale behind this predictor is that if the model matches the true system, then $\varepsilon_i(t) \to e(t)$ as $t \to \infty$ provided the system (2.23) is asymptotically stable, and the initial values are within the region of attraction (Granger and Andersen 1978). A local criterion $J_i$ associated with each local model can be written as

$$J_i = \frac{1}{l}\sum_{t=1}^{l} \varepsilon_i^T(t)\Lambda_i(t)\varepsilon_i(t)$$

where $l$ is the number of observations available, and $\Lambda_i(t)$ is a weight-matrix that in addition to relative scaling of the prediction errors has the purpose of weighting only the observations that are relevant for local model $i$. A typical choice is

$$\Lambda_i(t) = \Lambda\rho_i(z(t-1))$$

where $\Lambda$ is a constant scaling-matrix (Murray-Smith 1994*b*). Consider the local ARMAX model structure (2.10). The local model structure is parameterized by a vector $\theta_i$ and a matrix $\Theta_i$. If all local model structures are linear functions of the parameters, like (2.10), the global model structure is linear in the parameters, and standard identification methods can be applied, e.g. (Söderström and Stoica 1988).

Assume first the information vector $\psi(t-1)$ does not contain any noise terms $e(t-1), ..., e(t-n_e)$. The predictor can now be written on the linear regression form

$$\hat{f}_i(\psi(t-1)) = \varphi_i^T(t-1)\overline{\theta}_i$$

where $\overline{\theta}_i$ is a parameter vector and $\varphi_i(t-1)$ is a regression matrix. The parameters can be estimated using the least squares (LS) method:

$$\hat{\overline{\theta}}_i = \left(\frac{1}{l}\sum_{t=1}^{l}\varphi_i(t)\Lambda_i(t)\varphi_i^T(t)\right)^{-1}\left(\frac{1}{l}\sum_{t=1}^{l}\varphi_i(t)\Lambda_i(t)y(t)\right) \tag{2.24}$$

In the general case when delayed noise, $e(t-1), e(t-2), ..., e(t-n_e)$, is included in $\psi(t)$, we use a prediction error (PE) method. Since these terms are replaced with corresponding $\varepsilon_i$-terms, the prediction error (2.23) is no longer a linear function of the parameters. Hence it is not possible to find an explicit solution like (2.24). The criterion (2.24) must be minimized numerically using e.g. the Newton-Raphson algorithm

$$\hat{\overline{\theta}}_i^{(k+1)} = \hat{\overline{\theta}}_i^{(k)} - \alpha_k\left(\nabla^2 J_i\left(\hat{\overline{\theta}}_i^{(k)}\right)\right)^{-1}\nabla J_i\left(\hat{\overline{\theta}}_i^{(k)}\right)$$

### 2.3.2 Identifying Local Model Parameters using a Global Criterion

We define the global prediction error to be

$$\begin{aligned}\varepsilon(t) \;=\; & y(t) - \hat{f}(y(t-1), ..., y(t-n_y), \\ & u(t-1), ..., u(t-n_u), \varepsilon(t-1), ..., \varepsilon(t-n_e))\end{aligned}$$

and a global identification criterion is

$$J = \frac{1}{l} \sum_{t=1}^{l} \varepsilon^T(t) \Lambda \varepsilon(t) \tag{2.25}$$

The LS- and PE-methods can be formulated in the same manner as above. Our experience is that in most cases the use of a global criterion is superior to the use of local criteria, in terms of prediction performance. However, an exception is when the model structure is overparameterized. Then it can be shown that the use of local criteria has a regularizing effect (Murray-Smith 1994*b*) that reduces the effective number of parameters somewhat, and therefore gives better prediction performance. The amount of regularization can be tuned with the overlap between the $\rho_i$-functions.

### 2.3.3 Identifying Model Validity Function Parameters

In general, the local model validity function parameters will enter the equations for the prediction error non-linearly. In particular, if these parameters are to be identified simultaneously with the local model parameters, we get a complex non-linear programming problem. We will not discuss this problem here, but refer to the vast literature on general-purpose non-linear programming algorithms, e.g. (Gill, Murray and Wright 1981).

We will like to point out that our experiments indicate that a rough choice of the location and scaling parameters in the model validity functions will in most cases be almost as accurate and at least as robust as the identification of local model validity function parameters when a global identification criterion is applied. We therefore recommend that the local model validity function parameters are not optimized using empirical data, but chosen with the aid of prior knowledge, whenever available. With local identification criteria, the exact choice of local model validity functions becomes more critical, and a rough empirical choice may not be sufficient.

## 2.4 Experimental Results

In this section we will present the results of some rather simple experiments that illustrate the modeling procedure, (Johansen and Foss 1995).

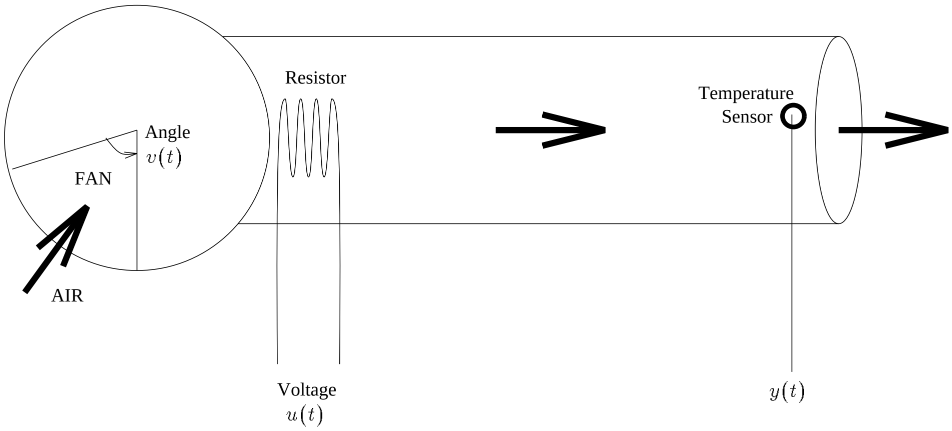


Figure 2.5: A heat transfer process: Air is pulled into the 30 cm tube through a valve by a fan. The air is heated by a resistor, and its temperature is measured at the outlet.

Consider the experimental setup illustrated in Fig. 2.5. The output of this system is the air temperature measurement $y(t)$, the input is the voltage over the resistor $u(t) \in [0, 10]$. In addition, the valve opening angle $v(t) \in [0°, 180°]$ is an independent measured variable that can be manipulated by hand only. On this system we have performed some experiments. First, consider the responses to a 10 volt step input, for $v_1 = 20°$, $v_2 = 50°$, and $v_3 = 100°$ plotted in Fig. 2.6a. From these curves, we make two observations:

1. First, there seems to be two dominating time-constants, at least for small valve opening angles $v$. The fast mode (time constant about 1 second) is related to the heat capacity of the air in the tube, while the slower one (time constant equal to a few minutes) is related to the heat capacity of the tube itself and the rest of the equipment. In the proceeding modeling and identification experiments, we will only attempt to find a model with good prediction performance on the horizon of the fastest of these modes, since this is the one that is interesting for control purposes.

2. Second, the steady-state gain seems to be a function of the valve opening angle $v$. On a shorter time-scale, cf. Fig 2.6b, we see that also the time-constant and time-delay are functions of $v$. Similar experiments with fixed $v$ but varying input step amplitude shows that the steady-state gain depends on the input signal level, too, cf. the steady-state response in Fig. 2.7.

For the purpose of identification, we use the data sequence in Fig 2.8. The sampling interval is $\Delta t = 0.11$ s, and the sequence contains about 1300 samples. The input $u(t)$ is a random signal exciting the dynamics locally about a sequence of random operating points that covers the full range of operation. The valve opening $v(t)$ varies over the full range of operation in a pseudo-random manner. For the purpose of model validation, we use another data sequence with somewhat different

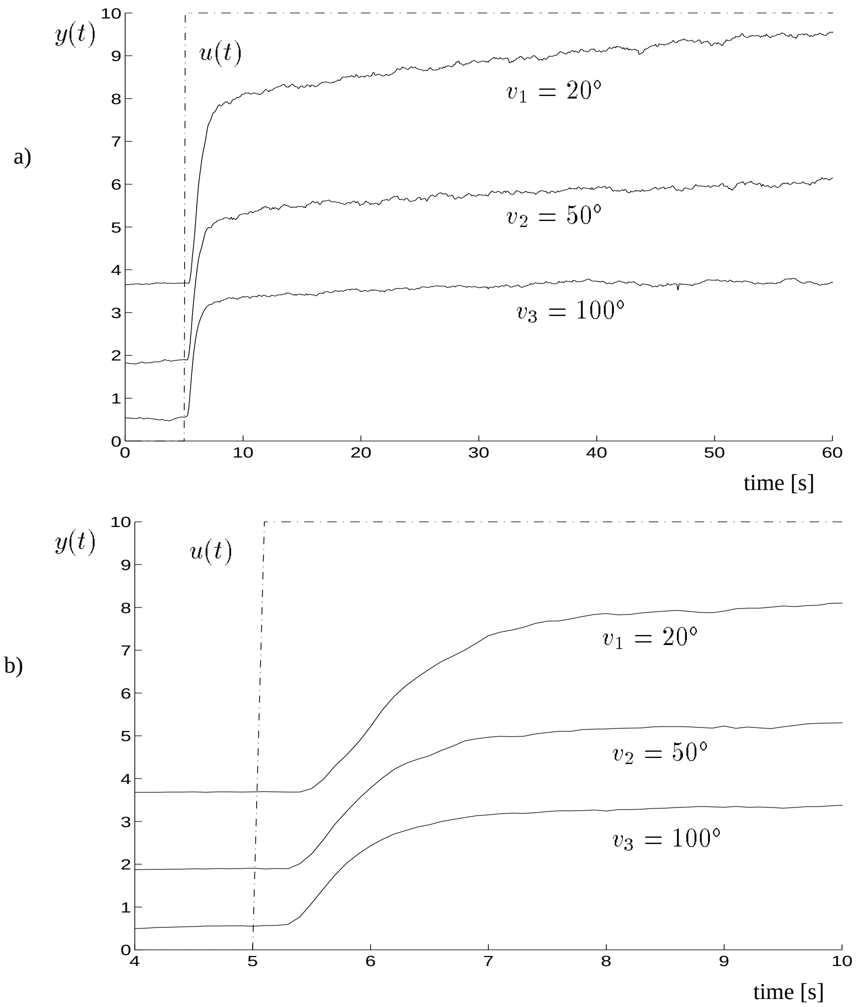


Figure 2.6: a) Response to a 10 volt step input, and b) the same response on a shorter time-scale.

excitation signals, see Fig. 2.9. The input varies in a qualitatively similar manner, while the valve opening angle varies more or less systematically to cover a wide range of frequencies and amplitude levels.

### 2.4.1 Identification of ARX Model

Using the identification data sequence and the least squares algorithm, we identify the following first-order ARX model

$$\begin{aligned} y(t) \,=\, & 0.2247 + 0.9284y(t-1) + 0.0352u(t-4) \\ & -1.286 \cdot 10^{-3}v(t-5) + e(t) \end{aligned} \tag{2.26}$$

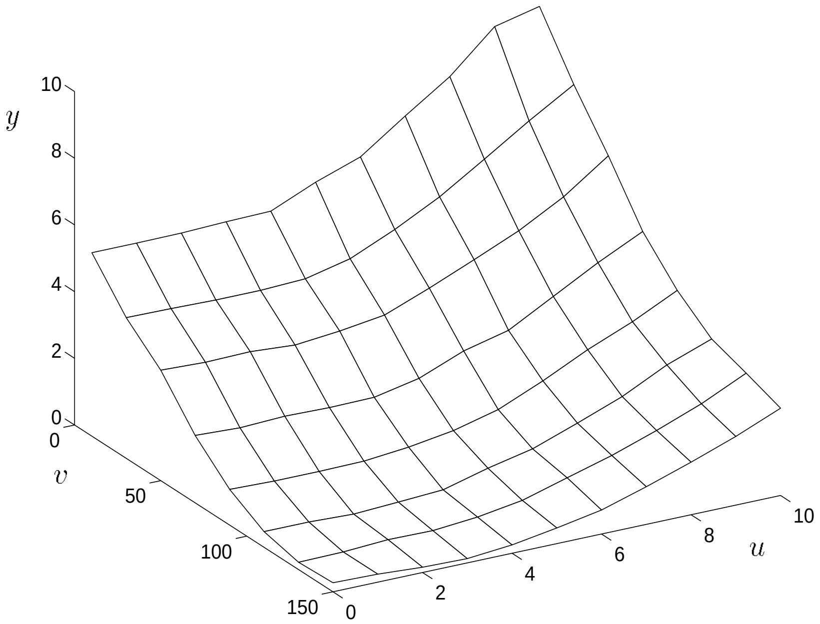


Figure 2.7: Steady-state response.

The order and time-delays are determined from the step-responses, cf. Fig. 2.6b.

### 2.4.2 Identification of Operating Regime based NARX Model

From the discussion above, it is clear that the coefficients of the ARX model (2.26) should be functions of at least $u$ and $v$. On the basis of the step-responses and the steady-state gain characteristic, cf. Figs. 2.6 and 2.7, we therefore try to combine four local ARX model structures into an NARX model structure, as outlined in section 2.2. The operating point is $z(t-1) = (u(t-4), v(t-5))^T$. The four local ARX models have local model validity functions

$$\rho_1(u,v) = \exp\left(-\frac{1}{2}\left(\left(\frac{u-3}{2}\right)^2 + \left(\frac{v-20^\circ}{40^\circ}\right)^2\right)\right)$$
$$\rho_2(u,v) = \exp\left(-\frac{1}{2}\left(\left(\frac{u-8}{2}\right)^2 + \left(\frac{v-20^\circ}{40^\circ}\right)^2\right)\right)$$
$$\rho_3(u,v) = \exp\left(-\frac{1}{2}\left(\left(\frac{u-3}{2}\right)^2 + \left(\frac{v-130^\circ}{40^\circ}\right)^2\right)\right)$$
$$\rho_4(u,v) = \exp\left(-\frac{1}{2}\left(\left(\frac{u-8}{2}\right)^2 + \left(\frac{v-130^\circ}{40^\circ}\right)^2\right)\right)$$

The interpolation functions are illustrated in Fig. 2.10. The local model valid-

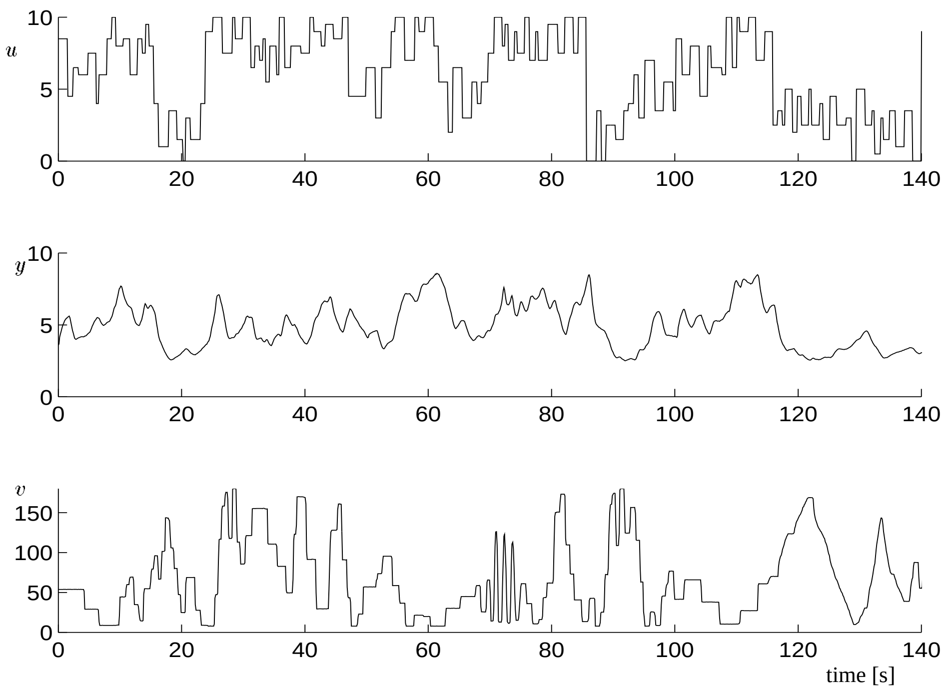


Figure 2.8: Data sequence used for identification.

ity functions are chosen to give a reasonable amount of overlap. The motivation behind this particular decomposition is that we believe a linear ARX model structure gives an adequate description of the system within each regime. Using the least squares algorithm based on the criterion that penalizes mismatch between the global model and the data gives the model

$$\begin{aligned} y(t) \;=\; & (0.3338 + 0.9196y(t-1) + 0.0049u(t-4) - 1.411\cdot 10^{-3}v(t-5)) \\ & \qquad\qquad \cdot w_1(u(t-4), v(t-5)) \\ & +(0.2029 + 0.9368y(t-1) + 0.0397u(t-4) - 2.389\cdot 10^{-3}v(t-5)) \\ & \qquad\qquad \cdot w_2(u(t-4), v(t-5)) \\ & +(0.5433 + 0.8347y(t-1) + 0.0139u(t-3) - 0.946\cdot 10^{-3}v(t-5)) \\ & \qquad\qquad \cdot w_3(u(t-4), v(t-5)) \\ & +(0.5181 + 0.8601y(t-1) + 0.0522u(t-3) - 2.135\cdot 10^{-3}v(t-5)) \\ & \qquad\qquad \cdot w_4(u(t-4), v(t-5)) + e(t) \end{aligned} \tag{2.27}$$

which we will call "the 4-local model". The time-delays are chosen to be smaller at large valve opening angle, cf. Fig 2.6b. We see that the model has captured the facts that the dynamics is faster at large valve-opening, and that the gain is larger at high input voltages.

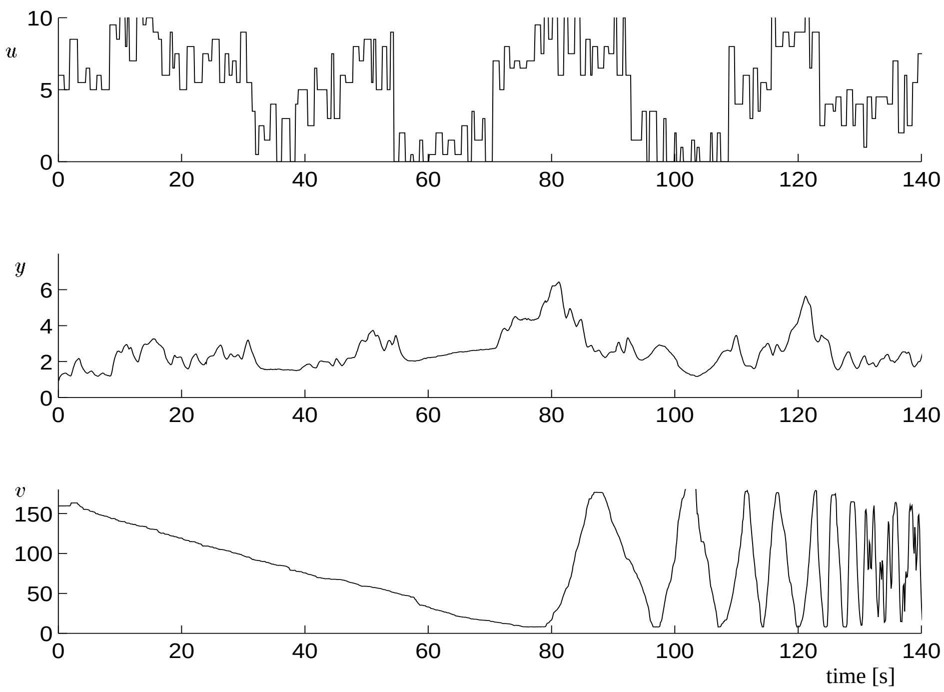


Figure 2.9: Data sequence used for validation.

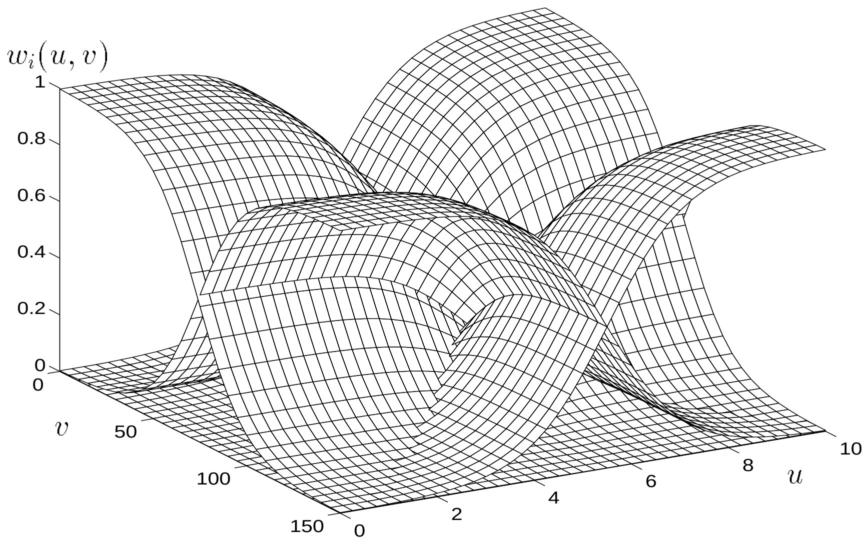


Figure 2.10: Interpolation functions for the local models.

| Symbol | Variable | Value | Unit |
|---|---|---|---|
| $v$ | Valve opening | | $^\circ$ |
| $\rho$ | Mass density for air | 1.2 | g/l |
| $V$ | Volume of tube | 0.7 | l |
| $c_p$ | Specific heat capacity for air | 1.0 | J/g K |
| $q$ | Volumetric air flow rate | | l/s |
| $k$ | "Fan coefficient" | | l/s |
| $G$ | Conductance in resistor | | $\Omega^{-1}$ |
| $U$ | Effective heat transfer coefficient | | J/K s |
| $Q$ | Heat added through fan | | W |
| $T$ | Air temperature in tube | | °C |
| $T_0$ | Air temperature in environment | 24 | °C |
| $T_1$ | Temperature in equipment | 28 | °C |
| $\tau_v$ | Time-delay, valve opening to output | 0.44 | s |
| $\tau_u$ | Time-delay, input to output | 0.33 | s |

Table 2.1: Symbols, constants and variables used in the mechanistic model.

### 2.4.3 Identification of a Semi-mechanistic Model

A simple energy balance for this system is

$$\begin{aligned}\rho V c_p \frac{d}{dt} T(t) &= \rho c_p q(v(t-\tau_v))(T_0 - T(t)) + U(T_1 - T(t)) \\ &\quad + G u^2(t-\tau_u) + Q \qquad (2.28)\end{aligned}$$

where the symbols are defined in Table 2.1. The first term in (2.28) is the energy lost because the outlet temperature is higher than the inlet temperature. The second term is the energy lost through the tube walls, while the third and fourth terms are heat added by the power source and fan, respectively. The volumetric air flow rate $q$ is a function of the valve opening angle $v$ and the fluid dynamics in the fan housing. We choose the empirical correlation $q = kv$, where $k$ is an unknown empirical constant, since a higher order correlation does not appear to give a model with significantly better prediction performance. In addition, we need the calibrated measurement equation

$$y(t) = 0.8173(T(t) - 26.6)$$

where $y(t)$ is the output voltage. Again, we identify the unknown physical parameters using the least squares algorithm, which gives

$$\begin{aligned}\hat{G} &= 0.0318 \ \ \Omega^{-1} \\ \hat{U} &= 0.5616 \ \ J/Ks \\ \hat{k} &= 1.4920 \cdot 10^{-3} \ \ l/s \\ \hat{Q} &= 2.0346 \ \ W\end{aligned}$$

Notice that this model is most correctly interpreted as a semi-mechanistic model. For example, an error in $T_0$ or $T_1$ will strongly influence the estimate $\hat{Q}$, which therefore has lost its physical interpretation as the heat added by the fan. Discretizing the model (2.28), we get

$$\begin{aligned} T(t) &= T(t-1) + \Delta t \left( \frac{kv(t-5)}{V}(T_0 - T(t-1)) \right. \\ &\quad \left. + \frac{U}{\rho V c_p}(T_1 - T(t-1)) + \frac{G}{\rho V c_p} u^2(t-4) + \frac{Q}{\rho V c_p} \right) \end{aligned}$$

We may re-parameterize this model to a semi-mechanistic NARX model on the form

$$\begin{aligned} y(t) &= 0.3028 + 0.9265 y(t-1) + 0.540 \cdot 10^{-3} v(t-5) \\ &\quad -0.234 \cdot 10^{-3} y(t-1) v(t-5) + 2.401 \cdot 10^{-3} u^2(t-4) + e(t) \end{aligned}$$

Lindskog and Ljung (1994) have discussed the design of semi-mechanistic models of this form.

### 2.4.4 Discussion of Identification Results

The prediction performance of the three identified models are compared by simulating the models' responses to the validation data input sequence, cf. Fig. 2.11. It is evident that both the empirical 4-local and semi-mechanistic models perform better than the ARX model, as one should expect. The semi-mechanistic model predicts slightly better than the 4-local model, but the difference is not significant. Notice that because the validation data were logged on a day with higher environment temperature, there was a significant, but constant, offset between the output of the simulations and the measured output. As this is easily compensated for with integral action in a controller, this offset was removed in Fig. 2.11 to simplify the comparison.

| Model | 1-step-ahead prediction error | 10-steps-ahead prediction error |
|---|---|---|
| ARX | 0.0788 | 0.4669 |
| 4-local | 0.0787 | 0.3447 |
| Semi-mechanistic | 0.0654 | 0.3182 |

Table 2.2: Root average squared prediction errors estimated with the validation data sequence.

The average 1-step-ahead and 10-steps-ahead prediction errors are estimated using the validation data sequence for the three models. The results are shown in Table 2.2. Again, we see that the semi-mechanistic model performs somewhat better than the 4-local model. The 10-steps-ahead prediction errors are illustrated in Fig. 2.12. They reveal that there is a significant amount of unmodeled dynamics left. This may be due to the slow time-constant we have neglected, and other effects that are not adequately modeled. In particular, there seem to be some slowly time-varying bias that might be adequately modeled as integrated white

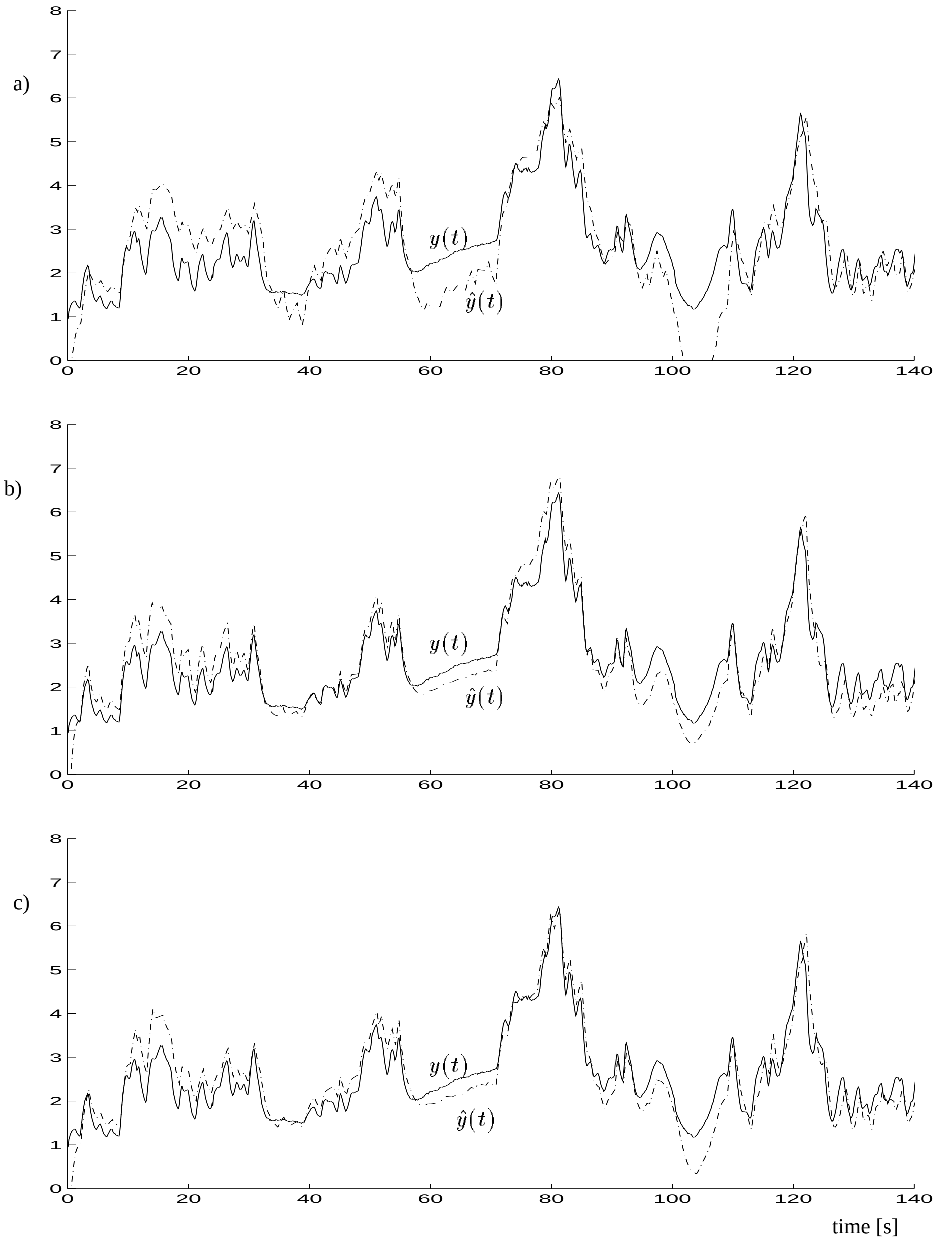


Figure 2.11: Simulation of the identified models' responses to the validation data input sequence. a) ARX model, b) 4-local model, and c) semi-mechanistic model.

noise. It is interesting to observe that the prediction error with the 4-local model consists of more bias but less variance than both the ARX and the mechanistic models. Further experiments has shown that the introduction of a noise model

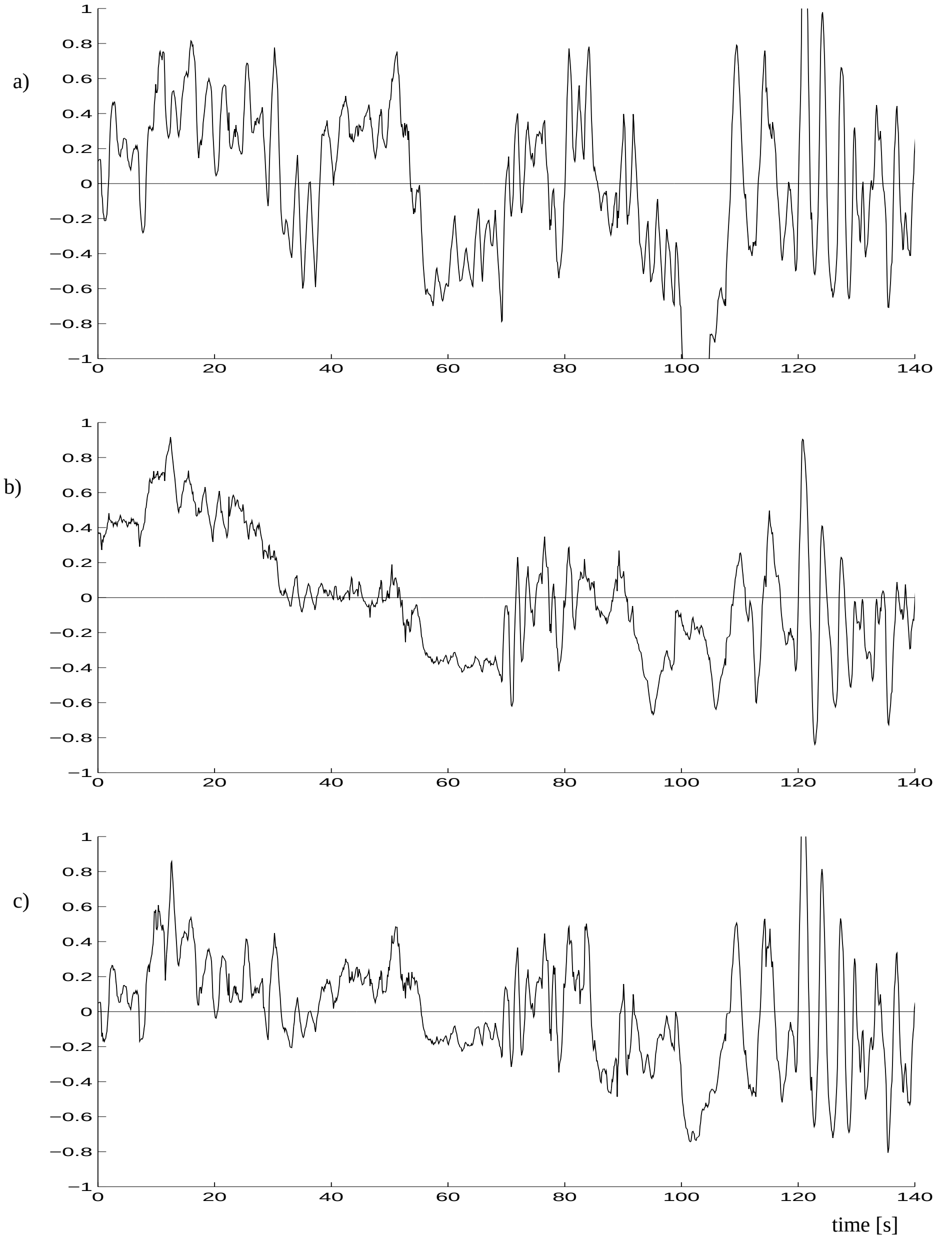


Figure 2.12: 10-steps-ahead prediction errors for the 3 models' responses to the validation data input sequence, a) with ARX model, b) with the 4-local model, and c) with the semi-mechanistic model.

that tries to incorporate the slowly varying bias in all three model alternatives gives most significant improvements for the 4-local model. However, the reason why the 4-local model has less variance is unclear, but we find it unlikely that this is a feature of the operating regime based modeling framework.

The prior knowledge applied to define the structure of the 4-local model is entirely empirical and gained by examining process data. Hence, the 4-local model is entirely empirical, but certainly not a black box model. The regimes can be linked to qualitative different behaviors of the system. For example, in Regime 1, the gain is small and the dynamics are slow, while in Regime 4, the gain is higher, and the dynamics are faster.

We have chosen local ARX models. Clearly, this is not the only possible choice. We might for example have expected a $u^2$-term since it is obvious that it is the power and not the voltage that is important in the energy balance. This would give local NARX model structures, and would have allowed us to reduce the dimension of the operating point and the number of local models.

## 2.5 Discussion

To summarize, we have investigated how non-linear systems can be modeled using NARMAX models based on local ARMAX models. We have described a practical modeling procedure, which is supported with some theoretical results and experiments.

The main theoretical results are

1. The interpolation method used for combining a number of local models is optimal, in a certain sense. While this is a nice result, one should not put too much emphasis on it, since the optimality is obviously a result of the particular choice of objective criterion. On the other hand, the chosen objective criterion captures the desired properties of the interpolation method.

2. The system function can be uniformly approximated to an arbitrary accuracy on any compact operating range by making the decomposition into operating regimes sufficiently fine. The proof is constructive and gives detail on how the regimes, local models, and local model validity functions may be chosen. Moreover, we have given an upper bound on the approximation accuracy when the model is constructed using this procedure. However, the constructive procedure is sub-optimal and is not recommended in practice. It is important to remember that these results are approximation results that assumes the "true function $f$" is known. In practise, the model is estimated on the basis of empirical data, and noise, the quality and amount of the empirical data available will limit the attainable accuracy. Hence, the approximation results are not directly applicable, but provide nevertheless a useful basis for understanding the various design trade-offs that must be made in the practical modeling and identification procedure. In the literature, there exists a wide range of results similar to Corollary 1 for other model representations, e.g. (Cybenko 1989, Barron 1993, Parzen 1962, Park

and Sandberg 1993). Related results on the relations between the complexity of the system, the approximation bound, and the complexity of a piecewise linear model are given in (Omohundro 1987).

3. With local linear model structures, the operating regimes should be characterized using variables that capture the system's non-linearities. This is a somewhat trivial result, but is still perhaps our best guide for the choice of operating point vector.

The experimental results has the purpose of illustrating the operating regime based modeling procedure. We have compared the results of this procedure with standard approaches, namely an ARX model and a simple mechanistic model. The operating regime based modeling approach gave a transparent model that was clearly superior to the ARX model and comparable to the semi-mechanistic model.

The problem of choosing the local model order parameters $n_y, n_u$, and $n_e$ has not been discussed. The reason is certainly not that this is a trivial problem. The general problem of order selection is widely discussed in the system identification literature, and there exists several statistical criteria and general guidelines, e.g. (Box and Jenkins 1970, Ljung 1987), see also Chapter 6. In the operating regime based modeling framework, there may be a need for different order of the different local models. Sbarbaro (1994) has suggested to represent continuous-time local linear models with Laguerre networks in order to reduce the sensitivity with respect to the choice of order.

The ARMAX model representation is perhaps the most widely used linear empirical dynamic model representation. For the NARMAX model representation, there exists a wide range of generic representations for the function $f$. Neural networks have over the past few years gained considerable popularity, at least in academia. The most popular feed-forward type networks can be used to build black box NARMAX models, typical examples can be found in (Chen, Billings and Grant 1990*a*, Nguyen and Widrow 1990, Sjöberg, Hjalmarsson and Ljung 1994). Common to most types of neural networks is that the internal representation is very complex and far from transparent, which may explain why so few industrial applications are reported. Moreover, from an identification point of view, the non-linear parameterization of many neural network based model representations is a serious drawback, in particular as long as there exists powerful linearly parameterized alternative model structures. Recently, there has been some interest in the application a prior knowledge for structuring neural nets (Mavrovouniotis and Chang 1992), initialization of neural network parameters and interpretation of the resulting model through linearizations (Scott and Ray 1993). A further step was taken in (Kramer et al. 1992, Thompson and Kramer 1994, Su et al. 1992, Psichogios and Ungar 1992, Aoyama and Venkatasubramanian 1993, Brown, Ruchti and Feng 1993) where certain combinations of neural network structures and mechanistic model structures was suggested. However, we believe these frameworks still does not provide the flexibility and elegance of the operating regime based modeling approach.

Other related empirical model structures are splines (Wahba 1990, Friedman 1991, Kavli 1993, Lane et al. 1992), radial basis function series expansions (Moody and

Darken 1989, Chen, Billings, Cowan and Grant 1990*b*), some fuzzy models (Wang and Mendel 1992), and wavelets (Benveniste et al. 1994). All these approaches lead to local basis-functions, which for low-dimensional problems automatically gives a transparent representation, since each basis-function and its coefficient can be uniquely related to a particular part of the information space. A fundamental problem with all local modeling methods is the curse of dimensionality, which may lead to lack of transparency of the resulting model, and exponentially increasing memory demand and computational complexity as a function of the key dimension. With the present approach, we have shown that this problem may sometimes be reduced considerably, since the operating point may be of lower dimension than the information vector and each local model contains more information than a basis-function. Moreover, the problem can be further reduced through the use of prior knowledge for regime decomposition. One reason is that for most practical problems, one does not need a model with uniform accuracy over the operating range. One reason for this is that there will be operating conditions that are physically infeasible for the dynamic system to operate under. However, for very high dimensional and complex empirical modeling problems, one may prefer projection based methods (De Veaux, Psichogios and Ungar 1993), such as sigmoidal neural nets, or projection pursuit regression (Friedman and Stuetzle 1981).

The title of this chapter is "Operating Regime based Input/Output Modeling". To keep the presentation simple, we have focused entirely on the NARMAX model representation. We do not think that the extension to other input/output model representations, like FIR model representations, or differential equation based model representations, will reveal any suprises.

In summary, decomposing the system's operating range into operating regimes and using local ARMAX models to describe the system behavior in each operating regime is appealing for the following reasons:

- ARMAX models are well understood and widely used in industry, and is therefore a convenient basis for building NARMAX models. If a given process is currently modeled with an ARMAX model, a non-linear model may quite naturally be introduced by extending this model to an NARMAX model based on two or three local ARMAX models.
- The class of (mathematical) systems that can be accurately represented is shown to be large, and the model structure is linearly parameterized.
- The concept and modeling procedure is straightforward, and the model structure is transparent. This is important, since the model structure can be easily interpreted, analyzed, and validated, and prior knowledge can be incorporated.
- Describing a system by means of operating regimes is common practise in engineering.

# Chapter 3

# Operating Regime based State-space Modeling

In this chapter we will discuss the design of non-linear state-space models in the operating regime based modeling framework. Local state-space models introduce some additional features and difficulties compared to local input/output models that were discussed in Chapter 2.

This chapter is organized as follows: First the model representation is introduced in section 3.1. In section 3.2 we discuss how process knowledge can be applied to develop a model structure in terms of operating regimes and local model structures. In section 3.3 we present some examples to illustrate the procedures, see also (Johansen and Foss 1993*b*), and in section 3.4 there is a discussion.

## 3.1 State-Space Model Representation

Here we address the problem of representing a model of the form

$$\dot{x} = f(x, u, v) \tag{3.1}$$

$$y = g(x) + w \tag{3.2}$$

where

| | | |
|---|---|---|
| $x \in \mathcal{X} \subset R^n$ | – | State vector |
| $u \in \mathcal{U} \subset R^r$ | – | Control vector |
| $v \in \mathcal{V} \subset R^n$ | – | Disturbance vector |
| $y \in \mathcal{Y} \subset R^m$ | – | Measurement vector |
| $w \in \mathcal{W} \subset R^m$ | – | Measurement noise vector |

A state-space model of the system consists of knowledge of $f : \mathcal{X} \times \mathcal{U} \times \mathcal{V} \rightarrow R^n$, and $g : \mathcal{X} \rightarrow R^m$, and some structural or statistical knowledge of the signals $v$ and $w$. In this chapter we will be concerned with representation and design of the functions $f$ and $g$.

Within a sufficiently small operating regime, a simple (for example linear) local model

$$\dot{x} = f_i(x, u, v) \tag{3.3}$$
$$y = g_i(x) + w \tag{3.4}$$

will give an adequate description of the system, provided the system satisfies some smoothness condition. This model will be valid within this particular operating regime, and more or less invalid outside this regime. We introduce the *operating point* $z$, which is a vector of variables that characterizes the system's different modes of operation, and the set of operating points $Z$. An *operating regime* is defined as a subset $Z_i \subset Z$. The operating point is typically a function of the state, inputs, and disturbances

$$z = H(x, u, v)$$

Next, we assume that for each local model there exists a smooth *local model validity function* $\rho_i : Z \to [0,1]$ that is constructed such that its value is close to one for operating points where the local model is an accurate description of the system, and close to zero elsewhere. If there are $N$ operating regimes with local models and validity functions for each regime, one may apply the following interpolation to get a global model

$$\dot{x} = \sum_{i=1}^{N} f_i(x, u, v) w_i(z) \tag{3.5}$$
$$y = \sum_{i=1}^{N} g_i(x) w_i(z) + w \tag{3.6}$$
$$w_i(z) = \frac{\rho_i(z)}{\sum_{j=1}^{N} \rho_j(z)} \tag{3.7}$$

The *interpolation function* $w_i : Z \to [0,1]$ is a normalization of the model validity function $\rho_i$, which has the property that $\sum_{i=1}^{N} w_i(z) = 1$ for all $z \in Z$. To get a complete global model, it must be assumed that at any operating point $z \in Z$, not all local model validity functions vanish.

### 3.1.1 Local Models of Different Structure

In general, only the input and output vectors need to be common for two different local models of the same system, as these are the only variables that are directly related to the *system*. States, disturbances, noise, parameters, and internal variables are directly related to the *model*. We have chosen to use the same superset of variables in all local models. Of course, each local model will only use the relevant subset of the variables. The motivation is the simplified notation, and the fact that dealing with all the variables simultaneously is required in the interpolation (3.5)-(3.7) anyway.

Consider the following two local models

$$\begin{aligned}\dot{x}^1 &= f_1(x^1, u, v)\\ \dot{x}^2 &= f_2(x^2, u, v)\end{aligned}$$

where $x^1 = (\xi_1\ \xi_2)^T$ and $x^2 = (\xi_2\ \xi_3)^T$. Typical reasons for different state-spaces may be that $\xi_3$ is not a state in the first local model either because it is *constant*, or *irrelevant* in that operating regime. Again, the reason for this may be of physical nature, like

- The temperature of boiling water is constant, while it may vary when it is not boiling.
- In a certain range of flow-rates there may be gas bubbles in e.g. a pipeline or reactor. The phase defined by the gas bubbles may be described by states that are not relevant at other flow rates, simply because the gas bubbles will not exist under these flow conditions.

Alternatively, the reason may be linked to the model representation, for example

- The local model structure in one regime is a black-box model structure, while it is a mechanistic model structure in another regime. Clearly, the state-spaces may be completely different.
- A simple model of low order is sufficient in one regime, but a high order model is required in another regime.

Anyway, when the model is applied for prediction during transition between two regimes with different state-spaces, one needs to answer the question: "What are the "initial" state of the local model associated with the regime just being entered?". A simple answer exists in cases when there exists a one-to-one mapping between the state-spaces, or from the input/output space to the state-space, which will typically be the case when the model is in some canonical form. In the following, we will describe a more general solution that mainly assumes observability of the states associated with each local model in a sufficiently large region containing the operating regime.

If $\xi_3$ is constant in the first operating regime, we have $\dot{\xi}_3 = 0$ in that regime, and $\xi_3$ enters as a constant parameter in the function $f_1$. Let us assume in addition that $\xi_1$ is constant in the second operating regime. Defining an augmented state vector $x = (\xi_1, \xi_2, \xi_3)^T$, we get the local models

$$\dot{x} = \begin{pmatrix}\dot{\xi}_1\\ \dot{\xi}_2\\ \dot{\xi}_3\end{pmatrix} = \begin{pmatrix} f_{11}(\xi_1, \xi_2, \xi_3, u, v)\\ f_{12}(\xi_1, \xi_2, \xi_3, u, v)\\ 0\end{pmatrix}$$

$$\dot{x} = \begin{pmatrix}\dot{\xi}_1\\ \dot{\xi}_2\\ \dot{\xi}_3\end{pmatrix} = \begin{pmatrix} 0\\ f_{22}(\xi_1, \xi_2, \xi_3, u, v)\\ f_{23}(\xi_1, \xi_2, \xi_3, u, v)\end{pmatrix}$$

A global model is now

$$\begin{pmatrix}\dot{\xi}_1\\ \dot{\xi}_2\\ \dot{\xi}_3\end{pmatrix} = \begin{pmatrix} f_{11}(\xi_1, \xi_2, \xi_3, u, v)\cdot w_1(z) & + \ 0 \cdot w_2(z)\\ f_{12}(\xi_1, \xi_2, \xi_3, u, v)\cdot w_1(z) & + \ f_{22}(\xi_1, \xi_2, \xi_3, u, v)\cdot w_2(z)\\ 0\cdot w_1(z) & + \ f_{23}(\xi_1, \xi_2, \xi_3, u, v)\cdot w_2(z)\end{pmatrix} \tag{3.8}$$

The other possibility is that $\xi_3$ is irrelevant in the first operating regime. This means that $f_1$ is not dependent on $\xi_3$, but it does not mean that $\xi_3$ is constant. Instead of augmenting with an equation $\dot{\xi}_3 = 0$, we model $\xi_3$ as a random walk process $\dot{\xi}_3 = \nu_3$ in this operating regime. Similarly, we model $\xi_1$ as a random walk process in the second operating regime. Now we get a global model

$$\begin{pmatrix} \dot{\xi}_1 \\ \dot{\xi}_2 \\ \dot{\xi}_3 \end{pmatrix} = \begin{pmatrix} f_{11}(\xi_1, \xi_2, u, v) \cdot w_1(z) & + & \nu_1 \cdot w_2(z) \\ f_{12}(\xi_1, \xi_2, u, v) \cdot w_1(z) & + & f_{22}(\xi_2, \xi_3, u, v) \cdot w_2(z) \\ \nu_3 \cdot w_1(z) & + & f_{23}(\xi_2, \xi_3, u, v) \cdot w_2(z) \end{pmatrix}$$

The variables $\nu_1$ and $\nu_3$ are unknown, but using an extended Kalman-filter to estimate the irrelevant process states, it is sufficient to assume these variables to be random white noise processes with known variance. If the model is applied on-line, the measured output can be used in the Kalman-filter. Otherwise, the global model output is used.

If the irrelevant variable $\xi_3$ is not observable in the first regime, it is impossible to update an estimate of it using an extended Kalman-filter, since there is no information available about $\xi_3$ in that regime. The best thing to do is to turn off (parts of) the estimator until the process enters an regime where $\xi_3$ is observable.

Another potential problem that in particular arises when one tries to transfer state or output information from an empirical local model to a mechanistic local model is that the empirical model may suggest state values in the mechanistic model that are non-physical. If this is a potential problem, it must be handled in an ad.hoc. manner by e.g. projecting the variables to a physically feasible region, since such mismatch between physical reality and model prediction is a fundamental problem with empirical models.

### 3.1.2 Local Linear Models

An interesting special case is when the local models are all linear with equal structure

$$\begin{aligned} \dot{x} &= a_i + A_i \Delta x_i + B_i \Delta u_i + C_i \Delta v_i \\ y &= d_i + D_i \Delta x_i + w \end{aligned}$$

Here we linearize around points $x_i$, $u_i$, $v_i$, and define $\Delta x_i = x - x_i$, $\Delta u_i = u - u_i$, $\Delta v_i = v - v_i$, $a_i = f(x_i, u_i, v_i)$, $d_i = g(x_i)$, $A_i = \partial f / \partial x(x_i, u_i, v_i)$, $B_i = \partial f / \partial u(x_i, u_i, v_i)$, $C_i = \partial f / \partial v(x_i, u_i, v_i)$, and $D_i = \partial g / \partial x(x_i)$. Then (3.3)-(3.4) becomes

$$\dot{x} = \sum_{i=1}^{N} (a_i + A_i \Delta x_i + B_i \Delta u_i + C_i \Delta v_i) w_i(z) \tag{3.9}$$

$$y = \sum_{i=1}^{N} (d_i + D_i \Delta x_i) w_i(z) + w \tag{3.10}$$

The model representation (3.9)-(3.10) with local linear models can be viewed as a quasi-linear model representation where the matrices are functions of the operating point $z$

$$\begin{aligned}
\dot{x} &= \sum_{i=1}^{N}(a_i + A_i(x - x_i) + B_i(u - u_i) + C_i(v - v_i))w_i(z) \\
&= \sum_{i=1}^{N}(a_i - A_i x_i - B_i u_i - C_i v_i)w_i(z) + \\
&\quad \left(\sum_{i=1}^{N} A_i w_i(z)\right) x + \left(\sum_{i=1}^{N} B_i w_i(z)\right) u + \left(\sum_{i=1}^{N} C_i w_i(z)\right) v \\
&= a(z) + A(z)x + B(z)u + C(z)v
\end{aligned} \tag{3.11}$$

Of course, the measurement equation can be rewritten in the same manner

$$y = d(z) + D(z)x + w$$

It is easy to see that any $f$ and $g$ can be written in the form

$$\dot{x} = f(x, u, v) = \tilde{a}(x, u, v) + \tilde{A}(x, u, v)x + \tilde{B}(x, u, v)u + \tilde{C}(x, u, v)v \tag{3.12}$$

$$y = g(x) + w = \tilde{d}(x) + \tilde{D}(x)x + w \tag{3.13}$$

This offers a guide for the choice of $Z$ and mapping $H$, which is essentially the same as the theoretical result in Chapter 2, but perhaps more intuitive. Comparing (3.11) with (3.12), it is clear that $z = H(x, u, v)$ can be chosen such that

$$\tilde{A}(x, u, v) = A(z), \quad \tilde{B}(x, u, v) = B(z), \quad \text{etc.}$$

In other words, such that $z$ captures the non-linearities. Notice that because the representation (3.12)-(3.13) is non-unique. Hence, there are several possible choices for $H$.

## 3.2 Incorporating Process Knowledge

In section 1.3 we stated that a major motivation for the operating regime based modeling approach is that local modeling is simpler than global modeling because locally there are less relevant phenomena, and their interactions are simpler. Let us illustrate this fact with an almost trivial example. Consider a chemical reactor where two globally relevant phenomena are *reaction* and *flow pattern*. Clearly, these are interacting since the flow pattern may influence the reaction kinetics, and the composition may influence the flow pattern. Developing a model that is valid for a wide range of flow patterns and compositions will be a laborous task because of the large number of mass- and heat-transfer phenomena that must be taken into account simultaneously. A local model, on the other hand, need only be valid for one particular flow pattern and a small range of compositions, and is typically significantly simpler to develop because the interactions between reaction and flow pattern are simpler. Of course, one should compare the cost of developing the required number of local models to the cost of developing one global model directly. When the number of operating regimes is large, this may appear as a major drawback of the operating regime based modeling approach. However, it is not necessarily so. The reason is that having developed one local model structure, the remaining ones follow often more easily. For example, it will often make sense to let a number of the local model structures be equal. A very simple, but generally applicable approach is to use the same local linear model structure in all regimes.

The basic modeling procedure outlined in section 2.2 can also be applied for state-space models. A major task is to identify and characterize the system's operating regimes, and select local model structures. Typically, this is based on one of the following two approaches:

1. Different mechanisms or physical phenomena are dominant in different operating regimes, giving rise to different local model structures.
2. Different local model parameters in different operating regimes, but the same local model structures.

Choosing a priori a linear model structure for each operating regime leads to the second approach. With this approach, an operating regime must be defined to be a region in which the system behaves approximately linearly. Hence, the decomposition into operating regimes is motivated and given by the system's non-linearities, not by the different physical phenomena. Compared to the first approach, this is a more empirical and less knowledge-demanding approach. There may be some cases that falls more or less into both these categories, for example when the parametric variations in the local model structure are explicitly related to physical phenomena or mechanisms.

Often, there will be a number of model variables that are candidates for being used to characterize operating regimes (i.e. included in the vector $z$). For example, consider a set of linear differential equations describing the mass- and energy balances of a chemical reactor. Such a model is clearly a local model, since chemical reactors usually exhibit strongly non-linear behavior. Different regimes may be characterized using variables like temperature and flow-rate, since the parameters of the linear local model structure will be different at high and low temperature, because the reaction rate is usually highly temperature dependent. Likewise, the flow-rate will influence the holdup-time, mixing, and flow patterns, and thereby the linear model parameters. In addition, there may be different operating regimes because a different set of chemical reactions may dominate under different conditions, often characterized by temperature, composition, or catalyst properties. Although this example is simple, it illustrates two important points: First, the choice of operating regimes and local model structures are closely interconnected. For example, if the local linear models in the example are replaced by non-linear local models where the reaction rate constants are replaced by Arrhenius-type terms, then each local model would be valid over a much wider, possibly global, temperature range. Hence, the dimension of the operating point space can usually be reduced at the cost of increasing the complexity of the local model structures. Second, the example illustrates that the process knowledge used for decomposition into operating regimes may be quite elementary, like knowledge of conditions under which different chemical reactions dominate, and the knowledge that the reaction rates depend strongly on temperature. Supplementary knowledge, like "The reaction rate will be approximately doubled if the temperature increases by 10 K" will be sufficient to decide roughly on how to design regimes along the temperature axis. Knowledge about how accurate the model need to be under different conditions can be used for the same purpose. For example, for on-line dynamic optimization purposes, a detailed model may be needed to describe the reactor under normal operating conditions in order to get close to optimal operation. Under other operating conditions, like startup, shutdown, product shifts, and abnormal operation, a rough model may be sufficient for the control system to bring the system safely and reliably into the desired operating mode.

It is not difficult to imagine that the problem may not always be as transparent as in the example above. If the number of model variables is much larger, and there are significant dependences among these variables, and the system behavior is not well understood, then one may need to take a different approach to the choice of operating space and regimes. One will typically need to analyse the correlation structure and collinearities in the empirical data as proposed by Sugeno and Kang (1986), see also (Nakamori et al. 1992). Projection pursuit tools (Huber 1985) like principal component analysis will guide the search for a small number of variables, or combination or variables, that contain sufficient information to characterize the system behavior in terms of operating regimes in relation to the local models. Such a projection pursuit algorithm may search for a projection within a class of projection operators, e.g. linear ones. A suitable optimization criterion will give penalty on large prediction errors and other undesirable properties of the model. For example, it is always desirable with an operating space of low dimension, since this will make the model structure more transparent, and the curse of dimensionality will be avoided. However, transparency also requires interpretability of the variables characterizing the operating regimes. Unfortunately, this is sometimes in conflict with the requirement of low dimension, so a tradeoff must be made.

When a rough decomposition into regimes is designed, one must choose how much the local model validity functions should overlap. The main motivation for using a smooth interpolation of the local models is that usually the system has some smoothness properties, i.e. the behavior changes smoothly and the phenomena appear or disappear gradually as the operating point changes. However, one may occasionally come across systems that are non-smooth, in the sense that they exhibit abrupt changes in behavior or mechanisms (jump phenomena, catastrophes, bifurcations). Reasons for this may be diverse, like in a combustion system that certainly behave differently depending on the presence of an ignition source, or in a flow system where a unit may be suddenly by-passed. There also exist phenomena that appear gradually, but rapidly compared to the interesting dynamics, as the operating point changes. Such systems may also be looked upon as non-smooth in some cases. This may include phenomena related to phase transitions and some changes in fluid flow patterns. In such cases it may be of interest use discrete variables to describe the operating point, or design non-smooth local model validity functions to capture these phenomena in a best possible way, see also (Hilhorst 1992, Söderman et al. 1993). However, if it is believed that the system has sufficient smoothness properties, then smoothly overlapping local model validity functions seems reasonable. If we use the model's ability to approximate the system behavior as a criterion for the choice of overlap, our experience indicate that there may exists a well defined optimum, see also Chapter 2. Unfortunately, the optimal overlap will depend on the modeling problem. On the other hand, we have also experienced that a rough choice will often give a close-to-optimal result. An important factor that may influence the choice of overlap is that a large overlap will lead to decreased interpretability of the model if the parameters are identified using a global criterion (Murray-Smith 1994*b*). The reason for this is essentially that an increased overlap will increase the dependences and interactions between the local models. If the local model parameters are fitted using empirical data, a large overlap may imply that a given data point will be relevant for a large number regimes. This may lead to local models that cannot be interpreted as local approximations to the system, since the parameters of some local models may be highly influenced by remote data points. Often, one local model will have a behavior that partially compensates for the behavior of another local model in the overlapping region. Hence, the global model may have a behavior that closely approximates the system behavior, but none of the local models approximate the behavior of the system locally. Clearly, this makes the interpretation somewhat more difficult. If the local model parameters are identified using local criteria, then the choice of overlap becomes more critical, since a large overlap will have a regularizing effect (Murray-Smith 1994*a*).

## 3.3 Examples

### 3.3.1 Simulation Example: Population Dynamics

A classical non-linear modeling problem is a predator/prey system. Here we will apply the local modeling framework to such a system. Let $x_1$ be the concentration of prey, and $x_2$ the concentration of predators. The basic mechanistic knowledge and assumptions one has about this system can be qualitatively expressed as the set of rules

IF $x_1$ is large THEN $x_2$ increases

IF $x_1$ is small THEN $x_2$ decreases

IF $x_2$ is large THEN $x_1$ decreases

IF $x_2$ is small THEN $x_1$ increases

As illustrated in Fig. 3.1, this gives four regions of the state space where the system behaves qualitatively different. Within each region, the system behavior is qualitatively simple. Adding some more detail, we can write the model as

IF $x_1$ is larger than $\overline{x}_1$ THEN $\dot{x}_2 = a_{22}x_2$

IF $x_1$ is smaller than $\overline{x}_1$ THEN $\dot{x}_2 = -a_{21}x_2$

IF $x_2$ is larger than $\overline{x}_2$ THEN $\dot{x}_1 = -a_{12}x_1$

IF $x_2$ is smaller than $\overline{x}_2$ THEN $\dot{x}_1 = a_{11}x_1$

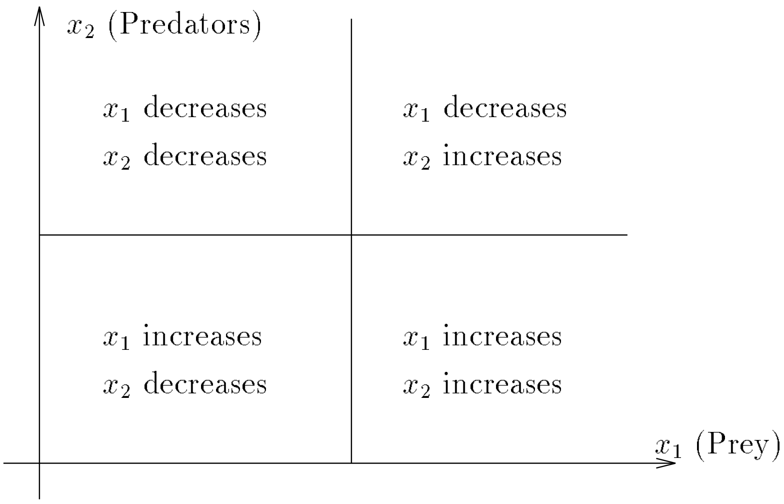


Figure 3.1: A high-level description of the predator/prey model.

where $\overline{x}_1$ and $\overline{x}_2$ are the average concentrations of prey and predators over a long period of time, respectively. Since there is no reason to believe that the behavior changes abruptly, the "smaller than" and "larger than" statements should be interpreted in a "soft" way. This gives four local operating regimes as illustrated in Fig. 3.2a, with local model validity functions as in Fig. 3.2b, and interpolation functions as in Fig. 3.2c. With $\overline{x}_1 = \overline{x}_2 = a_{11} = a_{12} = a_{21} = a_{22} = 1$, this model has trajectories as illustrated in Fig. 3.3a, where each closed cycle corresponds to different initial conditions. We see that the behavior of this model is qualitatively similar to the classical Lotka-Volterra model (Volterra 1931)

$$\begin{aligned}\dot{x}_1 &= b_{11}x_1 - b_{12}x_1x_2 \\ \dot{x}_2 &= -b_{21}x_2 + b_{22}x_1x_2\end{aligned}$$

which has trajectories as illustrated in Fig. 3.3b, with $b_{11} = b_{12} = b_{21} = b_{22} = 1$.

Now, these two models have similar behavior, but the underlying model representation is quite different, even though the same basic knowledge about the system mechanisms is applied during the modeling processes. Clearly, the Lotka-Volterra model is the most compact. On the other hand, the operating regime based model is definitely a mechanistic model, and the representation is perhaps closer to the way most people (including scientists and engineers) understand the behavior of this system. The modeling assumptions are more explicit and the model parameters are perhaps more transparent than in the Lotka-Volterra model. Hence, the operating regime based modeling framework has elevated the modeling process to a more transparent level than the equation-based modeling approach.

It is interesting to observe that a very similar operating regime based alternative to the Lotka-Volterra model was coined by Kolmogoroff (1936). This much less known work is also based on four regimes (called zones) with qualitatively different

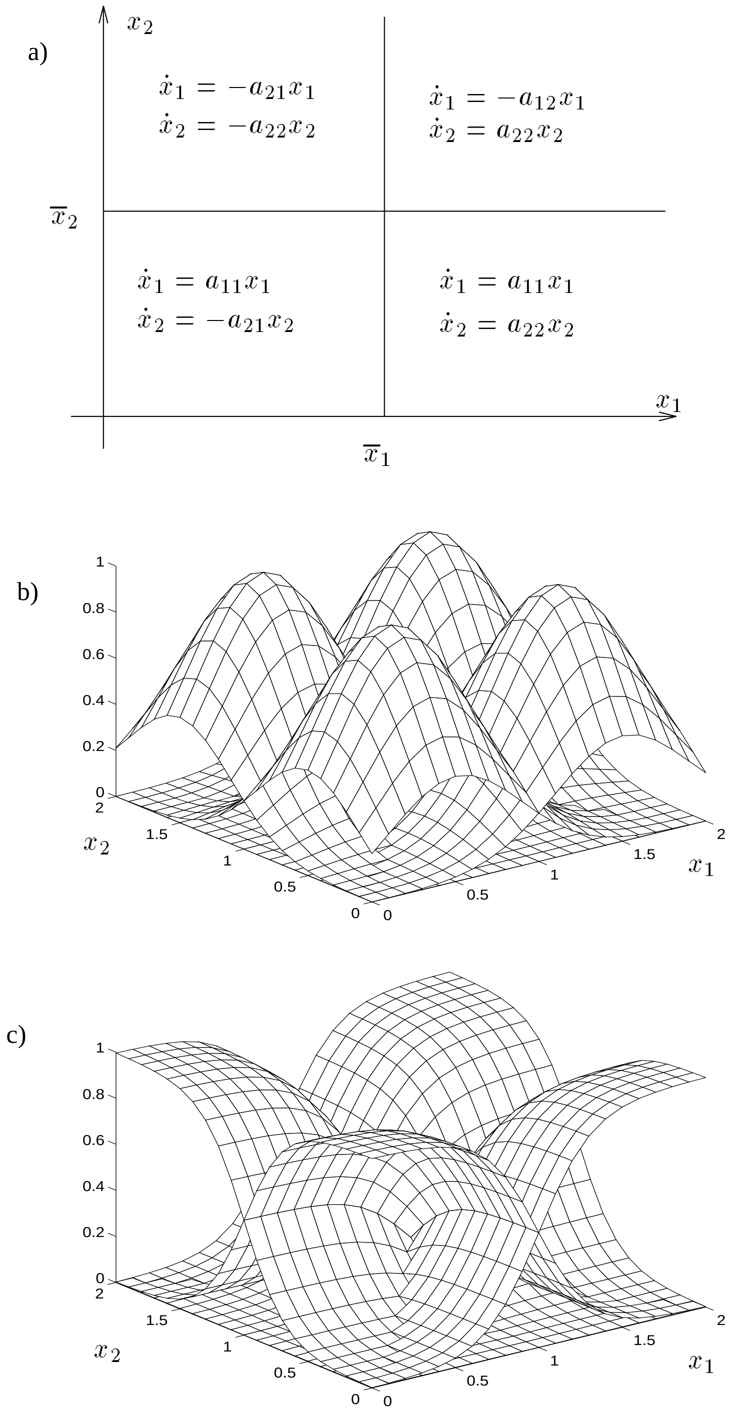


Figure 3.2: a) The operating regimes and local models, b) local model validity functions, and c) interpolation functions.

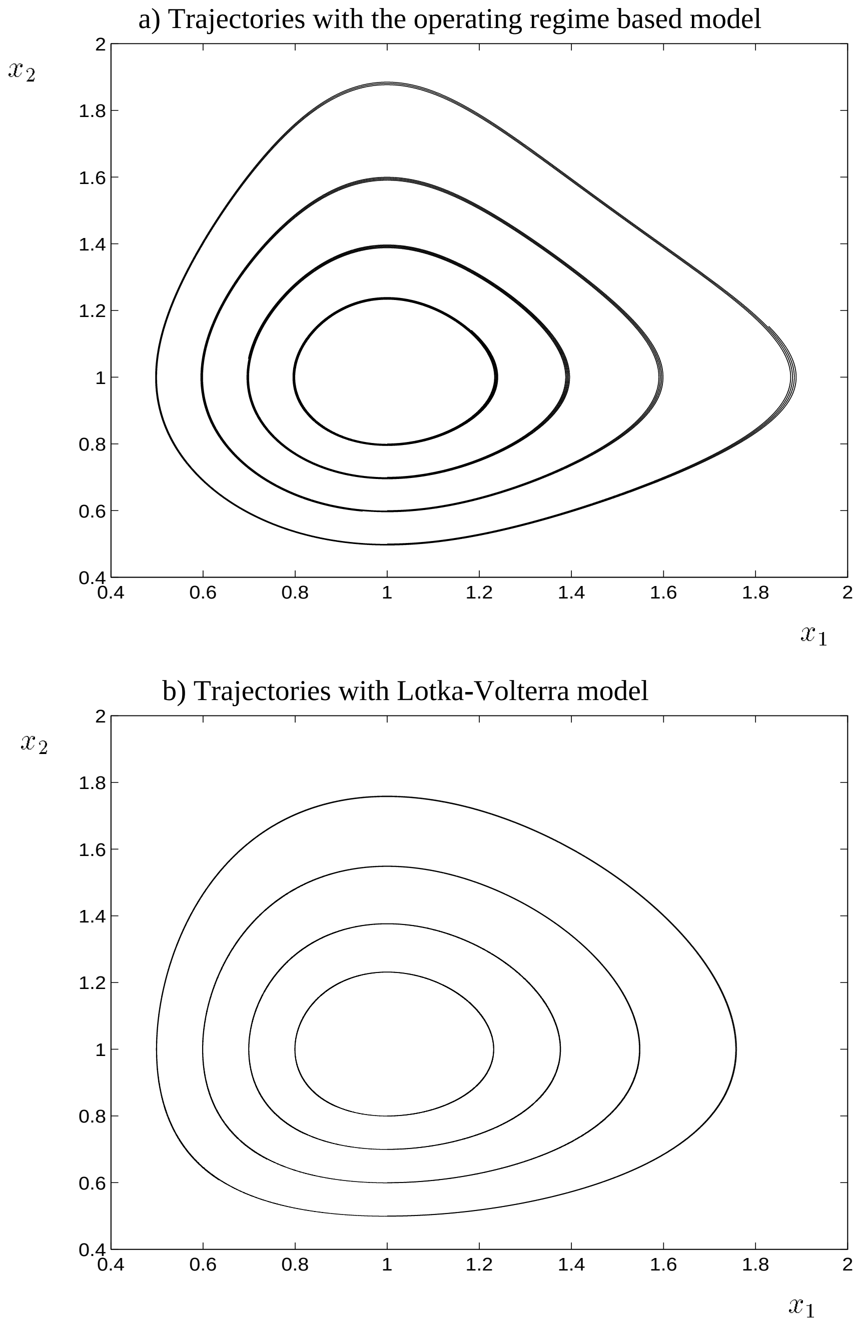


Figure 3.3: a) Trajectories with operating regime based model, and b) Trajectories with the Lotka-Volterra model.

behavior of the system, similar to Fig. 3.2. Rescigno and Richardson (1973) concludes that this is "...a general qualitative theory for the interactions between two species" and they show how the modeling principle can be applied to somewhat more complex systems. From (Rescigno and Richardson 1973) it appears that such model representations have had some impact on the understanding and analysis of such systems in the area of mathematical biology.

### 3.3.2 Simulation Example: A Batch Fermentation Reactor

Batch fermentation processes are examples of processes where an operating regime based model appears to be a may be quite natural approach. A typical batch consists of the following sequence of phases: *Initial lag phase* where the cells adapt to their new environment. *Growth phase* where the cell culture grows, *production phase* when the enzymes in the cells catalyses various chemical reactions, and *termination* when the reactions have stopped (Bailey and Ollis 1986). The production phase can in many cases be further decomposed into several more specific regimes. This will be the case if there is a sequence of chemical reaction steps, or if the cells under different environmental conditions form different enzymes and therefore catalyses different chemical reactions. Moreover, there may also be different regimes associated with various faulty modes of operation. It is natural to let the regimes overlap. For example, the chemical reactions in the production regimes may start immediately, but may only be significant after the cell culture has grown for some period. The cell growth may continue through the production regime. Often, there will be complex relations between the different regimes, since some chemical or biological components may inhibit some chemical reactions, cell growth or enzyme formation, while stimulating others. A more detailed discussion of bioreactors and local models can be found in (Zhang et al. 1994).

Building global mechanistic models of such processes may be a resource demanding task, e.g. (Bailey and Ollis 1986). Here we present a semi-realistic simulation example that illustrates how operating regimes and local models can be used to simplify the modeling problem. We will study the fermentation of glucose to gluconic acid by the micro-organism *Pseudomonas ovalis* in a well stirred batch reactor. The main overall reaction mechanism can be described by

$$\begin{aligned} \text{Cells} + \text{Glucose} + O_2 &\rightarrow \text{More Cells} \\ \text{Glucose} + O_2 &\overset{\text{Cells}}{\rightarrow} \text{Gluconolactone} \\ \text{Gluconolactone} + H_2O &\rightarrow \text{Gluconic Acid} \end{aligned}$$

The first reaction is the reproduction of cells, using glucose and oxygen. The second reaction is the production of the intermediate product gluconolactone, again using glucose and oxygen. This reaction is enzyme-catalysed by the cells, while the last reaction forms the final product, gluconic acid. In the simulation study, we use the following state-space model to simulate the "true system" (Ghose and Ghosh 1976):

$$\dot{\chi} = \mu_m \frac{sc}{k_s c + k_o s + sc} \chi$$

| Symbol | Description |
|---|---|
| $\chi$ | Cell concentration $[UOD/ml]$ |
| $p$ | Gluconic acid concentration $[g/l]$ |
| $l$ | Gluconolactone concentration $[g/l]$ |
| $s$ | Glucose concentration $[g/l]$ |
| $c$ | Dissolved oxygen concentration $[g/l]$ |
| $OUR$ | Oxygen uptake rate $[g/lh]$ |
| $\mu_m$ | $0.39\ h^{-1}$ |
| $k_s$ | $2.50\ g/l$ |
| $k_o$ | $0.55 \cdot 10^{-3}\ g/l$ |
| $k_p$ | $0.645\ h^{-1}$ |
| $v_l$ | $8.30mg\ UOD^{-1}\ h^{-1}$ |
| $k_l$ | $12.80\ g/l$ |
| $k_la$ | $\approx 180.0\ h^{-1}$ |
| $Y_s$ | $0.375\ UOD/mg$ |
| $Y_o$ | $0.890\ UOD/mg$ |
| $c^\star$ | $6.85 \cdot 10^{-3}\ g/l$ |

Table 3.1: Symbols and constants in state-space model of the fermenter.

$$\begin{aligned}
\dot{p} &= k_p l \\
\dot{l} &= v_l \frac{s}{k_l + s}\chi - 0.91 k_p l \\
\dot{s} &= -\frac{1}{Y_s}\mu_m \frac{sc}{k_s c + k_o s + sc}\chi - 1.011 v_l \frac{s}{k_l + s}\chi \\
\dot{c} &= k_l a(c^\star - c) - \frac{1}{Y_o}\mu_m \frac{sc}{k_s c + k_o s + sc}\chi - 0.09 v_l \frac{s}{k_l + s}\chi
\end{aligned}$$

with initial conditions $\chi(0) = \chi_0$, $p(0) = 0$, $l(0) = 0$, $s(0) = s_0$ and $c(0) = c^\star$. The symbols are defined in Table 3.1. It is assumed that the volume is constant, and that glucose and oxygen are rate-limiting. Simulating 10 $h$ batches using this model, "measuring" all states with 0.5 $h$ intervals, and adding "random measurement-noise" with a signal-to-noise ration of 30 dB to the "measured" states, two sets of data are collected, by varying $\chi_0, s_0$ and the agitation speed (affecting $k_la$). The first set contains data from 100 batches, and is used for estimation, while the second independent set, containing 30 batches, shown in Fig. 3.4, is used for model validation only. We define the following normalized variables:

$$\begin{aligned}
\chi_n &= \frac{\chi}{3\ UOD/ml} \\
p_n &= \frac{p}{50\ g/l} \\
l_n &= \frac{l}{15\ g/l} \\
s_n &= \frac{s}{50\ g/l}
\end{aligned}$$

$$c_n = \frac{c}{0.01\ g/l}$$
$$OUR_n = \frac{OUR}{g/lh}$$

where OUR is the oxygen uptake rate, and the normalized state-vector is defined by

$$x = \begin{pmatrix} \chi_n \\ p_n \\ l_n \\ s_n \\ c_n \end{pmatrix}$$

Using knowledge of the overall reaction mechanism, we define three operating regimes for this process, which are entered in the following sequence:

1. **Cell growth regime.** Qualitatively, this regime is characterized by cell reproduction and some limited chemical production.
2. **Cell growth and production.** All three reactions in the reaction scheme are significant.
3. **Final production.** This regime is characterized by shortage of glucose. Hence, only the last chemical reaction takes place.

In the following, we will describe two possible decompositions of this process' operating range into regimes, based on two different operating point vectors.

## Characterizing Operating Regimes using $\chi$ and OUR

The operating point is defined by

$$z(t) = \begin{pmatrix} OUR_n(t) \\ \chi_n(t) \end{pmatrix}$$

The three operating regimes can be characterized in the following way:

1. Small cell population $\chi$, and a small OUR. Since the two first reactions both consume oxygen, OUR is a measure of the total rate of those reactions.
2. Since all reactions progress at a high rate, the OUR is high.
3. Since only the last reaction takes place, and no oxygen is consumed in this reaction, the regime is characterized by a small OUR. In addition, $\chi$ tend to be large.

We use the following local linear model structure

$$\dot{x} = a_i + A_i x$$

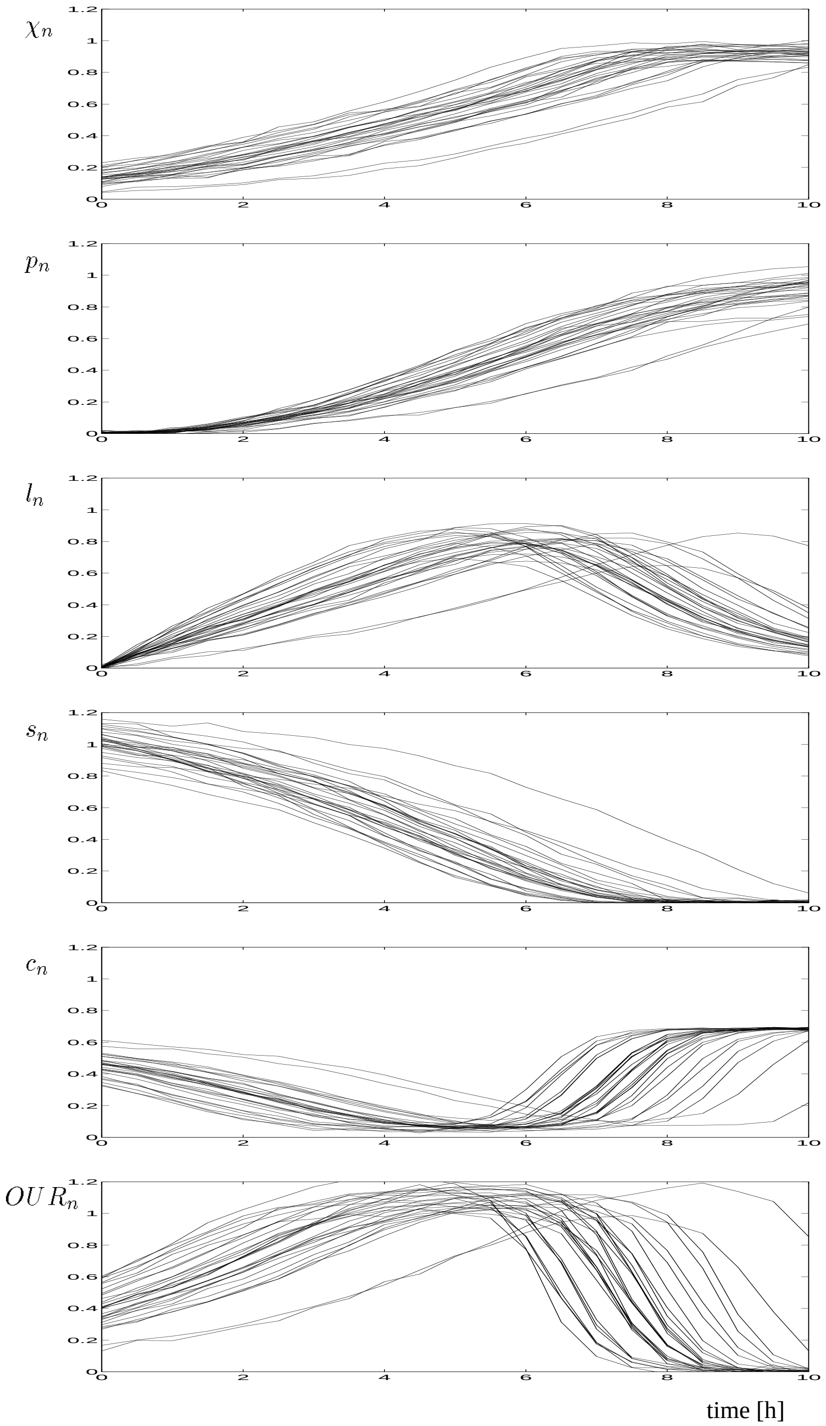


Figure 3.4: Data collected from simulated batch runs of fermenter.

where $a_i$ is a $5 \times 1$ vector, and $A_i$ a $5 \times 5$ matrix. Identification of the unknown local model parameters, using a standard least squares algorithm based on a global criterion, gives the following local model parameters

$$a_1 = \begin{pmatrix} 0.3550 \\ 0.0079 \\ -0.0184 \\ 0.0163 \\ -0.0018 \end{pmatrix}, \quad A_1 = \begin{pmatrix} -0.4457 & 0 & 0 & -0.0335 & -0.4547 \\ 0 & 0 & 0.1825 & 0 & 0 \\ 1.1308 & 0 & -0.5746 & 0.0405 & 0 \\ -0.3529 & 0 & 0 & -0.0411 & 0.0119 \\ -0.3764 & 0 & 0 & -0.0228 & 0.0140 \end{pmatrix}$$

$$a_2 = \begin{pmatrix} 0.0527 \\ 0.0154 \\ 0.0679 \\ -0.2020 \\ -0.0631 \end{pmatrix}, \quad A_2 = \begin{pmatrix} -0.1242 & 0 & 0 & -0.0371 & 0.2206 \\ 0 & 0 & 0.1813 & 0 & 0 \\ -0.2938 & 0 & 0.1734 & 0.1751 & 0 \\ 0.0947 & 0 & 0 & 0.0114 & 0.2345 \\ 0.0194 & 0 & 0 & -0.0149 & 0.0157 \end{pmatrix}$$

$$a_3 = \begin{pmatrix} 0.2362 \\ -0.0016 \\ -0.0245 \\ 0.0547 \\ 0.9224 \end{pmatrix}, \quad A_3 = \begin{pmatrix} -0.0775 & 0 & 0 & -0.3114 & -0.2400 \\ 0 & 0 & 0.1731 & 0 & 0 \\ 0.0170 & 0 & -0.4674 & 0.7240 & 0 \\ -0.0334 & 0 & 0 & -1.6363 & -0.0372 \\ 0.0982 & 0 & 0 & -1.4034 & -1.4749 \end{pmatrix}$$

where the zeros are structural and follows immediately by inspecting the overall reaction mechanism. These local models may be viewed as approximate linearizations of a mass balance for this process. Notice that the local model structures may be viewed as semi-empirical, since their structure is based on some limited mechanistic knowledge.

Using Gaussian local model validity functions, with an ad.hoc. choice of scaling matrices and location vectors to represent the regimes, we get interpolation functions as shown in figure 3.5. The process' transition between the operating regimes is illustrated in figure 3.6. The application of fuzzy sets to characterize the three regimes gives similar results (Johansen and Foss 1993*b*).

### Characterizing Operating Regimes using $c$ and $s$

As discussed in section 3.1.2, a general method for finding a suitable operating point vector is to use those state-variables that captures the non-linearities. Of course, some prior knowledge is required here. In the fermenter example, we suggest

$$z(t) = \begin{pmatrix} c_n(t) \\ s_n(t) \end{pmatrix}$$

as an operating point satisfying this condition. Since $c$ and $s$ are the concentrations of the rate-limiting substrates, this can be expected without knowing the equations

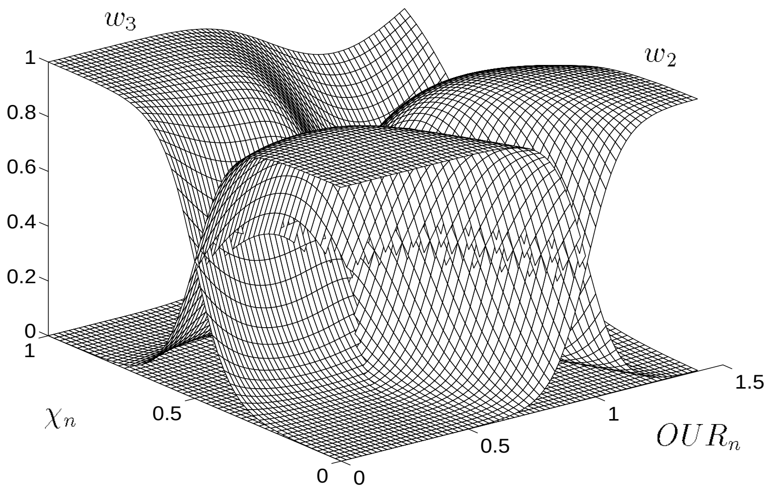


Figure 3.5: Interpolation functions $w_1$, $w_2$ and $w_3$ for the three local models of the fermenter, using $OUR$ and $\chi$ to characterize the operating regimes.

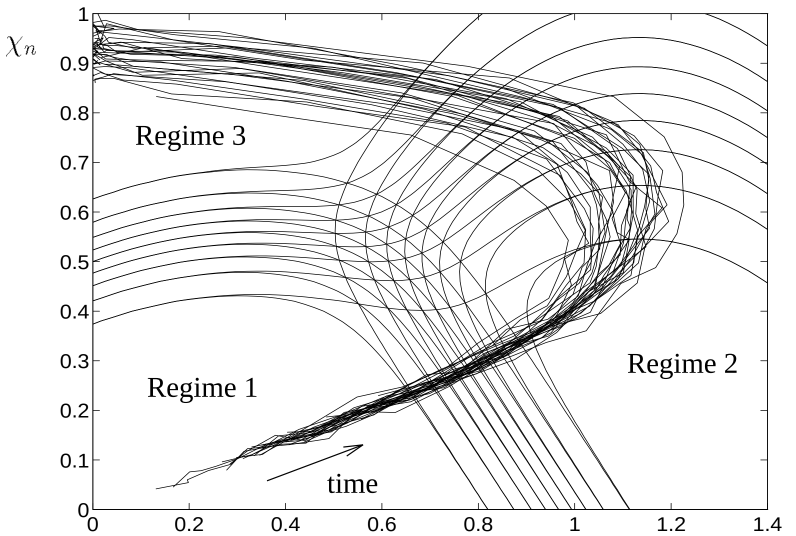


Figure 3.6: The trajectories show the evolution of the operating point $(OUR, \chi)$ for some batches. The trajectories start at a low $\chi$ and $OUR$ and end up at high $\chi$. The level curves for the interpolation functions indicate the region of transition between the operating regimes.

for the "true system". It follows directly from a simple mass balance based on the reaction mechanisms and the assumption that the reaction rates depend only on $c$ and $s$. We use four local models with a rough choice of Gaussian local model validity functions. These model validity functions gives interpolation functions as shown in figure 3.7, and transitions between the regimes as illustrated in figure 3.8. The local model parameters are identified using the same procedure as before:

$$a_1 = \begin{pmatrix} 0.2716 \\ 0.0051 \\ -0.0745 \\ 0.0313 \\ -0.0354 \end{pmatrix}, \quad A_1 = \begin{pmatrix} -0.2816 & 0 & 0 & -0.0130 & -0.3652 \\ 0 & 0 & 0.1922 & 0 & 0 \\ 1.0788 & 0 & -0.4926 & -0.1009 & 0 \\ -0.4071 & 0 & 0 & -0.0525 & 0.0131 \\ -0.1158 & 0 & 0 & -0.0639 & 0.0832 \end{pmatrix}$$

$$a_2 = \begin{pmatrix} 0.1192 \\ 0.0120 \\ -0.1109 \\ 0.0299 \\ -0.0346 \end{pmatrix}, \quad A_2 = \begin{pmatrix} -0.0117 & 0 & 0 & -0.0468 & 0.0497 \\ 0 & 0 & 0.1847 & 0 & 0 \\ 0.7084 & 0 & -0.4225 & 0.3289 & 0 \\ -0.2827 & 0 & 0 & -0.1462 & 0.0481 \\ 0.0703 & 0 & 0 & -0.1357 & -0.2770 \end{pmatrix}$$

$$a_3 = \begin{pmatrix} 0.2342 \\ -0.0221 \\ 0.2656 \\ 0.0000 \\ 0.1376 \end{pmatrix}, \quad A_3 = \begin{pmatrix} -0.0499 & 0 & 0 & -0.1231 & -0.2770 \\ 0 & 0 & 0.2131 & 0 & 0 \\ -0.5251 & 0 & -0.100 & 0.4651 & 0 \\ -0.0811 & 0 & 0 & -0.5285 & 0.1107 \\ 0.5058 & 0 & 0 & -1.7822 & -0.3871 \end{pmatrix}$$

$$a_4 = \begin{pmatrix} 0.2213 \\ -0.0015 \\ 0.0141 \\ 0.0031 \\ 0.7119 \end{pmatrix}, \quad A_4 = \begin{pmatrix} -0.1058 & 0 & 0 & -0.1677 & -0.1800 \\ 0 & 0 & 0.1730 & 0 & 0 \\ -0.0218 & 0 & -0.4703 & 0.0590 & 0 \\ 0.0310 & 0 & 0 & -2.0339 & -0.0471 \\ 0.0190 & 0 & 0 & -0.1516 & -1.0619 \end{pmatrix}$$

Notice that there is a close relationship between the two candidate operating point vectors $(OUR, \chi)$ and $(c, s)$. In fact, there is a one-to-one relationship between $OUR$ and $c$, and $\chi$ will be increasing with about the same rate as $s$ is decreasing, because the process is operating in batch mode. The use of four operating regimes is motivated by the observation that this improves the prediction capabilities of the global model somewhat, compared to three operating regimes. It is in particular when the glucose concentration is very low that the model benefits from the additional regime. The reason for this is that the production regime can naturally be decomposed into two regimes. Early in the production regime, it is the oxygen uptake rate that limits the rate of the reactions, until eventually glucose becomes rate limiting.

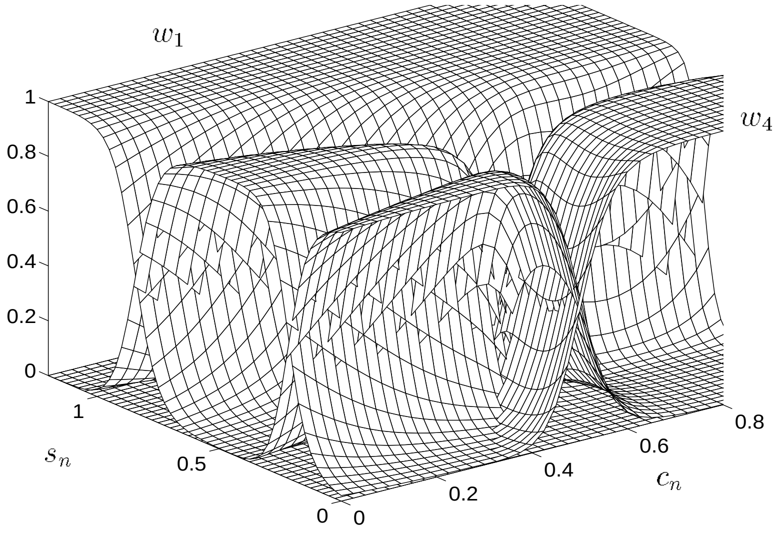


Figure 3.7: Interpolation functions $w_1$, $w_2$, $w_3$, and $w_4$ for the four local models of the fermenter, using $c$ and $s$ to characterize the operating regimes.

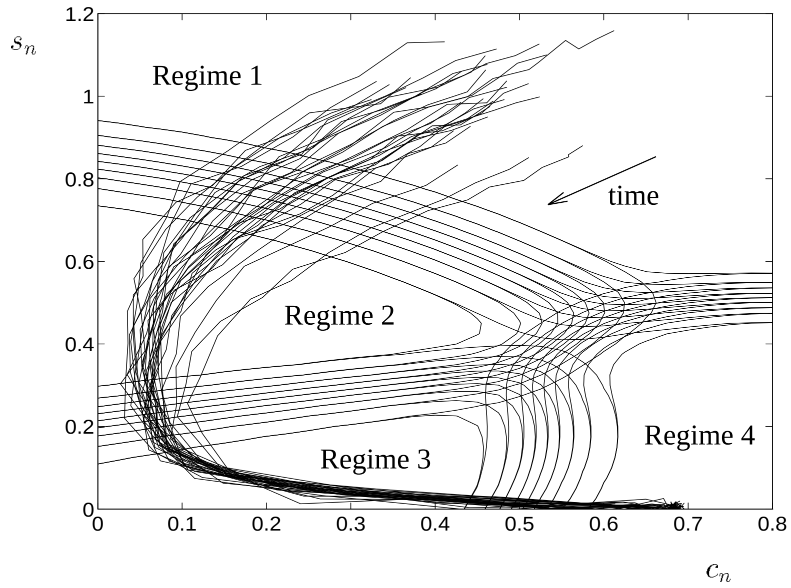


Figure 3.8: Evolution of the operating point $(c, s)$ for some batches. The trajectories start at high $s$ and $c$ and moves down to a low $s$ and high $c$ through the different regimes. The level curves for the interpolation functions indicate the regions of transition between the operating regimes.

### Results

The models are validated by simulation of the validation data. The results of simulations of a typical batch using the identified models are shown in figures 3.9-3.11, together with the evolution of the "true system" and a simulation using an identified global linear model $\dot{x} = a + Ax$

$$a = \begin{pmatrix} 0.2158 \\ 0.0045 \\ -0.0454 \\ -0.1311 \\ 0.1217 \end{pmatrix}, \quad A = \begin{pmatrix} -0.0805 & 0 & 0 & -0.0691 & -0.1976 \\ 0 & 0 & 0.1857 & 0 & 0 \\ -0.1768 & 0 & 0.0962 & 0.2544 & 0 \\ -0.0482 & 0 & 0 & -0.0706 & 0.2771 \\ 0.0631 & 0 & 0 & -0.2274 & -0.1028 \end{pmatrix}$$

In addition, the root average squared prediction errors estimated on the validation data are given in Tab. 3.2. We see that the two non-linear models always outperform the linear model, and the non-linear predictions are close to the evolution of the true system.

| Model | Average Error |
|---|---|
| Three local linear models, (OUR,$\chi$) as operating point | 0.0133 |
| Four local linear models, $(c, s)$ as operating point | 0.0134 |
| Global linear model | 0.0303 |

Table 3.2: Summary of simulation results, root average squared prediction error with the different models.

The process knowledge applied is primarily the knowledge of the overall reaction mechanism. Using this knowledge, operating regimes and local model structures are designed. The parameters of the local models are identified using a large data set that contains measurement of all states. Notice that we have used very limited knowledge about mass-transport, kinetics, and stoichiometry. It is therefore fair to say that the model structure is found on the basis of considerably less, and more qualitative, process knowledge compared to what would have been required to develop a mechanistic model.

## 3.4 Discussion

In this chapter we have focused on practical procedures for development of state-space models based on local models. The main results are

1. Combining local models with different state-spaces is less straightforward than when the state-spaces are the same. One must necessarily transfer state information between the different state spaces. For the more complicated cases, we suggest the application of a state-estimator as a general tool.
2. The operating regime based modeling framework typically gives transparent model structures that allow interpretation, analysis, and the incorporation of prior knowledge. The regimes can often be explicitly related to a particular set of phenomena, or different system behavior.

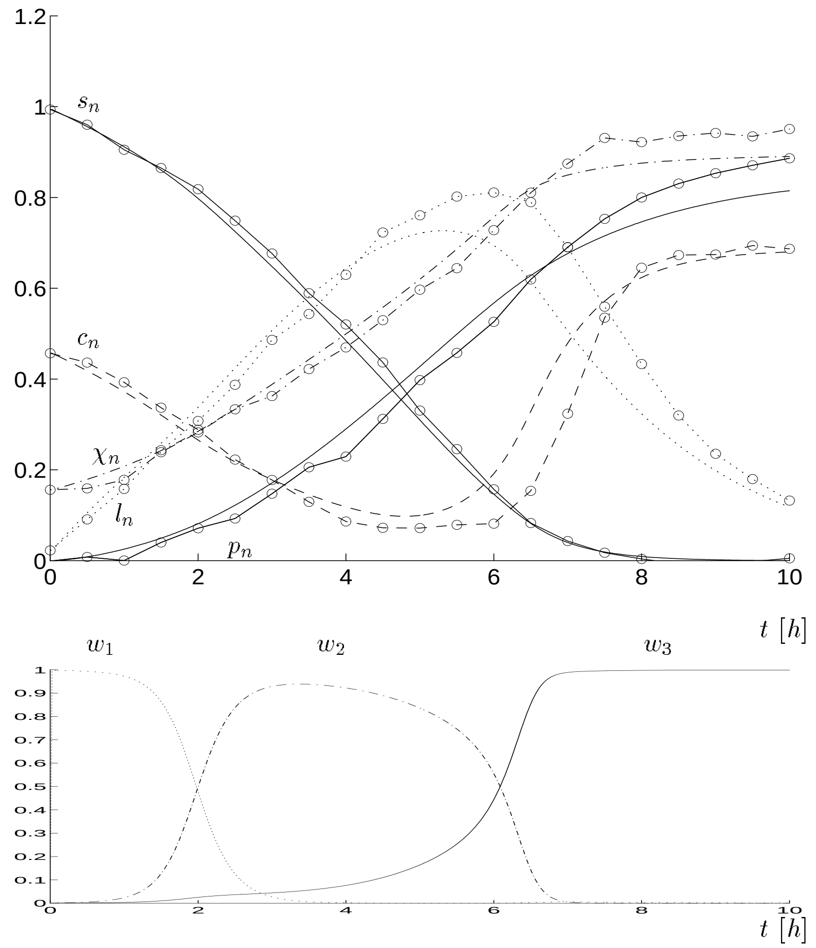


Figure 3.9: Top: Prediction using $OUR$ and $\chi$ as operating point. Trajectories with circles are generated by the "true system", while the other trajectories are simulations using the model. Bottom: The relative weight of the three local models in the interpolation.

3. We have illustrated by the means of simulation examples and philosophical arguments that the process knowledge required to design operating regimes and select local model structures may often be rather elementary and of quite qualitative nature.

It is interesting to compare this framework with the mechanistic modeling approach. The demand for process knowledge may be significantly less, and more qualitative and empirical knowledge can be incorporated directly.

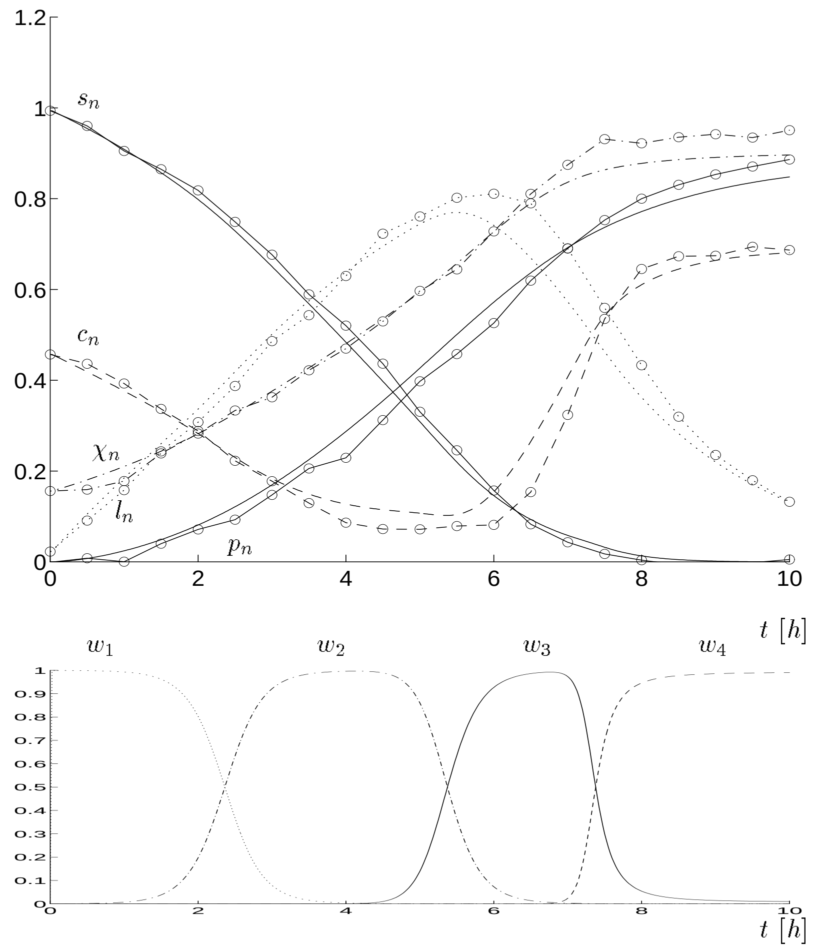


Figure 3.10: Top: Prediction using $c$ and $s$ as operating point. Trajectories with circles are generated by the "true system", while the other trajectories are simulations using the model. Bottom: The relative weight of the four local models in the interpolation.

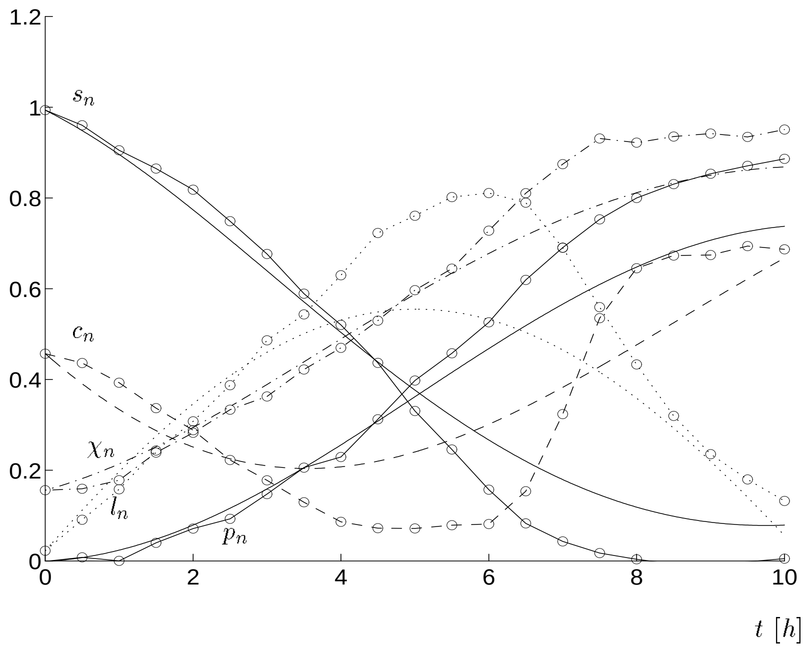


Figure 3.11: Prediction using global linear model. Trajectories with circles are generated by the "true system", while the other trajectories are simulations using the model.

# Chapter 4

# Model Structure Identification using Empirical Data from a Limited Operating Range

A fundamental assumption underlying most empirical modeling and identification methods is that the identification data sequence is complete, i.e. contains a sufficient amount of data from all relevant operating conditions (Ljung 1994). However, in many industries it is impossible to satisfy this condition, due to economic, engineering, or safety constraints on the operating conditions one is allowed to perform experiments under. Using standard system identification methods for the identification of model structure, without paying attention to the fact that a fundamental assumption is violated, may in such cases lead to an unreliable model, in the sense that its extrapolation capability can be poor. This problem is related to the robustness of the model structure parameterization, and identifiability. It is typical that mechanistic model structures are more robustly parameterized than empirical model structures.
In this chapter, we describe some criteria for model structure identification that with the help of some prior knowledge about the system's future operating range, is able to partially take into account this aspect. The application of such a criterion will improve the robustness of the structure identification algorithm and result in a more reliable model with better extrapolation capabilities. This procedure is also outlined in (Johansen and Foss 1994*a*).

## 4.1 Estimating Expected Prediction Performance with Bootstrapping

Let a candidate model structure $S$ be given. This model structure is essentially a set of equations (a combination of differential, difference and algebraic). The equations are parameterized by a parameter vector $\theta$ that is estimated using a standard prediction error estimator $\hat{\theta} = \hat{\theta}(\mathcal{D}_l)$, e.g. (Söderström and Stoica 1988), on the basis of a data sequence

$$\mathcal{D}_l = ((u(1), y(1)), (u(2), y(2)), ..., (u(l), y(l)))$$

The problem we address is to estimate on the basis of the data sequence $\mathcal{D}_l$ the expected (future) prediction performance of the model defined by $(S, \hat{\theta}(\mathcal{D}_l))$.

Let an unknown future data sequence be denoted $\mathcal{D}_t^\star$. The one-step-ahead prediction $\hat{y}(\mathcal{D}_{t-1}^\star; S, \hat{\theta}(\mathcal{D}_l))$ is computed by solving the model equations with the parameter estimate $\hat{\theta}(\mathcal{D}_l)$, either from a known or estimated state at time $t-1$. The state estimate may come from an extended Kalman-filter (Gelb 1974). Let us assume that the future output $y^\star(t)$ is made up of a deterministic component $y^\star(\mathcal{D}_{t-1}^\star)$ and an unpredictable white noise component $e(t)$:

$$y^\star(t) = y^\star(\mathcal{D}_{t-1}^\star) + e(t)$$

Clearly, the best we can hope for, is that the model can predict the deterministic component well. We define the expected squared prediction error using the model structure $S$ as

$$\begin{aligned}
\Sigma(S) &= E\left(y^\star(t) - \hat{y}(\mathcal{D}_{t-1}^\star; S, \hat{\theta}(\mathcal{D}_l))\right)\left(y^\star(t) - \hat{y}(\mathcal{D}_{t-1}^\star; S, \hat{\theta}(\mathcal{D}_l))\right)^T \\
&= E\left(y^\star(\mathcal{D}_{t-1}^\star) - \hat{y}(\mathcal{D}_{t-1}^\star; S, \hat{\theta}(\mathcal{D}_l)) + e(t)\right) \\
&\qquad \cdot \left(y^\star(\mathcal{D}_{t-1}^\star) - \hat{y}(\mathcal{D}_{t-1}^\star; S, \hat{\theta}(\mathcal{D}_l)) + e(t)\right)^T
\end{aligned}$$

where $E$ denotes expectation with respect to the joint probability distribution of the identification data sequence $\mathcal{D}_l$ and the future data sequence $\mathcal{D}_t^\star$ the model is to be applied to. Both $\mathcal{D}_l$ and $\mathcal{D}_t^\star$ are viewed as stochastic variables, so $\Sigma(S)$ is the ensemble average over all possible identification data sequences of length $l$ and future data sequences. Provided the future data $\mathcal{D}_t^\star$ is uncorrelated with the identification data $\mathcal{D}_l$, we get

$$\begin{aligned}
\Sigma(S) = E&\left(y^\star(\mathcal{D}_{t-1}^\star) - E\left(\hat{y}(\mathcal{D}_{t-1}^\star; S, \hat{\theta}(\mathcal{D}_l))\middle|\mathcal{D}_{t-1}^\star\right)\right) \\
&\cdot \left(y^\star(\mathcal{D}_{t-1}^\star) - E\left(\hat{y}(\mathcal{D}_t^\star; S, \hat{\theta}(\mathcal{D}_l))\middle|\mathcal{D}_{t-1}^\star\right)\right)^T \\
+E&\left(\hat{y}(\mathcal{D}_{t-1}^\star; S, \hat{\theta}(\mathcal{D}_l)) - E\left(\hat{y}(\mathcal{D}_{t-1}^\star; S, \hat{\theta}(\mathcal{D}_l))\middle|\mathcal{D}_{t-1}^\star\right)\right) \\
&\cdot \left(\hat{y}(\mathcal{D}_{t-1}^\star; S, \hat{\theta}(\mathcal{D}_l)) - E\left(\hat{y}(\mathcal{D}_{t-1}^\star; S, \hat{\theta}(\mathcal{D}_l))\middle|\mathcal{D}_{t-1}^\star\right)\right)^T \\
+E&\left(e(t)e^T(t)\right) \qquad (4.1)
\end{aligned}$$

Eq. (4.1) is a bias/variance decomposition of the expected squared prediction error. The first term is the systematic error (squared bias), that is non-zero if the model structure $S$ is not sufficiently rich to exactly describe the deterministic component of the system output. The second term is the random error (variance),

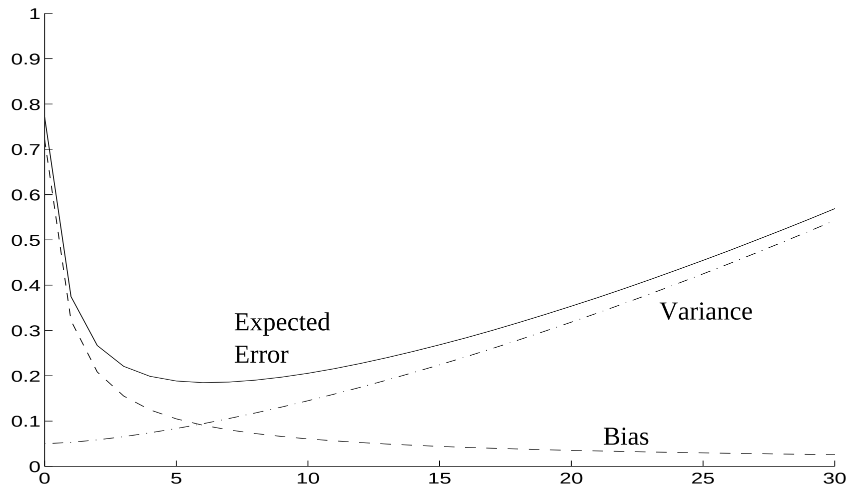


Figure 4.1: Typical relationship between bias, variance and model structure complexity, when it is assumed that model structure complexity can be measured by a real number.

which arises because the optimal parameters cannot be exactly determined on the basis of the finite data sequence $\mathcal{D}_l$. The third term is due to the noise, which is independent of the model and impossible to predict. Typically, if the length of the data sequence $\mathcal{D}_l$ is kept fixed, and the complexity of the model structure $S$ increases, the bias will decrease, while the variance will increase, see Fig. 4.1. The decrease in bias is a direct consequence of the increased degrees of freedom in the model structure, while the increase in variance is caused by the growing number of parameters in the model structure that must be estimated. If the model structure is over-parameterized, the parameter estimate will be highly sensitive to the particular realization of the identification data sequence, i.e. another data sequence of the same length logged under the same experimental conditions may lead to a very different parameter estimate. Hence, there is a tradeoff between bias and variance, which suggests that there exists a model with an optimal complexity within the set of model structures, cf. Fig 4.1.

Let us first consider the case when $\mathcal{D}_l^\star$ and $\mathcal{D}_l$ have the same underlying probability distribution. $\Sigma(S)$ will reflect the expected prediction performance of the model and would be suited as the basis for an identification criterion. Unfortunately, $\Sigma(S)$ cannot be computed, since we assume the underlying probability distribution is unknown. However, $\Sigma(S)$ can be estimated on the basis of the data sequence $\mathcal{D}_l$ using bootstrapping, e.g. (Efron and Tibshirani 1986, Carlstein 1992).

The bootstrap strategy is generally applicable, and the idea is simply to approximate the unknown probability distribution with a probability distribution that is estimated on the basis of the available empirical data. Suppose we want to estimate a statistic $T = T(p)$ that depends on an unknown probability distribution $p$.

The bootstrap estimate is $\widehat{T(p)} = T(\hat{p})$ where $\hat{p}$ is a parametric or non-parametric estimate of $p$.
In our case, a bootstrap estimate of the expected squared prediction error is

$$\hat{\Sigma}^{BOOT}(S) = \frac{1}{l}\sum_{t=1}^{l}\frac{1}{B}\sum_{b=1}^{B}\left(y(t) - \hat{y}(\mathcal{D}_{t-1}; S, \hat{\theta}_b)\right)\left(y(t) - \hat{y}(\mathcal{D}_{t-1}; S, \hat{\theta}_b)\right)^T$$

In this estimate, the underlying probability distribution is approximated by the empirical distribution of the residuals (Carlstein 1992). The expectation in the expression for $\Sigma(S)$ is approximated using Monte Carlo integration, and the estimated probability distribution. $B$ is the number of Monte Carlo runs. In more detail, suppose $\hat{\theta}(\mathcal{D}_l)$ is defined by

$$\hat{\theta}(\mathcal{D}_l) = \arg\min_{\theta}\frac{1}{l}\sum_{t=1}^{l}(y(t) - \hat{y}(\mathcal{D}_{t-1}; S, \theta))^T(y(t) - \hat{y}(\mathcal{D}_{t-1}; S, \theta))$$

In order to approximately mimic the random nature of the identification data, $\hat{\theta}_b$ is defined by

$$\hat{\theta}_b = \arg\min_{\theta}\frac{1}{l}\sum_{k=1}^{l}\left(y(t_k^b) - \hat{y}(\mathcal{D}_{t_k^b-1}; S, \theta)\right)^T\left(y(t_k^b) - \hat{y}(\mathcal{D}_{t_k^b-1}; S, \theta)\right)$$

where $t_1^b, t_2^b, ..., t_l^b$ are independent random samples with probability

$$P(t_k^b = i) = \begin{cases} 1/l, & i \in \{1, 2, ..., l\} \\ 0, & \text{otherwise} \end{cases}$$

In other words, $\hat{\theta}_b$ minimizes a criterion that is based on random resamples of the original residuals.
The criterion

$$J_{BOOT}(S) = \operatorname{trace}\left(\hat{\Sigma}^{BOOT}(S)\right)$$

is well suited for model structure identification when the future data sequence $\mathcal{D}_t^\star$ has the same underlying probability distribution as the identification data sequence $\mathcal{D}_l$.
An alternatice bootstrap estimate can be based on the bias/variance decomposition (4.1):

$$\begin{aligned}\hat{\Sigma}_{BV}^{BOOT}(S) = & \frac{1}{l}\sum_{t=1}^{l}\left(y(t) - \frac{1}{B_1}\sum_{b_1=1}^{B_1}\hat{y}(\mathcal{D}_{t-1}; S, \hat{\theta}_{b_1})\right) \\ & \cdot\left(y(t) - \frac{1}{B_1}\sum_{b_1=1}^{B_1}\hat{y}(\mathcal{D}_{t-1}; S, \hat{\theta}_{b_1})\right)^T \\ & + \frac{1}{l}\sum_{t=1}^{l}\frac{1}{B_2}\sum_{b_2=1}^{B_2}\left(\hat{y}(\mathcal{D}_{t-1}; S, \hat{\theta}_{b_2}) - \frac{1}{B_1}\sum_{b_1=1}^{B_1}\hat{y}(\mathcal{D}_{t-1}; S, \hat{\theta}_{b_1})\right) \\ & \cdot\left(\hat{y}(\mathcal{D}_{t-1}; S, \hat{\theta}_{b_2}) - \frac{1}{B_1}\sum_{b_1=1}^{B_1}\hat{y}(\mathcal{D}_{t-1}; S, \hat{\theta}_{b_1})\right)^T\end{aligned}$$

where $B_1$ and $B_2$ are the number of Monte Carlo runs and $\hat{\theta}_{b_1}$ and $\hat{\theta}_{b_2}$ are defined in the same way as $\hat{\theta}_b$. However, this is not recommended in practise, but will be a useful starting point in the following, when we will develop two modified criteria. Notice that the variance (second term) can be computed exactly in the special case when the predictor is linearly parameterized.

Let us now return to the general problem when the future data has a different underlying probability distribution than the identification data. A typical situation is when there are identification data available from a limited region of the full operating range only. It appears to be impossible to accurately estimate the bias of a model structure in some operating regime without any data or prior knowledge in that operating regime. However, we may indeed estimate the variance of the model structure over the entire range of operating conditions the model structure is defined for. Hence, we modify the estimate $\hat{\Sigma}_{BV}^{BOOT}$ such that the variance over the interesting range of operation is weighted according to the future probability distribution $p_{\text{future}}(\mathcal{D}_t^\star)$

$$
\begin{aligned}
\tilde{\Sigma}^{BOOT}(S) = & \frac{1}{l}\sum_{t=1}^{l}\left(y(t) - \frac{1}{B_1}\sum_{b_1=1}^{B_1}\hat{y}(\mathcal{D}_{t-1}; S, \hat{\theta}_{b_1})\right) \\
& \cdot \left(y(t) - \frac{1}{B_1}\sum_{b_1=1}^{B_1}\hat{y}(\mathcal{D}_{t-1}; S, \hat{\theta}_{b_1})\right)^T \\
& + \int_{\mathcal{D}_t^\star} \frac{1}{B_2}\sum_{b_2=1}^{B_2}\left(\hat{y}(\mathcal{D}_t^\star; S, \hat{\theta}_{b_2}) - \frac{1}{B_1}\sum_{b_1=1}^{B_1}\hat{y}(\mathcal{D}_t^\star; S, \hat{\theta}_{b_1})\right) \\
& \cdot \left(\hat{y}(\mathcal{D}_t^\star; S, \hat{\theta}_{b_2}) - \frac{1}{B_1}\sum_{b_1=1}^{B_1}\hat{y}(\mathcal{D}_t^\star; S, \hat{\theta}_{b_1})\right)^T p_{\text{future}}(\mathcal{D}_t^\star)d\mathcal{D}_t^\star
\end{aligned}
$$

In practise, this integral is also approximated using Monte Carlo integration. Compared to $\hat{\Sigma}_{BV}^{BOOT}$, the estimate of the term containing the bias is unchanged, but ideally we would also like to weight this term according to $p_{\text{future}}$, too. This is feasible only if an estimate of the probability distribution of the identification data $p_{\text{past}}$ is available:

$$
\begin{aligned}
\overline{\Sigma}^{BOOT}(S) = & \frac{1}{l}\sum_{t=1}^{l}\left(y(t) - \frac{1}{B_1}\sum_{b_1=1}^{B_1}\hat{y}(\mathcal{D}_{t-1}; S, \hat{\theta}_{b_1})\right) \\
& \cdot \left(y(t) - \frac{1}{B_1}\sum_{b_1=1}^{B_1}\hat{y}(\mathcal{D}_{t-1}; S, \hat{\theta}_{b_1})\right)^T \frac{p_{\text{future}}(\mathcal{D}_{t-1})}{p_{\text{past}}(\mathcal{D}_{t-1})} \\
& + \int_{\mathcal{D}_t^\star} \frac{1}{B_2}\sum_{b_2=1}^{B_2}\left(\hat{y}(\mathcal{D}_t^\star; \hat{\theta}_{b_2}) - \frac{1}{B_1}\sum_{b_1=1}^{B_1}\hat{y}(\mathcal{D}_t^\star; \hat{\theta}_{b_1})\right) \\
& \cdot \left(\hat{y}(\mathcal{D}_t^\star; \hat{\theta}_{b_2}) - \frac{1}{B_1}\sum_{b_1=1}^{B_1}\hat{y}(\mathcal{D}_t^\star; \hat{\theta}_{b_1})\right)^T p_{\text{future}}(\mathcal{D}_t^\star)d\mathcal{D}_t^\star
\end{aligned}
$$

## 4.2 Discussion

The above estimates provides at least a partial solution to the problem, and criteria like

$$\tilde{J}_{BOOT}(S) = \text{trace}\left(\tilde{\Sigma}^{BOOT}(S)\right)$$

or

$$\overline{J}_{BOOT}(S) = \text{trace}\left(\overline{\Sigma}^{BOOT}(S)\right)$$

will favor model structures that have low variance everywhere, and low bias in at least the region from which we have data available.

Since the criteria do not take into account the bias of the model structure in regimes from which there is no data available, it is important that these criteria are applied with care, and that a subjective assessment of this bias is taken into consideration when the model structure is selected.

Bootstrapping with sequential structure in the data may be complicated, see (Carlstein 1992) for a survey of available techniques that more or less preserve the sequential structure in the resample. It is shown that bootstrapping using the residuals, as applied here, at least gives a consistent estimate for AR processes (Freedman 1984). However, a strong theoretical justification of the application of bootstrapping in cases with non-stationary and sequentially correlated stochastic processes is not available, to our knowledge. It is the general applicability and simple computer implementation that has made bootstrapping a very popular statistical technique.

The main "user's choice" is the future data probability distribution $p_{\text{future}}(\mathcal{D}_t^\star)$. For theoretical developments we are usually concerned with identification criteria that impose on the error according to the data distribution, because it allows us to derive nice results. While such weighting may be reasonable in many applications, it is not difficult to imagine cases where this is far from what we want. A typical example is a safety critical process, where on one hand there may not be much data available from certain critical operating conditions, while these are the conditions under which the model should be most reliable. The distribution $p_{\text{future}}(\mathcal{D}_t^\star)$ is a convenient tool for at least partially take such considerations into account, when it is interpreted as a weight-function for desired model accuracy.

Unfortunately, $p_{\text{future}}(\mathcal{D}_t^\star)$ is usually not known, and only a very rough approximation may be available. This is the major drawback with this approach, in our view. However, on the few examples we have tested this procedure on, it is clear that improvements are achieved with a rough choice of $p_{\text{future}}(\mathcal{D}_t^\star)$ compared to neglecting the fact that the future operating range is larger that reflected in the identification data.

Parametric and non-parametric estimation of density functions like $p_{\text{past}}(\mathcal{D}_t^\star)$ is a topic that is treated in detail in the literature, see e.g. (Silverman 1986). We will therefore not dwell with this problem here.

# Chapter 5

# Model Structure Identification using Separate Validation Data - Asymptotic Properties

The decomposition of a model set into a parametric and a structural level is common. The parameters are typically identified using a prediction error criterion, e.g. (Ljung 1987), while there are several possible criteria for structure identification. These include FPE (Akaike 1969), AIC (Akaike 1974), MDL (Rissanen 1978), Bayesian criteria (Kashyap 1977), cross-validation (Stoica, Eykhoff, Janssen and Söderström 1986), and the unbiasness criterion (Ivakhnenko and Yurachkovsky 1988). Common for all these criteria is that they try to estimate the expected prediction performance when the identified models of different structures are used for prediction. However, the simplest and perhaps most popular such estimate is the use of separate validation data to assess the prediction performance of the different identified models. The major drawback of this criterion, compared to the above mentioned ones, is that a longer data sequence may be needed.

Suppose a bounded sequence of $2l$ input/output observations from a dynamical system is available:

$$Z_{2l} = (z_1, z_2, ..., z_{2l})$$

where $z_t = (u_t, y_t)$, $u_t \in R^r$ is the input vector, and $y_t \in R^m$ is the output vector, at time $t$. The problem we address is the one of identifying a model of the system on the basis of these observations, and in particular study the model as $l$ approaches infinity. The algorithm we will study is based on hierarchical optimization, since the model set is decomposed into two levels, namely model structure and model parameters. Referring to Fig. 1.2, we will in this chapter assume that a model set is given. We will present a novel convergence result for the hierarchical identification procedure based on separate validation data, under similar assumptions as in the asymptotic analysis of the parameter identification problem given by Ljung (1978). Indeed, such a result should be expected, since the use of separate validation data should give a good estimate of the expected prediction performance. It should be noted that some asymptotic results are given for related criteria based on separate validation data in (Stepashko 1988, Aksenova 1989), in a significantly less general context than this chapter. In particular, these results only hold for a finite number of model structures, and the observations are assumed to be independent.

For easy reference, the telations between the assumptions and propositions in this chapter are as illustrated in Fig. 5.1. The material in this chapter is from (Johansen and Weyer 1995).

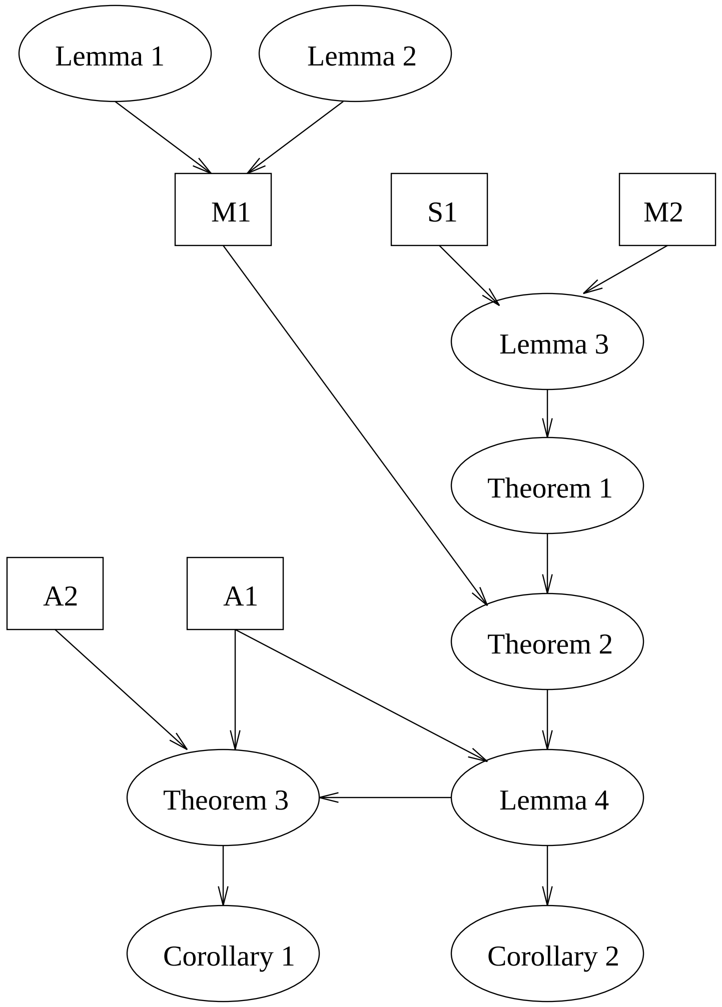


Figure 5.1: How the assumptions and propositions are related.

## 5.1 Main Assumptions and Preliminary Results

Let $\mathcal{S}$ be a set of model structures. A model structure $S \in \mathcal{S}$ is a set of possibly non-linear equations, which may be algebraic, difference, differential, or a combination. The equations may contain some unknown parameters $\theta$ that are known to belong to a parameter set $\Theta_S$, which is a subset of a finite-dimensional Euclidean vector space, by assumption. We will not explicitly require $\Theta_S$ to be compact, but as we shall see, this is typically needed to satisfy the assumptions. This defines a model set

$$\mathcal{M} = \{M = (S, \theta) \mid S \in \mathcal{S}, \theta \in \Theta_S\}$$

Suppose that for each model $M \in \mathcal{M}$ we are able to make a prediction $\hat{y}_t(M, Z_{t-1})$ of the output $y_t$ on the basis of the data sub-sequence $Z_{t-1}$. Notice that the dependence of the model $M$ on the predictor may be arbitrary. For example, the predictor may be based on a suboptimal state-estimator. The one-step-ahead prediction error using the model $M$ is defined by $\varepsilon_t(M, Z_t) = y_t - \hat{y}_t(M, Z_{t-1})$. In the following, we may let the dependence on $Z_{t-1}$ be implicit when it is more convenient, and use the notation $\hat{y}_t(M)$ and $\varepsilon_t(M)$.

Before we proceed, we need to introduce some notation. Let the $\gamma$-norm of a vector $x \in R^n$ be defined by $||x||_\gamma = (|x_1|^\gamma + ... + |x_n|^\gamma)^{1/\gamma}$ for all $\gamma \geq 1$, and the Euclidean norm be denoted by $||x|| = ||x||_2$.

The following assumption states the key technical property of the model set. It essentially limits the complexity of the model set, in the sense that there must exists a finite number of good approximations to all models in the model set.

**Assumption M1.** The model set $\mathcal{M}$ can be covered by a finite $\epsilon$-net with $N(\epsilon)$ elements, i.e. for any $\epsilon > 0$ there exists a finite set of models $\mathcal{M}_\epsilon = \{M_1, M_2, ..., M_{N(\epsilon)}\}$ (not necessarily a subset of $\mathcal{M}$) such that for any $M \in \mathcal{M}$ there exists an $M_i(M) \in \mathcal{M}_\epsilon$ with the property

$$\sup_{t, Z_{t-1}} ||\hat{y}_t(M, Z_{t-1}) - \hat{y}_t(M_i(M), Z_{t-1})|| \leq \epsilon \tag{5.1}$$

Please apologize for possible confusion by using $\epsilon$ and $\varepsilon$ in two different meanings.

□

Before we proceed, let us discuss the limitations imposed by assumption M1 on the model set.

**Lemma 1** *Suppose*

1. *$\mathcal{S}$ contains only a finite number of model structures.*
2. *For every $S \in \mathcal{S}$ the parameter set $\Theta_S$ is compact.*
3. *For every $S \in \mathcal{S}$ the predictor $\hat{y}_t(S, \theta, Z_{t-1})$ is uniformly continuous in $\Theta_S$.*

*Then $\mathcal{M}$ can be covered by a finite $\epsilon$-net.*

**Proof.** Let $S \in \mathcal{S}$ be arbitrary. From the uniform continuity there exists for every $\epsilon > 0$ a $\delta > 0$ that depends only on $\epsilon$, with the property that

$$||\theta_1 - \theta_2|| < \delta \text{ implies } \sup_{t, Z_{t-1}} ||\hat{y}_t(S, \theta_1, Z_{t-1}) - \hat{y}_t(S, \theta_2, Z_{t-1})|| < \epsilon \qquad (5.2)$$

uniformly in $\Theta_S$. Since $\Theta_S$ is compact, it can be covered by a finite number of $\delta$-balls. The parameter vectors at the center of these $\delta$-balls define an $\epsilon$-net for the model set $\mathcal{M}_S = \{M = (S, \theta) \mid \theta \in \Theta_S\}$. Since $\mathcal{S}$ contains only a finite number of model structures, the result follows.

□

Uniform continuity may exclude certain types of models and systems. Consider for example the simple model (predictor) $\hat{y}_t(S, \theta, Z_{t-1}) = \theta y_{t-1}$. We get

$$\sup_{t, Z_{t-1}} ||\hat{y}_t(S, \theta_1, Z_{t-1}) - \hat{y}_t(S, \theta_2, Z_{t-1})|| \leq \delta \sup_t ||y_{t-1}|| \; < \; \epsilon \qquad (5.3)$$

Since $\delta$ cannot depend on $t$, the system output must be bounded in order for uniform continuity to hold. However, we are free to choose the predictors ourselves, and may include for example saturation effects for very large outputs in the predictor to enforce this.

**Lemma 2** *Suppose*

1. *For all* $t$, $z_t \in \mathcal{Z}$, *where* $\mathcal{Z}$ *is compact.*
2. *There exists a constant* $K$ *such that*

$$\sup_{t, Z_{t-1}, M} ||\hat{y}_t(M, Z_{t-1})|| \leq K \qquad (5.4)$$

3. *For every* $k \geq 1$ *there exists a uniformly continuous function* $f_k : \mathcal{M} \times \mathcal{Z}^k \longrightarrow R^m$ *with the property*

$$\sup_{t, Z_{t-1}} ||\hat{y}_t(M, z_{t-1}, ..., z_1) - f_k(M, z_{t-1}, ..., z_{t-k})|| \leq C\lambda^k \qquad (5.5)$$

*for some constants* $C$ *and* $\lambda \in (0, 1)$.

*Then* $\mathcal{M}$ *can be covered by a finite* $\epsilon$*-net.*

**Proof.** Let $M \in \mathcal{M}$ and $\epsilon > 0$ be arbitrary. Choose $k$ so large that

$$\sup_{t, Z_{t-1}} ||\hat{y}_t(M, z_{t-1}, ..., z_1) - f_k(M, z_{t-1}, ..., z_{t-k})|| \leq \epsilon/2 \qquad (5.6)$$

Provided the set of functions $\mathcal{F}_k = \{f_k(M, \cdot) : \mathcal{Z}^k \rightarrow R^m \mid M \in \mathcal{M}\}$ can be covered with a finite $\epsilon/2$-net, this $\epsilon/2$-net will also be an $\epsilon$-net for $\mathcal{M}$.

From the uniform continuity of $f_k$ there exists a $\delta > 0$ such that

$$||Z^1 - Z^2|| < \delta \quad \text{implies} \quad \sup_{M \in \mathcal{M}} ||f_k(M, Z^1) - f_k(M, Z^2)|| < \epsilon/2 \qquad (5.7)$$

Cover $\mathcal{Z}^k$ with a finite number of boxes $\mathcal{Z}_1, \mathcal{Z}_2, ..., \mathcal{Z}_N$, where each box is contained within a $\delta$-ball, and

$$\mathcal{Z} \subset \bigcup_i \mathcal{Z}_i \quad \text{and} \quad \mathcal{Z}_i \cap \mathcal{Z}_j = \emptyset \ \text{ for } i \neq j \tag{5.8}$$

An approximation $\hat{f}_k(M, \cdot)$ to $f_k(M, \cdot)$ is now

$$\hat{f}_k(M, Z) = \sum_{i=1}^{N} \beta_i(M)\mu_i(Z)$$

where $\mu_i$ is the characteristic function for $\mathcal{Z}_i$. From (5.7) and assumption 2, it is clear that a finite number of values for the parameter vector $\beta_i$ is sufficient to achieve a uniform approximation accuracy of $\epsilon/2$. Hence, this defines a finite $\epsilon/2$-net for $\mathcal{F}_k$.

**Remarks.** It will later be required that the $\epsilon$-net for $\mathcal{M}$ must consists of predictors that are smooth functions of past data, cf. M2. This may be achieved by replacing the $\mu_i$-functions by smoothed versions. It should be mentioned that the use of a piecewise constant approximation in the proof is only one possibility, and several other approximators are expected to work as well.

□

Assumption 3 requires that the dependence of the predictor on past data must decay exponentially. Roughly speaking, this means that the predictor must be exponentially stable. Moreover, the predictor must be time-invariant, which is a quite restrictive assumption that need not hold when the predictor contains a state estimator, for example. On the other hand, boundedness of the data and prediction (cf. assumptions 1 and 2) are quite reasonable assumptions.

We would also like to stress that Lemma 1 and 2 are only examples of sufficient conditions for the existence of $\epsilon$-nets that cover $\mathcal{M}$, and weaker conditions may very well exist. It must also be emphasized that it is only sufficient to prove the existence of an $\epsilon$-net. However, as we have seen, the proofs are typically constructive.

The asymptotic analysis will be based on a law of large numbers that holds uniformly in the model set. We will in the next paragraphs state a set of assumptions that together with M1 ensure that such a law holds. We will introduce a stochastic framework that allows us to deal with non-stationary stochastic processes that may contain deterministic components and sequential correlations. In the following, let $E$ denote expectation with respect to input/output observations, and following Ljung (1987) we define the operator $\overline{E}$ by

$$\overline{E}x_t = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} Ex_t$$

for a vector sequence $x = (x_t)$. Notice that it is implicitly assumed that the limit exists in all subsequent applications of this operator.

**Definition Ljung (1978)/Hjalmarsson (1993).** A sequence $(x_t)$ is said to be exponentially forgetting of order $\gamma$ if there exists a doubly indexed sequence $(x_t^s)$ where $x_t^s$ is independent of the subsequences $(x_0, x_1, ..., x_s)$ and $(x_p^p, x_{p+1}^p, ..., x_s^p)$ for all $p \leq s$, $x_t^t = 0$, and for every $t \geq s$

$$E\,||x_t - x_t^s||_\gamma^\gamma \leq C\lambda^{t-s} \tag{5.9}$$

for some constants $C$ and $\lambda \in (0,1)$.

□

This means that it must be possible to approximate the sequence $(x_t)$ with a sequence $(x_t^s)$ that is independent of the remote past. Moreover, the approximation must be so good that the "remote past is forgotten at an exponential rate". Ljung (1978) uses this as a definition of exponential stability of stochastic dynamical systems, and has shown that in many cases it contains the classical definition of exponential stability. Clearly, this definition also has similarities to assumption 3 in Lemma 2. The definition is made in a stochastic framework using the expectation operator, while in Lemma 2 the assumption is in a deterministic framework using the supremum norm.

**Assumption S1.** The sequences $(u_t)$ and $(y_t)$ are exponentially forgetting of order 4.

□

On the model set $\mathcal{M}$ and the $\epsilon$-net $\mathcal{M}_\epsilon$ we need to make an additional assumption to ensure that the squared prediction error is an exponentially forgetting sequence when S1 holds.

**Assumption M2.** Define the sequence $h(M) = (||\varepsilon_t(M)||^2)$, and suppose

1. The function $\hat{y}_t(M, Z_{t-1})$ is differentiable with respect to $Z_{t-1}$ for all $M \in \mathcal{M} \cup \mathcal{M}_\epsilon$.
2. For all $t \geq 0$ and $k \geq 1$, there exists a convex function $g_{t,k}$ such that for all models $M \in \mathcal{M} \cup \mathcal{M}_\epsilon$
$$\left|\left|\frac{\partial}{\partial z_{t-k}} h_t(M, z_{t-1}, ..., z_1)\right|\right| \leq g_{t,k}(z_{t-1}, ..., z_1)$$
3. There exist constants $C$ and $\lambda \in (0,1)$ such that for all $t \geq 0$ the set of functions $\{g_{t,k} \mid k \geq 1\}$ has the property
$$E||z_t||^4 < \infty \quad \text{for all } t \quad \text{implies} \quad E(g_{t,k}(z_{t-1}, ..., z_1))^4 \leq C\lambda^k$$
4. Moreover, $(h_t(M, 0, 0, ..., 0))$ is a bounded sequence for and all $M \in \mathcal{M} \cup \mathcal{M}_\epsilon$.

□

First, assumption M2 means that the sensitivity of the squared prediction error with respect to the past inputs and outputs $z_{t-k}$ must be bounded by a function that decays exponentially with $k$, which basically means that the predictor must be exponentially stable. In addition, M2 impose some growth conditions on the predictor. It is clearly sufficient that the derivative of the predictor is bounded. Furthermore, the prediction must depend smoothly on the past data. These are reasonable conditions that are likely to be fulfilled for many model sets, see also (Hjalmarsson 1993), p. 107.

**Lemma 3** *(Hjalmarsson 1993), pp. 46-48. Suppose Assumptions S1 and M2 hold, then for all $M \in \mathcal{M} \cup \mathcal{M}_\epsilon$ the sequence $(||\varepsilon_t(M)||^2)$ is exponentially forgetting of order 2. Moreover, the sequence $(\sup_{M \in \mathcal{M}} ||\varepsilon_t(M)||^2)$ is also exponentially forgetting of order 2.*

□

Exponential forgetting of order 2 is sufficient for the strong law of large numbers to hold:

**Theorem 1 (Law of Large Numbers.)** *Suppose a scalar sequence $(h_t)$ is exponentially forgetting of order 2. Then*

$$\left|\overline{E}h_t - \frac{1}{l}\sum_{t=1}^{l} h_t\right| \rightarrow \quad w.\ p.\ 1\ as \quad l \rightarrow \infty$$

**Proof.** Follows from a somewhat stronger result in (Hjalmarsson 1993), p. 66.

□

This law is a prerequisite for a uniform law of large numbers, which is the main tool in the asymptotic analysis.

**Theorem 2 (Uniform Law of Large Numbers.)** *Suppose $\mathcal{H}$ is a set of scalar sequences that are exponentially forgetting of order 2, and for each $\epsilon > 0$ there exists a finite set of exponentially forgetting sequences $\mathcal{H}_\epsilon$ containing lower and upper approximations $h^{\epsilon,L}$ and $h^{\epsilon,U}$ to each $h \in \mathcal{H}$ such that*

$$h_t^{\epsilon,L} \leq h_t \leq h_t^{\epsilon,U} \quad and \quad \overline{E}(h_t^{\epsilon,U} - h_t^{\epsilon,L}) < \epsilon \tag{5.10}$$

*Then*

$$\sup_{h \in \mathcal{H}} \left|\overline{E}h_t - \frac{1}{l}\sum_{t=1}^{l} h_t\right| \rightarrow 0\ w.\ p.\ \ 1\ as \quad l \rightarrow \infty \tag{5.11}$$

**Proof.** Follows from a theorem in (Pollard 1984), p. 8, and Theorem 1.

□

## 5.2 Structural and Parametric Identification

We define the expected squared prediction error

$$\overline{V}(M) = \overline{E}\,||\varepsilon_t(M)||^2 \tag{5.12}$$

The ultimate goal is to find a model in $\mathcal{M}$ that makes $\overline{V}$ as small as possible. To better understand the nature of this problem, consider

$$\overline{E}_Z\overline{V}(\hat{M}_{2l}) = \overline{E}_Z\overline{E}\left|\left|\varepsilon_t(\hat{M}_{2l})\right|\right|^2 \tag{5.13}$$

where $\hat{M}_{2l} = (\hat{S}_{2l}, \hat{\theta}_{2l}) \in \mathcal{M}$ is a model estimated on the basis of the data $Z_{2l}$ using an arbitrary method, and the expectation $\overline{E}_Z$ is with respect to the data sequence $Z_{2l}$ used for identification, that is, the average over the ensemble of possible data sequences $Z_{2l}$ of fixed length $2l$. The composite operator $\overline{E}_Z\overline{E}$ will in addition average over the ensemble of possible (future) data in order to assess the expected squared prediction error. In other words, $\overline{E}_Z\overline{V}(\hat{M}_{2l})$ will be the expected squared prediction error of the model $\hat{M}_{2l}$ when both the fixed length data sequence $Z_{2l}$ used for identification and the future data are viewed as random variables. We have the following bound on the expected squared prediction error:

$$\begin{aligned}\overline{E}_Z\overline{V}(\hat{M}_{2l}) \leq\ & \overline{E}3\,||y_t - \hat{y}_t(M^\star)||^2 \\ & +\overline{E}_Z\overline{E}3\left|\left|\hat{y}_t(S^\star,\theta^\star) - \hat{y}_t(\hat{S}_{2l},\theta^\star_{2l})\right|\right|^2 \\ & +\overline{E}_Z\overline{E}3\left|\left|\hat{y}_t(\hat{S}_{2l},\theta^\star_{2l}) - \hat{y}_t(\hat{S}_{2l},\hat{\theta}_{2l})\right|\right|^2 \end{aligned} \tag{5.14}$$

where $M^\star = (S^\star, \theta^\star)$ is the (not necessarily unique) best model in $\mathcal{M}$, in the sense that it minimizes $\overline{V}(M)$ over $\mathcal{M}$. Likewise, $\theta^\star_{2l}$ is the (not necessarily unique) best parameter vector for the structure $\hat{S}_{2l}$, in the sense that it minimizes $\overline{V}(M)$ over the set of models

$$\mathcal{M}_{\hat{S}_{2l}} = \left\{M = (\hat{S}_{2l}, \theta)\ \middle|\ \theta \in \Theta_{\hat{S}_{2l}}\right\}$$

induced by the model structure $\hat{S}_{2l}$. The first term in this bias/variance decomposition is the variance of the residuals that are impossible to predict using any model in $\mathcal{M}$, both due to unpredictable random error and because the model set $\mathcal{M}$ may not be sufficiently rich to exactly reproduce the deterministic component of the system output. The second term is the error caused by a possibly too simple model structure $\hat{S}_{2l}$ (bias), while the third term is variance due to parametric error. In general, model structures that are too complex will give large parametric errors. In particular, models with more parameters than the number of observations used to fit the model may give unbounded parametric error. Clearly, it is desirable with both small structural and parametric errors, but for a finite $l$, these are mutually exclusive, cf. Fig 4.1. A small structural error will in general require a complex model structure, while a small parametric error will require a simple model structure with few parameters compared to $l$. This is the well known bias/variance trade-off. However, asymptotically (as $l \to \infty$) it is possible to simultaneously reduce both these terms. This requires that the model is chosen to give a delicate balance between structural and parametric error.

Minimization of the expected squared prediction error $\overline{V}(M)$ will in many cases reflect the desired properties of the model, and is therefore suited as a criterion for model identification. Unfortunately, the underlying probability measure is assumed to be unknown, and the expectation cannot be computed. Consider instead the empirical average prediction error

$$V_l(S,\theta) = \frac{1}{l}\sum_{t=1}^{l} ||\varepsilon_t(S,\theta)||^2 \tag{5.15}$$

estimated on the basis of the first half of the data sequence $Z_{2l}$. Within the hierarchical identification framework in Fig. 1.2, we minimize for each model structure $S \in \mathcal{S}$ the criterion $V_l(S, \theta)$ with respect to $\theta$ in order to find a parameter estimate.

**Assumption A1.** For every $l$ and $S \in \mathcal{S}$ the following global minimum exists with probability one

$$\hat{\theta}_{S,l} = \arg \min_{\theta \in \Theta_S} V_l(S, \theta)$$

□

Notice that we do not assume that the minimum is unique, but use the convention that if more than one value of $\theta$ minimizes $V_l(S, \theta)$, then is $\hat{\theta}_{S,l}$ an arbitrary minimum. Non-uniqueness should indeed be expected, because $\mathcal{M}$ may contain model structures of high complexity, possibly with an unbounded number of parameters. If this is the case, then for $l$ smaller than the number of parameters in the model structure, the parameter identification problem is under-determined and there may be several parameter vectors that give exact match to the data. Notice that A1 automatically holds if $\Theta_S$ is compact, and $V_l$ is a continuous function of $\theta$.

We define $\hat{M}_{S,l} = (S, \hat{\theta}_{S,l})$. This is the (not necessarily unique) model with structure $S$ that on the basis of the $l$ observations is believed to be the best. Next, we need to choose among the models in the set $\{\hat{M}_{S,l} \mid S \in \mathcal{S}\}$. We define a new criterion on the basis of the second half of the data sequence $Z_{2l}$

$$V_l'(S, \hat{\theta}_{S,l}) = \frac{1}{l-k} \sum_{t=l+k+1}^{2l} \left|\left| \varepsilon_t(S, \hat{\theta}_{S,l}) \right|\right|^2 \tag{5.16}$$

throwing away the first $k$ samples to ensure that the remaining samples are not too much correlated with $\hat{\theta}_{S,l}$, cf. the definition of exponentially forgetting sequences.

**Assumption A2.** For every $l$, the global minimum

$$\hat{S}_{2l} = \arg \min_{S \in \mathcal{S}} V_l'(S, \hat{\theta}_{S,l})$$

exists with probability one.

□

Again, this minimum need not be unique, and we use the convention that in the case of non-uniqueness, $\hat{S}_{2l}$ is any minimum. In summary, the model estimate $\hat{M}_{2l}$ is defined by the following hierarchical optimization procedure

$$\text{Parameter Identification:} \quad \hat{\theta}_{S,l} = \arg \min_{\theta \in \Theta_S} V_l(S, \theta) \quad \text{for all } S \in \mathcal{S} \tag{5.17}$$

$$\text{Structure Identification:} \quad \hat{S}_{2l} = \arg \min_{S \in \mathcal{S}} V_l'(S, \hat{\theta}_{S,l}) \tag{5.18}$$

$$\text{Identified Model:} \quad \hat{M}_{2l} = (\hat{S}_{2l}, \hat{\theta}_{\hat{S}_{2l},l}) \tag{5.19}$$

We are now in position to prove the main result:

**Theorem 3** *Suppose Assumptions A1 and A2 hold, and*

$$\sup_{M\in\mathcal{M}} \left|V_l(M) - \overline{V}(M)\right| \to 0 \quad w.\ p.\ 1 \quad as \quad l\to\infty \tag{5.20}$$

$$\sup_{S\in\mathcal{S}} \left|V_l'(S,\hat{\theta}_{S,l}) - \overline{V}(S,\hat{\theta}_{S,l})\right| \to 0 \quad w.\ p.\ 1 \quad as \quad l\to\infty \tag{5.21}$$

*Then*

$$\overline{V}(\hat{M}_{2l}) \to \inf_{M\in\mathcal{M}} \overline{V}(M) \quad w.\ p.\ 1 \quad as \quad l\to\infty \tag{5.22}$$

**Proof.** Successive applications of the triangle inequality gives

$$\begin{aligned}
\left|\overline{V}(\hat{M}_{2l}) - \inf_{M\in\mathcal{M}} \overline{V}(M)\right| &= \left|\overline{V}(\hat{M}_{2l}) - \inf_{S\in\mathcal{S}}\inf_{\theta\in\Theta_S} \overline{V}(S,\theta)\right| \\
&\le \left|\overline{V}(\hat{M}_{2l}) - \inf_{S\in\mathcal{S}} \overline{V}(S,\hat{\theta}_{S,l})\right| \\
&\quad + \left|\inf_{S\in\mathcal{S}} \overline{V}(S,\hat{\theta}_{S,l}) - \inf_{S\in\mathcal{S}}\inf_{\theta\in\Theta_S} \overline{V}(S,\theta)\right| \\
&\le \left|\overline{V}(\hat{M}_{2l}) - V_l'(\hat{M}_{2l})\right| \\
&\quad + \left|V_l'(\hat{M}_{2l}) - \inf_{S\in\mathcal{S}} V_l'(S,\hat{\theta}_{S,l})\right| \\
&\quad + \left|\inf_{S\in\mathcal{S}} V_l'(S,\hat{\theta}_{S,l}) - \inf_{S\in\mathcal{S}} \overline{V}(S,\hat{\theta}_{S,l})\right| \\
&\quad + \left|\inf_{S\in\mathcal{S}} \overline{V}(S,\hat{\theta}_{S,l}) - \inf_{S\in\mathcal{S}} V_l(S,\hat{\theta}_{S,l})\right| \\
&\quad + \left|\inf_{S\in\mathcal{S}} V_l(S,\hat{\theta}_{S,l}) - \inf_{S\in\mathcal{S}}\inf_{\theta\in\Theta_S} V_l(S,\theta)\right| \\
&\quad + \left|\inf_{S\in\mathcal{S}}\inf_{\theta\in\Theta_S} V_l(S,\theta) - \inf_{S\in\mathcal{S}}\inf_{\theta\in\Theta_S} \overline{V}(S,\theta)\right|
\end{aligned}$$

The second and fifth terms are zero, with probability one, cf. (5.17) and (5.18). It is straightforward to prove that for any set $\mathcal{X}$, and functions $f,g:\mathcal{X}\to R$ that are bounded from below, that

$$\left|\inf_{x\in\mathcal{X}} f(x) - \inf_{x\in\mathcal{X}} g(x)\right| \le \sup_{x\in\mathcal{X}} |f(x)-g(x)| \tag{5.23}$$

Hence, with probability one

$$\begin{aligned}
\left|\overline{V}(\hat{M}_{2l}) - \inf_{M\in\mathcal{M}} \overline{V}(M)\right| &\le 2\sup_{S\in\mathcal{S}} \left|V_l'(S,\hat{\theta}_{S,l}) - \overline{V}(S,\hat{\theta}_{S,l})\right| \\
&\quad +2\sup_{S\in\mathcal{S}}\sup_{\theta\in\Theta_S} \left|V_l(S,\theta) - \overline{V}(S,\theta)\right|
\end{aligned} \tag{5.24}$$

□

The result tells us that the expected squared prediction error when the model $\hat{M}_{2l}$ is used for prediction will tend towards the smallest possible with probability one as $l\to\infty$. In particular,

**Corollary 1** *Suppose the assumptions in Theorem 3 hold, and in addition*

1. *There exists a unique $M^\star \in \mathcal{M}$ such that $M^\star = \arg\min_{M \in \mathcal{M}} \overline{V}(M)$.*
2. *The parameterizations of the model structures are such that a metric on $\mathcal{M}$ can be defined, and $\mathcal{M}$ is compact.*

*Then $\hat{M}_{2l} \to M^\star$ with probability one as $l \to \infty$.*

□

This result is commonly defined as consistency of the model estimate. The conditions are related to identifiability of the model set.

Next, we discuss assumptions under which (5.20) and (5.21) hold.

**Lemma 4** *Suppose S1, M1, M2 and A1 hold, then (5.20) and (5.21) hold.*

**Proof of (5.20):** The proof is inspired by a result in (Weyer, Williamson and Mareels 1994). Recall the definition of the sequence $h(M) = (||\varepsilon_t(M)||^2)$ and define the set of sequences $\mathcal{H} = \{h(M) \mid M \in \mathcal{M}\}$. Of course, the main tool is Theorem 2, and the proof is basically to show that $h(M)$ satisfies the assumptions of this theorem.

Pick an arbitrary $\epsilon > 0$ and $M \in \mathcal{M}$, and let $M_i(M) \in \mathcal{M}_{\epsilon'}$ denote the model in the $\epsilon'$-net $\mathcal{M}_{\epsilon'}$ that is closest to $M$ (in the sense of (5.1)), where $\epsilon'$ is yet unspecified. Now

$$h_t(M_i(M)) - 2\epsilon'\sqrt{h_t(M_i(M))} \leq h_t(M) \leq \left(\sqrt{h_t(M_i(M))} + \epsilon'\right)^2 \tag{5.25}$$

and since $\sqrt{x} \leq 1 + x$, lower and upper approximations to the sequence $h(M)$ are given by

$$\begin{aligned} h_t^{\epsilon,L} &= h_t(M_i(M)) - 2\epsilon'(1 + h_t(M_i(M))) \leq h_t(M) \\ &\leq h_t(M_i(M)) + 2\epsilon'(1 + h_t(M_i(M))) + (\epsilon')^2 = h_t^{\epsilon,U} \end{aligned} \tag{5.26}$$

Taking the expectation, we get

$$\overline{E}\left(h_t^{\epsilon,U} - h_t^{\epsilon,L}\right) = (\epsilon')^2 + 4\epsilon' + 4\epsilon'\overline{V}(M_i(M)) \tag{5.27}$$

Now

$$\overline{V}(M_i(M)) \leq \sup_{M \in \mathcal{M}} \overline{V}(M_i(M)) \leq (K + \epsilon')^2 \tag{5.28}$$

since it follows directly from Lemma 3 that $(\sup_{M \in \mathcal{M}} h_t(M))$ is exponentially forgetting of order 2, and it therefore exists a constant $K$ such that

$$\sup_{M \in \mathcal{M}} \overline{V}(M) \leq K^2 \tag{5.29}$$

Choosing $\epsilon'$ sufficiently small ensures

$$\overline{E}\left(h_t^{\epsilon,L} - h_t^{\epsilon,U}\right) < \epsilon \tag{5.30}$$

Hence, the conditions of Theorem 2 are fulfilled, and (5.20) follows.

**Proof of (5.21):** The terms in $V_l'$ will make up a sequence $(||\varepsilon_t(S,\hat{\theta}_{S,l})||^2)$ that is exponentially forgetting of order 2, and since the model subset $\{M = (S,\hat{\theta}_{S,l}) \mid S \in \mathcal{S}\} \subset \mathcal{M}$ can be covered by a finite $\epsilon$-net, the rest of the proof is similar to that of (5.20).

□

Eq. (5.20) says that the empirical average $V_l$ converges with probability one to its expectation $\overline{V}$ simultaneously for all possible models in the model set. This is often referred to as uniform convergence, or as a strong uniform law of large numbers. Its main implication is that $V_l$ will be a close approximation to $\overline{V}$ for all models in $\mathcal{M}$, for large $l$. Hence, minimization of $V_l$ and $\overline{V}$ are approximately equivalent. The result of Ljung (1978), which applies to a fixed model structure, is a special case:

**Corollary 2** *Suppose S1 and M2 hold. Consider a fixed model structure $S \in \mathcal{S}$, suppose $\Theta_S$ is compact and $\hat{y}_t(S,\theta)$ is differentiable with respect to $\theta$ for all $t$ and $\theta \in \Theta_S$. Then*

$$\sup_{\theta \in \Theta_S} \left| V_l(S,\theta) - \overline{V}(S,\theta) \right| \rightarrow 0 \quad w.\ p.\ 1 \quad as \quad l \rightarrow \infty$$

□

In addition to the use of a separate validation data sequence, it is for example known that the MDL criterion (Rissanen 1980) and Bayesian criterion (Kashyap 1977) lead to consistent model structure estimates. On the other hand, it is known that AIC and FPE are biased (Shibata 1976), and does therefore not lead to consistent model structure estimates.

Theorem 3 still holds if we introduce several structure levels in a hierarchy, using separate data sequences to formulate a criterion at each level in the hierarchy. We may also add a regularization term (Tikhonov and Arsenin 1977) in the criterion:

$$V_{l,\gamma}(M) = V_l(M) + \gamma(l)\Omega(M) \tag{5.31}$$

where $\gamma(l) > 0$ is a regularization parameter, and $\Omega$ is a stabilizer that has the purpose of making the problem better conditioned. To asymptotically reduce the bias introduced by this term, we let $\gamma(l) \rightarrow 0$ as $l \rightarrow \infty$. Optimization of the regularization parameter $\gamma(l)$ using yet another independent data sequence does not impose any difficulties in the analysis.

The data sequence is split into two sequences of approximately equal length. The analysis holds as long as the length of both sequences go to infinity as $l$ goes to infinity. In order to find a good trade-off, one should consider the size of the model structure set relative to the parameter sets, and take into account the convergence rate, cf. (5.24). Unfortunately, accurate estimates of the convergence rates are not easily attainable. There are some convergence results available in (Vapnik 1982, Pollard 1984), but they suffer from two major drawbacks: They are potentially very conservative, and applicable for independent stochastic processes only. It is not clear how to extend these results to exponentially forgetting processes.

## 5.3 Discussion and Concluding Remarks

A model set is decomposed into a structural and a parametric level. For each model structure, the parameters are estimated using a prediction error method. The model structures are compared by estimating their expected prediction performance on a separate validation data sequence, using the estimated parameters. Here we show that under reasonable assumptions, selecting the model structure with the best performance on the validation data asymptotically leads to a model with the best possible expected prediction performance within the given model set.

We have shown that this procedure essentially has the same asymptotic convergence properties as the parameter identification procedure of Ljung (1978). We have introduced one additional assumption, namely the existence of a finite $\epsilon$-net that covers the model set, but also illustrated that this assumption holds under some reasonable conditions on the model set, cf. Lemmas 1 and 2.

# Chapter 6

# Identification of Operating Regimes

In Chapters 2 and 3 we focused on regime decomposition and choice of local model structure on the basis of system knowledge. The aim of this chapter is to report on an algorithm that automatically identifies a decomposition into operating regimes and local models on the basis of empirical data. Only weak a priori knowledge, such as one or more alternative local model structures is required.

This chapter is organized as follows: First, in section 6.1 the problem is formulated. Then, in section 6.2, a heuristic search algorithm is presented. Statistical properties of the algorithm are discussed in section 6.3, and the algorithm is applied to simulated and experimental data sets in section 6.4. Next, in section 6.5, the knowledge requirements and possibilities for incorporation of a priori knowledge in the model are considered, before it is discussed how qualitative system knowledge can be extracted from the identified model. An overview of related work and a discussion of the limitations and possible improvements of the approach follows. A short version of the chapter can be found in (Johansen and Foss 1994*b*).

## 6.1 Problem Formulation

We address the problem of inducing a model of an unknown non-linear system on the basis of a sequence of $l$ input/output observations

$$\mathcal{D}_l = ((u(1), y(1)), (u(2), y(2)), ..., (u(l), y(l)))$$

where $u(t) \in R^r$ and $y(t) \in R^m$ are the input and output vectors of the system, respectively. We denote by $\mathcal{D}_t$ the subsequence of $\mathcal{D}_l$ containing data up to and including time $t \leq l$.

First, consider static models

$$y(t) = f(u(t)) + e(t) \tag{6.1}$$

where $e(t) \in R^m$ is zero-mean noise, and $f$ is an unknown function to be estimated. An approximation $\hat{f}$ to $f$ suggests the predictor

$$\hat{y}(t|u(t)) = \hat{f}(u(t))$$

which gives the prediction error

$$\varepsilon(t) = y(t) - \hat{y}(t|u(t)) = f(u(t)) - \hat{f}(u(t)) + e(t)$$

Next, consider a stable dynamic system represented by an NARMAX (non-linear ARMAX) model, (Chen and Billings 1989):

$$\begin{aligned} y(t) = f(&y(t-1), ..., y(t-n_y), u(t-1), ..., u(t-n_u), \\ &e(t-1), ..., e(t-n_e)) + e(t) \end{aligned} \tag{6.2}$$

where $e(t) \in R^m$ is zero-mean noise, and $n_y, n_u$, and $n_e$ are non-negative integers. Given an approximation $\hat{f}$ to the function $f$, a one-step-ahead predictor $\hat{y}(t|\mathcal{D}_{t-1})$ can be formulated:

$$\begin{aligned} \hat{y}(t|\mathcal{D}_{t-1}) &= \hat{f}(y(t-1), ..., y(t-n_y), u(t-1), ..., u(t-n_u), \varepsilon(t-1), ..., \varepsilon(t-n_e)) \\ \varepsilon(t) &= y(t) - \hat{y}(t|\mathcal{D}_{t-1}) \end{aligned}$$

The motivation behind this predictor is that while the noise sequence $e$ is unknown, $\varepsilon(t) \rightarrow e(t)$ as $t \rightarrow \infty$, if $\hat{f} = f$, the noise model is invertible, and the initial values are within the region of attraction.

Finally, we consider state-space models

$$x(t+1) = g(x(t), u(t)) + v(t) \tag{6.3}$$
$$y(t) = h(x(t)) + w(t) \tag{6.4}$$

where $x(t)$ is a state-vector, and $v(t)$ and $w(t)$ are zero-mean disturbance and noise vectors of appropriate dimensions. In this case, the model is defined by the functions $g$ and $h$. Again, using approximations $\hat{g}$ and $\hat{h}$, it is possible to construct a one-step-ahead predictor $\hat{y}(t|\mathcal{D}_{t-1})$ using for example the extended Kalman-filter approach, e.g. (Ljung 1987)

$$\begin{aligned} \hat{x}(t|\mathcal{D}_{t-1}) &= \hat{g}(\hat{x}(t-1|\mathcal{D}_{t-2}), u(t-1)) + K(t-1)\varepsilon(t-1) \\ \hat{y}(t|\mathcal{D}_{t-1}) &= \hat{h}(\hat{x}(t|\mathcal{D}_{t-1})) \\ \varepsilon(t) &= y(t) - \hat{y}(t|\mathcal{D}_{t-1}) \end{aligned}$$

where $K(t)$ is the Kalman-filter gain matrix. This matrix will depend explicitly on the time, the functions $\hat{g}$ and $\hat{h}$, and the covariance matrices of the disturbance and noise sequences.

### 6.1.1 A Generalized Framework

In all these cases, we can write the model equations on the form

$$\eta(t) = f(\xi(t)) + e(t) \tag{6.5}$$

where $\eta(t) \in R^{\overline{m}}$ is a generalized output-vector, $\xi(t) \in R^{\overline{r}}$ is a generalized input-vector, and $e(t) \in R^{\overline{m}}$ is zero-mean noise. We denote the space $R^{\overline{r}}$ the input space. In the static model case (6.1), the input and output vectors equals the generalized input and output vectors. In the NARMAX case (6.2), the generalized input vector contains delayed input and output vectors in addition to delayed noise vectors, while the generalized output equals the system output. If noise terms $e(t-1), ..., e(t-n_e)$ are present, the generalized input vector is partially unknown and cannot be found exactly from the data $\mathcal{D}_t$. For state-space models (6.3)-(6.4), neither the generalized input nor the generalized output vectors can be found exactly, because they contain the unmeasured state vector. The purpose of the formulation (6.5) with the generalized input and output vectors is to write the model in a generic form with one unknown function $f$. The problem we address is to estimate this function, and since this function immediately gives the model equations, this also solves the system identification problem. Notice that the fact that the generalized inputs and outputs may not be exactly known is not a too difficult problem, since the model parameters can still be estimated from the input/output data using a prediction error approach with the predictors described above, e.g. (Söderström and Stoica 1988).

### 6.1.2 Model Representation

We define the system's operating point at time $t$ as $z(t) = (z_1(t), ..., z_d(t))^T \in Z = R^d$, where typically $d \leq \overline{r}$ and the operating space $Z$ is a subspace or sub-manifold of the input space. It is assumed that $\xi$ and $z$ are related by a known bounded mapping $H$ such that $z = H(\xi)$. Typically, $Z$ and $H$ are designed such that the operating point $z(t)$ characterizes different modes of behavior of the system under different operating conditions, as we have discussed in Chapters 2 and 3. Suppose $Z$ is decomposed into $N$ disjoint sets $\{Z_i\}_{i \in I_N}$ (regimes) so that

$$Z = \bigcup_{i \in I_N} Z_i$$

for some index set $I_N = \{i_1, ..., i_N\}$ with $N$ elements. Assume that for each regime $Z_i$ we have a local model structure defined by the function $\hat{f}_i(\xi; \theta_i)$ parameterized by the vector $\theta_i$, and a local model validity function $\rho_i(z) \geq 0$ which indicates the relative validity of the local model structure as a function of $z$. In addition to being smooth, $\rho_i$ is designed to have the property that $\rho_i(z)$ is close to zero if $z \notin Z_i$. Furthermore, it is assumed that for all $z \in Z$ there exists an $i \in I_N$ such that $\rho_i(z) > 0$, to ensure completeness of the model. A global model can be formed as

$$\hat{f}(\xi) = \sum_{i \in I_N} \hat{f}_i(\xi; \theta_i) w_i(z) \tag{6.6}$$

$$w_i(z) = \rho_i(z) \Big/ \sum_{j \in I_N} \rho_j(z) \tag{6.7}$$

where the functions in the set $\{w_i\}_{i \in I_N}$ are called interpolation functions.

A model structure based on a decomposition into $N$ regimes is written

$$S_N = \left\{ \left( Z_i, \rho_i, \hat{f}_i \right) \right\}_{i \in I_N} \tag{6.8}$$

This is somewhat redundant, since there is a close (but not necessarily one-to-one) relationship between $Z_i$ and $\rho_i$.

With this representation, the modeling problem consists of the following subproblems:

1. Choose an operating space $Z$ and mapping $H$.
2. Decompose $Z$ into regimes, and choose local model structures.
3. Identify the local model parameters.

In Chapters 2 and 3 it is demonstrated with examples that in some cases, some coarse qualitative process knowledge is sufficient to carry out this procedure. In the following sections we propose an algorithm that requires significantly less system knowledge to decompose $Z$, choose local model structures, and construct interpolation functions.

### 6.1.3 Model Structure Identification Criteria

Let a model structure $S$ of the form (6.8) be given. Notice that in a model structure, the model parameters $\theta^T = (\theta_{i_1}^T, ..., \theta_{i_N}^T)$ are considered unknown. The model structure $S$ together with the admissible parameter set $\Theta_S$ induces a model set

$$\mathcal{M}_S = \{M = (S, \theta);\ \theta \in \Theta_S\}$$

In this section, we will discuss how different model structures can be compared using a sequence of empirical data to estimate its expected prediction performance. We introduce the notation

$$\begin{aligned} y(t) &= y^\star(\mathcal{D}_{t-1}) + e(t) \\ \varepsilon(t|S,\theta) &= y(t) - \hat{y}(t\,|\mathcal{D}_{t-1}, S, \theta\,) \end{aligned}$$

where $y^\star(\mathcal{D}_{t-1})$ is the deterministic (predictable) component of the system output, $e(t)$ is the stochastic (unpredictable) component, and $\varepsilon(t|S,\theta)$ is the residual. Let $\hat{\theta}_S$ be a parameter estimate that minimize the prediction error criterion

$$J_S(\theta) = \frac{1}{l} \sum_{t=1}^{l} \operatorname{trace}\left( \varepsilon(t\,|S,\theta\,)\varepsilon^T(t\,|S,\theta\,) \right) \tag{6.9}$$

Let an unknown *future* data sequence be denoted $\mathcal{D}_t^\star$, and assume $\mathcal{D}_t^\star$ and $\mathcal{D}_l$ are uncorrelated. Moreover, let $E_\mathcal{D}$ and $E_{\mathcal{D}^\star}$ denote expectations with respect to $\mathcal{D}_l$ and $\mathcal{D}_t^\star$, respectively. The prediction error is given by

$$\varepsilon^\star(t|S,\hat{\theta}_S) = y^\star(\mathcal{D}_{t-1}^\star) - \hat{y}(t|\mathcal{D}_{t-1}^\star, S, \hat{\theta}_S(\mathcal{D}_l)) + e(t)$$

where the dependence of $\mathcal{D}_l$ on $\hat{\theta}_S$ has been written explicitly. The expected squared prediction error is defined by

$$\Sigma(S) = E_{\mathcal{D}^\star} E_\mathcal{D} \varepsilon^\star(t|S,\hat{\theta}_S(\mathcal{D}_l))\varepsilon^\star(t|S,\hat{\theta}_S(\mathcal{D}_l))^T$$

Assuming $e(t)$ is white noise that is uncorrelated with $\mathcal{D}_{t-1}^\star$ and $\mathcal{D}_l$, we get the following bias/variance decomposition of this expected squared prediction error

$$\begin{aligned}
\Sigma(S) = {} & E_{\mathcal{D}^\star}\Big(y^\star(\mathcal{D}_{t-1}^\star) - E_\mathcal{D}\hat{y}(t|\mathcal{D}_{t-1}^\star, S, \hat{\theta}_S(\mathcal{D}_l))\Big) \\
& \cdot \Big(y^\star(\mathcal{D}_{t-1}^\star) - E_\mathcal{D}\hat{y}(t|\mathcal{D}_{t-1}^\star, S, \hat{\theta}_S(\mathcal{D}_l))\Big)^T \\
& + E_{\mathcal{D}^\star} E_\mathcal{D}\Big(\hat{y}(t|\mathcal{D}_{t-1}^\star, S, \hat{\theta}_S(\mathcal{D}_l)) - E_\mathcal{D}\hat{y}(t\left|\mathcal{D}_{t-1}^\star, S, \hat{\theta}_S(\mathcal{D}_l))\right.\Big) \\
& \cdot \Big(\hat{y}(t|\mathcal{D}_{t-1}^\star, S, \hat{\theta}_S(\mathcal{D}_l)) - E_\mathcal{D}\hat{y}(t|\mathcal{D}_{t-1}^\star, S, \hat{\theta}_S(\mathcal{D}_l))\Big)^T \\
& + E_{\mathcal{D}^\star} e(t)e^T(t)
\end{aligned} \tag{6.10}$$

The first term is the squared systematic error (squared bias) caused by a too simple model structure. The second term is the random error (variance) that is present because the best model in the model set $\mathcal{M}_S$ cannot in general be identified on the basis of the finite data sequence $\mathcal{D}_l$. Finally, the third term is the unpredictable component of the system output. Notice that the first term does not depend on the data $\mathcal{D}_l$, while the third term does not depend on neither the data $\mathcal{D}_l$, nor the model structure.

We apply a set of approximators that can approximate any smooth function uniformly on a compact subset of the input space, cf. Theorem 2 in Chapter 2. This is obviously a desirable property of the model set, but also a cause for some problems. The richness implies that there will exist models in the model set that will make the bias arbitrarily small. However, the finite amount of data will give large variance for such models. Such a model will be fitted not only to represent the system, but also the particular realization of the noise. In other words, the model may give very good prediction of $\mathcal{D}_l$, but poor prediction capability when applied to $\mathcal{D}_t^\star$. This is known as over-fitting, and is caused by too many degrees of freedom in the model. It is therefore important that a model with the correct number of degrees of freedom is found, in the sense that it balances bias and variance. A general guideline is the parsimony principle, e.g. (Ljung 1987), which states that the best model structure is the one with least degrees of freedom that adequately describes the behavior of the system. We will base the model structure identification algorithm on statistical criteria that reflect this principle.

The mean square error (MSE) criterion is defined by

$$J_{MSE}(S) = \mathrm{trace}(\Sigma(S))$$

Minimizing $J_{MSE}$ will lead to a parsimonious model structure, but with a finite sequence of data $\mathcal{D}_l$, the problem is ill-posed. The reason is simply that $J_{MSE}$ cannot be computed since the probability distribution for the prediction error is unknown. An alternative would be to minimize the average squared prediction error (PE) criterion with respect to the model structure

$$J_{PE}(S) = \text{trace}\left(\frac{1}{l}\sum_{t=1}^{l}\varepsilon(t|S,\hat{\theta}_S)\varepsilon(t|S,\hat{\theta}_S)^T\right)$$

For finite $l$, $J_{PE}$ may be a strongly biased estimate of $J_{MSE}$, since the prediction performance is measured using the same data the parameters are fitted to, and the law of large numbers is not valid because of the strong dependences between the terms. Hence, the use of $J_{PE}$ for structure identification, will not lead to a parsimonious model. We will in the following present several criteria that are far better estimates of $J_{MSE}$ than $J_{PE}$.

If a separate data sequence $\mathcal{D}_l^\star$ (independent of $\mathcal{D}_l$) is known, an unbiased estimate of $J_{MSE}$ can be found by computing the empirical average squared prediction error that results when the model fitted to the data $\mathcal{D}_l$ is used to predict the data $\mathcal{D}_l^\star$. This is the simplest and perhaps most reliable procedure, but suffers from the drawback that a significantly larger amount of data is required. Experiments and collection of data is a major cost for many modeling problems. We therefore proceed with some alternatives that allow the data $\mathcal{D}_l$ to be reused to find good estimates of $J_{MSE}$.

First we consider the final prediction error criterion (FPE) (Akaike 1969), given by

$$J_{FPE}(S) = \frac{1+p(S)/l}{1-p(S)/l}\,J_{PE}(S)$$

where $p(S)$ is the effective number of parameters (degrees of freedom) in the model structure. $J_{FPE}$ is an estimate of $J_{MSE}$, and penalizes model complexity relative to the length of the available data sequence through the term $p(S)/l$. It is assumed that the predictor is linearly parameterized. A non-linear generalization is given in (Larsen 1992). Closely related criteria are Mallows $C_p$ statistic (Mallows 1973), and the Akaike Information Criterion (AIC) (Akaike 1974), which is appropriate in the context of maximum likelihood estimation.

An alternative criterion can be formulated using cross validation (Stone 1974). The basic principle is extended to dynamic systems by Stoica et al. (1986) and Janssen, Stoica, Söderström and Eykhoff (1988). Let $v \leq l$ be a positive integer, and let $k = [l/v]$, i.e. the largest integer not greater than $l/v$. We define the index sets

$$\begin{aligned}\mathcal{I}_j &= \{(j-1)k+1,(j-1)k+2,...,kj\}, \qquad j=1,...,v-1\\ \mathcal{I}_v &= \{(v-1)k+1,(v-1)k+2,...,l\}\end{aligned}$$

which forms a partition of the time interval from 1 to $l$ into $v$ approximately equal-sized intervals. The $v$-fold cross-validation criterion is defined as

$$J_{CV}^v(S) = \text{trace}\left(\frac{1}{l}\sum_{j=1}^{v}\sum_{t\in\mathcal{I}_j}\varepsilon\left(t\,\middle|\,S,\hat{\theta}_S^{-\mathcal{I}_j}\right)\varepsilon^T\left(t\,\middle|\,S,\hat{\theta}_S^{-\mathcal{I}_j}\right)\right)$$

where the parameter estimate $\hat{\theta}_S^{-\mathcal{I}_k}$ minimizes the following criterion

$$J_{PE}^{-\mathcal{I}_j}(\theta) = \frac{1}{l - \#\mathcal{I}_j} \sum_{t \in \mathcal{I}_k^\star} \varepsilon(t\,|\mathcal{D}_{t-1}, S, \theta)\varepsilon^T(t\,|\mathcal{D}_{t-1}, S, \theta)$$

where $\mathcal{I}_k^\star = \{1, 2, ..., l\} \backslash \mathcal{I}_k$, and $\#\mathcal{I}_j$ is the number of elements in the set $\mathcal{I}_j$. Cross-validation may give a reasonable approximation to the use of independent data for selecting the model structure, at the cost of extra computations. The computational complexity can be considerably reduced when the prediction error is a linear function of the parameters, or using the approximate cross-validation criteria of Stoica et al. (1986) and Janssen et al. (1988). It is shown that the approximate criteria are asymptotically equivalent to FPE, as $l \to \infty$. Yet another approximation to cross-validation is the Generalized Cross Validation (GCV) criterion (Craven and Wahba 1979)

$$J_{GCV}(S) = \frac{1}{(1 - p(S)/l)^2} J_{PE}(S)$$

which is easily seen to be asymptotically equivalent to FPE, and assumes linear parameterization of the predictor.

Bootstrapping (Efron and Tibshirani 1986), is a general procedure for the estimation of statistical properties. In Chapter 4, we introduced a bootstrap criterion that estimates the expected squared prediction error, and two modified bootstrap criteria that gives improved estimates of the expected squared prediction error when the system operates under a wider range of operating conditions than reflected in the empirical data $\mathcal{D}_l$.

Any one of these criteria can be applied with the structure identification algorithm presented in the next section.

## 6.2 System Identification

Let a set of candidate local model structures $\mathcal{L} = \{f_1, f_2, ..., f_{N_L}\}$ be given, where $f_i$ is a parameterized function that determines a local model structure, cf. (6.8).

### 6.2.1 The Set of Model Structure Candidates

Assume the input- and output observations in $\mathcal{D}_l$ are bounded. Then the system's operating range $Z$ can be approximated by the $d$-dimensional box

$$Z_1 = \left[z_{1,1}^{\min}, z_{1,1}^{\max}\right] \times \cdots \times \left[z_{1,d}^{\min}, z_{1,d}^{\max}\right]$$

where $z(t) \in Z_1$ for all $t \in \{1, ..., l\}$, since $H$ is a bounded mapping. Notice that the resulting model will extrapolate and can be applied for operating points outside $Z_1$. Next, we consider the problem of decomposing $Z_1$ into regimes.

Consider the possible decompositions of the set $Z_1$ into two disjoint subsets $Z_{11}$ and $Z_{12}$ with the property $Z_1 = Z_{11} \cup Z_{12}$. We restrict the possibilities by the constraint that the splitting boundary is a hyper-plane orthogonal to one of the natural basis-vectors of $R^d$, i.e.

$$\begin{aligned} Z_{11} &= \{z \in Z_1 \mid z_{d_1} < \zeta_1\} \\ Z_{12} &= \{z \in Z_1 \mid z_{d_1} \geq \zeta_1\} \end{aligned}$$

for some dimension index $d_1 \in \{1, ..., d\}$ of $R^d$ and splitting point $\zeta_1 \in \left[z_{1,d_1}^{\min}, z_{1,d_1}^{\max}\right]$. Local model validity functions for the two regimes are defined by the recursion

$$\begin{aligned} \rho_{11}(z) &= \rho_1(z) b(z_{d_1} - \overline{z}_{11,d_1}; \lambda_{11}) \\ \rho_{12}(z) &= \rho_1(z) b(z_{d_1} - \overline{z}_{12,d_1}; \lambda_{12}) \end{aligned}$$

where $\overline{z}_{i,d_1} = 0.5\left(z_{i,d_1}^{\min} + z_{i,d_1}^{\max}\right)$ for $i \in \{11, 12\}$ is the center point of $Z_i$ in the $d_1$-direction. The function $b(r; \lambda)$ is a scalar basis-function with scaling parameter $\lambda$, and the local model validity function associated with the regime $Z_1$ is $\rho_1(z) = 1$. The scaling parameters are chosen by considering the overlap between the local model validity functions. For $i \in \{11, 12\}$, we choose $\lambda_i = 0.5\gamma\left(z_{i,d_1}^{\max} - z_{i,d_1}^{\min}\right)$ where $\gamma$ is a design parameter that typically takes a value between 0.25 and 2.0. There will be almost no overlap when $\gamma = 0.25$, and large overlap when $\gamma = 2.0$.

For each dimension index $d_1 \in \{1, ..., d\}$ we represent the interval $\left[z_{1,d_1}^{\min},\ z_{1,d_1}^{\max}\right]$ by a finite number of $N_1$ points uniformly covering the interval. Now $d_1, \zeta_1, f_{1,v}$, and $f_{2,w}$ defines a new model structure, where the regime $Z_1$ is decomposed according to the dimension index $d_1$ at the point $\zeta_1$, and the two local model structures are $f_{1,v}$ and $f_{2,w}$. Formally, the set of candidate model structures $\mathcal{S}_n$ with $n$ regimes is given by

$$\begin{aligned} \mathcal{S}_1 &= \{\{(Z_1, \rho_1, f_j)\};\ j \in \{1, 2, ..., N_L\}\} \\ \mathcal{S}_2 &= \left\{\left\{\left(Z_{11}^i, \rho_{11}^i, f_j\right), \left(Z_{12}^i, \rho_{12}^i, f_k\right)\right\};\ i \in \{1, 2, ..., dN_1\},\ j, k \in \{1, 2, ..., N_L\}\right\} \\ \mathcal{S}_3 &= \left\{\left\{\left(Z_{11}^i, \rho_{11}^i, f_j\right), \left(Z_{121}^m, \rho_{121}^m, f_k\right), \left(Z_{122}^m, \rho_{122}^m, f_n\right)\right\};\right. \\ &\qquad\qquad \left. i, m \in \{1, 2, ..., dN_1\},\ j, k, n \in \{1, 2, ..., N_L\}\right\} \\ &\quad \cup \left\{\left\{\left(Z_{111}^m, \rho_{111}^m, f_k\right), \left(Z_{112}^m, \rho_{112}^m, f_n\right), \left(Z_{12}^i, \rho_{12}^i, f_j\right)\right\};\right. \\ &\qquad\qquad \left. i, m \in \{1, 2, ..., dN_1\},\ j, k, n \in \{1, 2, ..., N_L\}\right\} \\ \mathcal{S}_4 &= \cdots \end{aligned}$$

The model structure set is now

$$\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3 \cup \cdots$$

which is illustrated as a search tree in Fig. 6.1. Strictly speaking, the model structure set is not a tree, since sometimes different sequences of decompositions lead to the same model structure. However, we choose to represent this as a tree, for simplicity. Now the structure identification problem can be looked upon as a multi-step decomposition process, where at each step one regime from the previous step is decomposed into two sub-regimes. Such an approach will lead to a sequence of model structures $S_1$, $S_2$, ..., $S_n$ where the model structure $S_{i+1}$ has more degrees of freedom than $S_i$. Due to the normalization of the local model validity functions, the model set is usually not strictly hierarchical, in the sense that $S_i$ cannot be exactly represented using $S_{i+1}$. However, the increasing degrees of freedom defines a hierarchical structure.

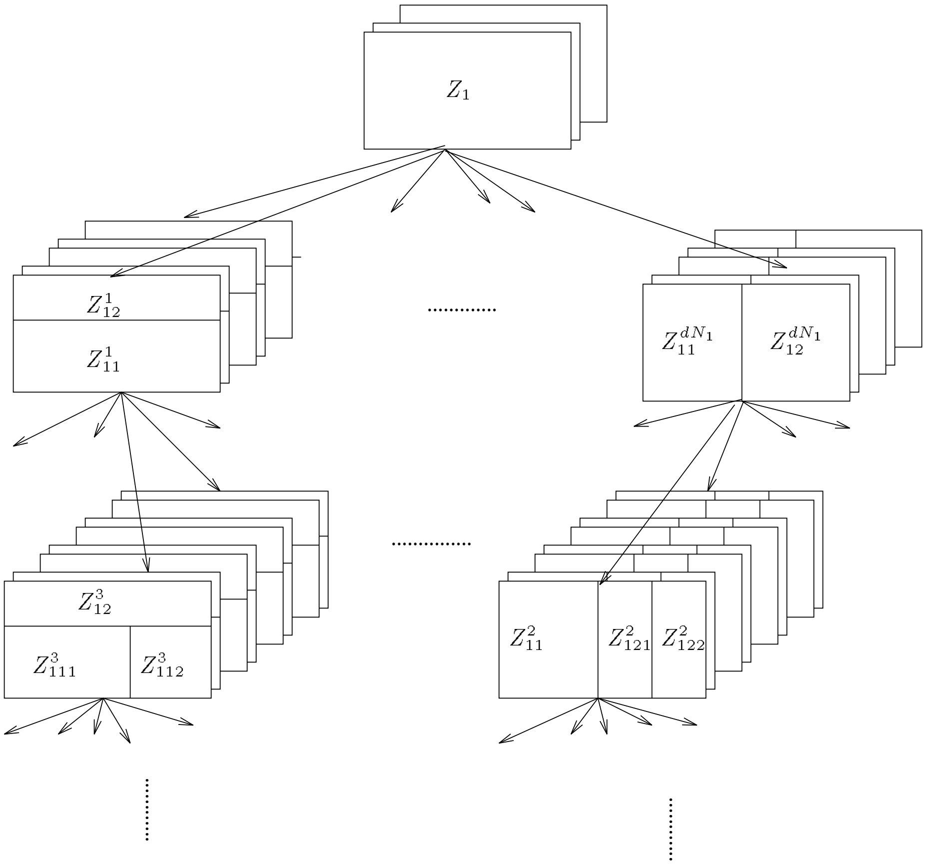


Figure 6.1: Model structure search tree illustrating possible decompositions into regimes and choice of local model structures. Each level in the tree correspond to the possible decompositions into one more regime than at the previous level, i.e. the model structure sets $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$ etc. The subset of model structures at each "super-node" in the tree corresponds to a fixed decomposition into regimes, but different combinations of local model structures.

### 6.2.2 Basic Search Algorithm

The problem is now to search the set $\mathcal{S}$ for the best possible model structure. The estimate of the parameters in model structure $S$ is defined by

$$\hat{\theta} = \arg\min_{\theta} J_S(\theta) \tag{6.11}$$

where it has been assumed that a unique minimum exists. This can be ensured by restricting the parameters to a compact set. Now, the chosen structure identification criterion is written $J'(S)$. We define for a given $n$

$$S_n = \arg\min_{S \in \mathcal{S}_n} J'(S) \tag{6.12}$$

when it again has been assumed that a unique minimum exists. Consider the following extended horizon search algorithm, where the integer $n^\star \geq 1$ is called the search horizon.

**Search Algorithm.**

1. Start with the regime $Z_1$. Let $n = 1$.
2. At each step $n \geq 1$, find a sequence of decompositions and local model structures $S_n, S_{n+1}, ..., S_{n+n^\star}$ that solves the optimization problem
$$\min_{S \in \mathcal{S}_{n+n^\star}} J'(S)$$
3. Restrict the search tree by keeping the decomposition that leads to $S_{n+1}$ fixed for the future.
4. If
$$J'(S_n) > \min_{k \in \{1,2,\ldots,n^\star\}} J'(S_{n+k})$$
then increment $n$ and go to 2. Otherwise, the model structure $S_n$ is chosen.

Referring to Fig. 6.1, this algorithm will search the tree starting at the top (corresponding to one local model covering the whole operating space), and selecting a decomposition at each level through a sequence of "locally exhaustive" searches of depth $n^\star$. In other words, this algorithm will make an $n^\star$-step-ahead optimal decomposition at each step, in the sense that the decomposition is optimal if there is going to be exactly $n^\star$ more decompositions. If $n^\star = 1$, this is a local search algorithm.

### 6.2.3 Heuristic Search Algorithm

Clearly, the performance of the algorithm is expected to improve as $n^\star$ increases, but the computational complexity makes $n^\star > 3$ not feasible for any practical problem with 1994 desktop computer technology, even if the local model structure set $\mathcal{L}$ contains as few as 1 or 2 possibilities.

**Example.** Consider the problem of identifying a state-space model of the form $x(t+1) = f(x(t))$, where $\dim(x) = 5$, and we apply local models of the form

$$\begin{pmatrix} x_1(t+1) \\ x_2(t+1) \\ \vdots \\ x_5(t+1) \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_5 \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{15} \\ a_{21} & a_{22} & & a_{25} \\ \vdots & & \ddots & \vdots \\ a_{51} & a_{52} & \cdots & a_{55} \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_5(t) \end{pmatrix}$$

Each of the 30 parameters can be replaced by a structural zero, which gives a set of local linear model structure with $2^{30} \approx 10^9$ elements. On the other hand, even if there is only one possible local model structure to choose among, the number of possible decompositions into no more than five regimes is

$$\#\mathcal{S}_1 + ... + \#\mathcal{S}_5 = 1 + dN_1 + 2(dN_1)^2 + 3!(dN_1)^3 + 4!(dN_1)^4$$

For $d = 2$ and $N_1 = 10$, this is approximately $4 \cdot 10^6$ candidate decompositions. With $n^\star < 4$, the model structure set is considerably reduced. In particular, $n^\star = 3$ gives about $10^5$ decompositions to search among, $n^\star = 2$ gives about 2500, while $n^\star = 1$ gives 80 candidate decompositions.

□

Because of the combinatorial nature of the model structure set, it is clearly of interest to implement some heuristics that cut down on the computational complexity, without sacrificing too much of the optimality of the algorithm. As we have seen, the number of candidate decompositions at each step in the search may be large. To reduce the number of candidates, we suggest to apply the following heuristics in the "locally exhaustive" search at the second step in the search algorithm:

**Heuristic 1**. *At each level in the search tree, proceed with only the most promising candidates.*

The best candidates are of course not known a priori, so there is always a possibility that this may lead to a sub-optimal model. We suggest to proceed with the best decomposition for each of the possible splitting dimensions. Instead of trying to find the best candidates, one can often more easily single out the "least promising candidate decompositions":

**Heuristic 2**. *Discard the candidate decompositions that give an increase in the criterion from one level in the search tree to the next.*

The number of remaining candidates will typically be larger than when using Heuristic 1, but the chance of discarding the optimal decomposition may be smaller. Some candidate decompositions may give rise to regimes where no substantial amount of data is available, and may therefore be classified a priori as not feasible:

**Heuristic 3**. *Discard candidate decompositions that lead to regimes with less data points relevant to the regime than the number of degrees of freedom in the corresponding local model structure and local model validity function.*

Counting the number of relevant data-points associated with each regime is controversial, since the interpolation functions overlap. We use the heuristic count $l_i = \sum_{t=1}^{l} w_i(z(t))$, which has the attractive property $\sum_{i \in I_N} l_i = l$.

**Heuristic 4**. *Use a (backward or forward) stepwise regression procedure to handle local model structure sets $\mathcal{L}$ of combinatorial nature, as implemented by (Sugeno and Kang 1988).*

Related to the example above, one should start with no structural zeroes, and then add one structural zero at a time, choosing the one that gives most significant improvement in the prediction performance. This should give less than $30 + 29 + 28 + ... + 1 = 465$ candidate model structures, which is quite different from $2^{30}$.

### 6.2.4 Parameter Identification

For each model structure $S$, the prediction error criterion $J_S(\theta)$ is used for parameter identification. Let us briefly concentrate on parameter identification algorithms for models with a fixed structure, i.e. solving the problem (6.11).

First, consider the simplest possible case, when the generalized output vector $\eta(t)$ and generalized input vector $\xi(t)$ are explicitly given by the data. In addition, we assume that all the local model structures are linearly parameterized, i.e. can be written in the linear regression form

$$\hat{f}_i(\xi(t);\theta_i) = \varphi_i^T(\xi(t))\theta_i \tag{6.13}$$

for some matrix $\varphi_i(\xi(t))$. Using (6.13), the global model is

$$\eta(t) = \sum_{i \in I_N} \varphi_i^T(\xi(t)) w_i(z(t))\theta_i + e(t)$$

which is also in the linear regression form. A standard off-line least squares algorithm, e.g. (Ljung 1987), is applied to estimate the parameters.

If any of the assumptions above is violated, a numerical search algorithm is applied to find the parameter estimate. This is simply because the predictor will be non-linearly parameterized. We use the Fletcher-Reeves-Powell conjugate gradient algorithm with a line search (Press, Flannery, Teukolsky and Vetterling 1988), which requires computation of the gradient

$$\frac{\partial}{\partial \theta} J_S(\theta) \tag{6.14}$$

In the case of an NARMAX model based on local linear ARMAX models, this gradient can be effectively computed using a filter similar to the ARMAX case, e.g. Ljung (1987). An explicit expression for the gradient can also be found for some other special cases. Otherwise, an approximation to (6.14) is found by numerical differentiation.

### 6.2.5 User Choices

The basis-function $b(r;\lambda)$ with scaling parameter $\lambda$ has the purpose of providing a smooth interpolation between the local models. The basis-function is assumed to have the property $b(r;\lambda) \geq 0$ for all $r \in R$ and $b(r,\lambda) \rightarrow 0$ as $|r| \rightarrow \infty$. Typical choices are bell-shaped kernel-functions like the unnormalized Gaussian $\exp(-r^2/2\lambda^2)$. It may appear that the choice of this function has significant impact on the model. However, it is our experience that both the algorithm and the model's prediction performance are quite insensitive with respect to this choice, and the specification of this function does not require any prior knowledge about the system. What is more important is the choice of $\lambda$, which is controlled by the user-specified parameter $\gamma$.

To compute the criterion $J_{GCV}$ or $J_{FPE}$, the effective number of parameters $p$ in the model structure must be known. If the choice of model structure is not based on the data $\mathcal{D}_l$, then the effective number of parameters is

$$p = \sum_{i \in I_N} \dim(\theta_i) \tag{6.15}$$

in the case of linear regression. However, the proposed algorithm for model structure identification makes some use of the data $\mathcal{D}_l$ for restricting the model structure set during the search. Hence, the $p$-value given by (6.15) will be too small. Counting the effective number of parameters in this case is controversial. We apply the heuristic

$$p = \kappa(N-1) + \sum_{i \in I_N} \dim(\theta_i)$$

where $\kappa \geq 0$ is a heuristic constant, which also can be interpreted as a smoothing parameter, since a large $\kappa$ will put a large penalty on model complexity, and will therefore give a smooth model. A typical choice of $\kappa$ is between 0 and 4 (Friedman 1991).

## 6.3 Statistical Properties

Consider the bias/variance decomposition (6.10). It is evident that a small bias requires in general a complex model structure with a large number of operating regimes, cf. Theorem 2 in Chapter 2. On the other hand, a small variance requires a model structure that is simple, with few parameters compared to the number of observations $l$. "The perfect model" is characterized by both a small bias and variance, which appears to be impossible to achieve simultaneously for a small $l$. This is known as the bias/variance dilemma, cf. Fig 4.1. However, as $l \to \infty$, it is possible to make the bias and variance vanish simultaneously. This will be discussed in section 6.3.1. For finite $l$, it is important to make the correct trade-off between bias and variance to find the best model structure, as we will discuss in section 6.3.2.

### 6.3.1 Asymptotic Properties

It is shown in Theorem 3 in Chapter 5 that both the bias and the variance will tend towards their smallest possible values, with probability one, provided

1. The prediction error criterion converges to its expectation, see (Ljung 1978) for conditions under which this holds.
2. The estimate $J'$ of the expected squared prediction error used for model structure identification converges to its expectation.
3. Global minima of the parameter and structure optimization problems are found with probability one.

4. The model set

$$\mathcal{M} = \{M = (S, \theta); S \in \mathcal{S}, \theta \in \Theta_S\}$$

can be covered by a finite $\epsilon$-net.

It is shown in Chapter 5 that the use of a separate validation data sequence for model structure identification gives a $J'$ that satisfies the second requirement. It is also known that FPE and AIC are slightly biased.

However, the main problem may be the third requirement, since neither the parameter optimization nor the structure optimization algorithms need result in global minima. An attractive feature of the model structure tree is that it appears not only to have a large amount of multiple global minima, but also many close-to-optimal local minima. It is easy to see that the restriction of the search to any sub-tree of the model structure tree does not exclude any possible decomposition. The worst thing that can happen is that the number of decompositions may be somewhat larger the necessary, or alternatively that the partition may not be as fine as desired. This leads to sub-optimality for finite amount of data, but not necessarily so asymptotically.

The fourth condition is somewhat technical, but does in general impose a restriction on the complexity of the model set. In practise, this is not a serious restriction, as discussed in Chapter 5.

### 6.3.2 Finite Sample Accuracy

For practical problems, the finite sample accuracy is obviously more interesting than the asymptotic properties. Consider for example the problem of modeling the unknown static system

$$y = 2e^{-u^2}\sin(u) + e \tag{6.16}$$

where $e \sim N(0, 0.15^2)$. Using some unknown probability density for $u$, 20 data sequences $\mathcal{D}^1_{100}, ..., \mathcal{D}^{20}_{100}$ are generated, each containing $l = 100$ realizations of the mapping (6.16). Running the identification algorithm on each of these 20 data sequences gives 20 different models. Some have the same structure, but have different parameters, while others have both different structure and parameters. For example, using local linear models, the search algorithm with $n^\star = 1$, the GCV criterion with $\kappa = 1$, the least squares algorithm, and Gaussian basis-function with $\gamma = 1$ gives 20 models as shown in Fig 6.2. The number of regimes is typically three to five, but we observe from Fig. 6.2 that the variability is fairly small, except in the regimes where data is sparse. Tests indicate that the variability may decrease slightly as $n^\star$ increases.

Usually, one has only one data sequence available. In one particular case, suppose only $\mathcal{D}^1_{100}$ is available. It is clearly of interest to compute some measure of the uncertainty of the model identified on the basis of this data set. Since no more independent data is available, the best we can do is to reuse the data $\mathcal{D}^1_{100}$. One effective data resampling technique is the bootstrap, e.g. (Efron and Tibshirani

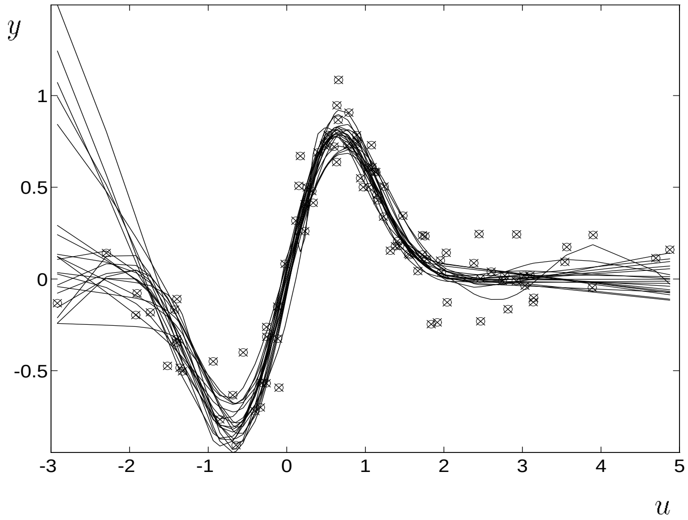


Figure 6.2: The curves are 20 models fitted using 20 different data sets generated by the same system, each containing 100 samples. The dots are some typical observations from the unknown static system $y = f(u) + e$.

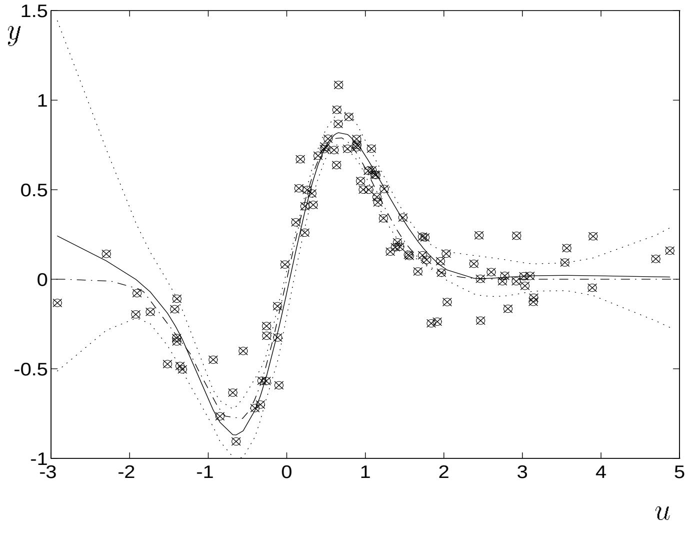


Figure 6.3: The 100 dots are observations from the unknown static system. The solid curve ( — ) is a model $\hat{f}(u)$ fitted using the 100 observations, while the dashed curve $(\cdot - \cdot -)$ is the "true system" $f(u)$. Approximate confidence sets estimated using bootstrapping with $B = 100$ are dotted $(\cdots)$.

1986). The idea is to generate a large number $B$ of new data sequences of length $l$ by randomly choosing data from $\mathcal{D}^1_{100}$ with replacement. This gives a set of $B$ models, and using this set of models, one may for example estimate the bias, variance, and confidence sets using Monte Carlo simulation. Fig. 6.3 shows $f(u)$, $\hat{f}(u)$ and approximate confidence sets corresponding to

$$\hat{f}(u) + \widehat{\text{bias}}\left(\hat{f}(u)\right) \pm 2\sqrt{\widehat{\text{var}}\left(\hat{f}(u)\right)}$$

The use of bootstrapping for the estimation of statistics is somewhat more complicated for dynamic systems because one needs to preserve the time-structure in the resamples, see the survey by Carlstein (1992), and also Chapter 4.

A theoretical analysis of the finite sample properties is outside the scope of this work, although a modification of the results of Kavli and Weyer (1995) would be feasible. These results makes use of the theory of Vapnik (1982), which makes weak assumptions on the model set and underlying probability measure, but is based on the strong assumption that the observations are independent. Essentially, this theory gives potentially very conservative confidence sets for the prediction error, parameter estimates, and model structure.

## 6.4 Examples

### 6.4.1 Simulation Example: A Batch Fermentation Reactor (Revisited)

Consider the fermentation of glucose to gluconic acid by the micro-organism *Pseudomonas ovalis* in a well stirred batch reactor. The main overall reaction mechanism is described by

$$\begin{aligned} \text{Cells} + \text{Glucose} + O_2 &\rightarrow \text{More Cells} \\ \text{Glucose} + O_2 &\xrightarrow{\text{Cells}} \text{Gluconolactone} \\ \text{Gluconolactone} + H_2O &\rightarrow \text{Gluconic Acid} \end{aligned}$$

The first reaction is the reproduction of cells, consuming the substrate glucose, and oxygen. The second reaction is the production of gluconolactone, again consuming glucose and oxygen. This reaction is enzyme-catalysed by the cells, while the last reaction forms the final product, gluconic acid. The model used to simulate the "true system" is the same as in section 3.3.2. The data sets used for identification and validation are also the same as in section 3.3.2.

#### Alternative 1. Semi-empirical Local Model Structure

We have specified only one possible local linear discrete-time state-space model structure

$$x(t+1) = a_i + A_i x(t)$$

with 12 structural zeros in the $A_i$ matrices. These zeros follow directly from the reaction mechanism, using the assumption that glucose and oxygen are rate-limiting. We choose the operating point $z = (s_n,\ c_n)^T$, which captures these non-linearities and characterizes the operating conditions of the process with respect to local linear models, see also section 3.3.2.

Running the search algorithm with $n^\star = 1$, the FPE criterion with $\kappa = 1$, the least squares estimator, and Gaussian basis-function with $\gamma = 1$, results in a model with five operating regimes, and root average squared one-step-ahead prediction error (PE) on the validation data PE=0.0135. Restricting the number of operating regimes to three, gives PE=0.0147, while an identified global linear model gives PE=0.0303. This clearly indicates that there exists significant non-linearities that have been captured by the two more complex models and not by the linear model. The five regimes are illustrated in Fig. 6.4. Perhaps the most interesting and

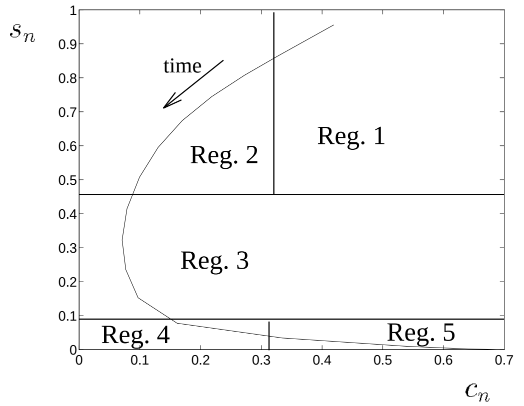


Figure 6.4: The decomposition into five regimes using the simulated fermenter data, with a typical simulated system trajectory projected onto the $(c_n, s_n)$-plane.

attractive feature of the method is that the identified model can be interpreted in a natural way. The five regimes correspond to the following phases in the batch

1. Initial phase.
2. Growth phase, where only the amount of micro-organisms is limiting the rate of the reactions.
3. Oxygen supply is rate-limiting.
4. Glucose is rate-limiting.
5. No glucose left, termination.

This gives a high-level qualitative description of the reactor. More low-level quantitative details on e.g. reaction kinetics can be added by examining the parameters of the local models. The model can be represented as the model tree shown in Fig. 6.5, see also (Strömberg et al. 1991, Omohundro 1987, Breiman, Friedman, Olshen and Stone 1984). A simulation of a typical batch from the test set is shown in Fig.

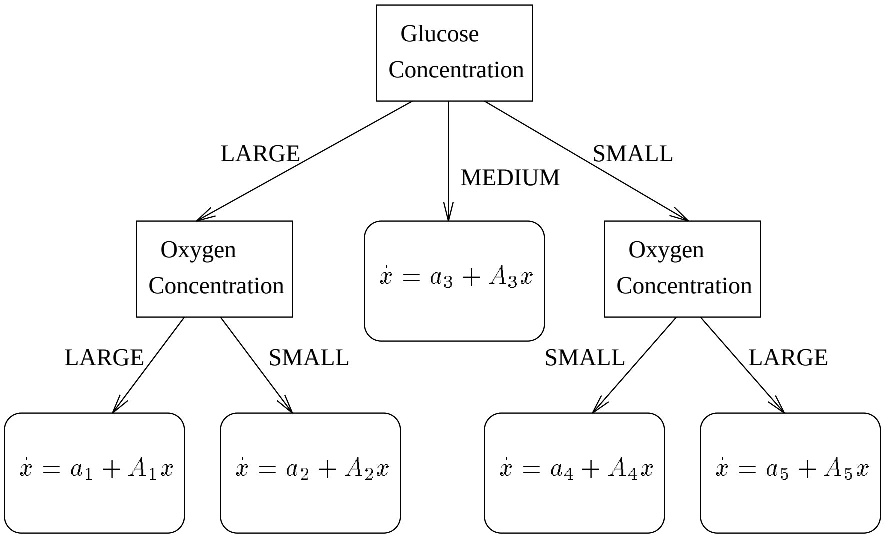


Figure 6.5: Model tree representation of fermenter model.

6.6, using the model with five local models, and the identified global linear model for comparison. Clearly, the results favor the non-linear model. Comparison with similar models with three or four regimes, designed by hand on the basis of system knowledge (Johansen and Foss 1993*b*), see also section 3.3.2 indicates that their performance are comparable. Results of simulations using various heuristics on the problem, when $N = 3$ is fixed a priori, are shown in Fig. 6.7. The heuristics give models that are only marginally suboptimal. Moreover, $n^\star = 3$ gives about 5% improvement over $n^\star = 1$. Even if the variability of the models is small measured by the chosen criterion, the structural difference appears to be large from the qualitatively different regimes in Fig. 6.7. However, from the typical trajectory projected onto the operating space, it is evident that due to the evolution of the batch, all three models are similar, since the three regimes correspond to the initial, intermediate, and final phases of the batch in all three cases.

Basically, the applied a priori knowledge is the overall reaction mechanism, used for structuring the $A_i$ matrices and for selection of the two variables used to characterize the operating regimes. It must be noted that the algorithm has also been applied without this knowledge, i.e. with full $A_i$ matrices and operating point $z = x$, resulting in only a slight decrease in the prediction accuracy of the identified model. In this case, it is interesting to observe that among the five components of the operating point, the algorithm chooses only $\chi_n$ and $c_n$ for decomposition into regimes. As noted in section 3.3.2, due to the batch nature of the process operation, the information in the state $\chi_n$ is highly redundant with the information in $s_n$. This is also evident from Fig. 6.6, which clearly shows the collinearity between these variables. Hence, the algorithm is forced to make a somewhat arbitrary choice about which variable to use for decomposition, a fact that makes the interpretation of the model more difficult. This problem is also commented upon in view of the MARS algorithm by De Veaux et al. (1993).

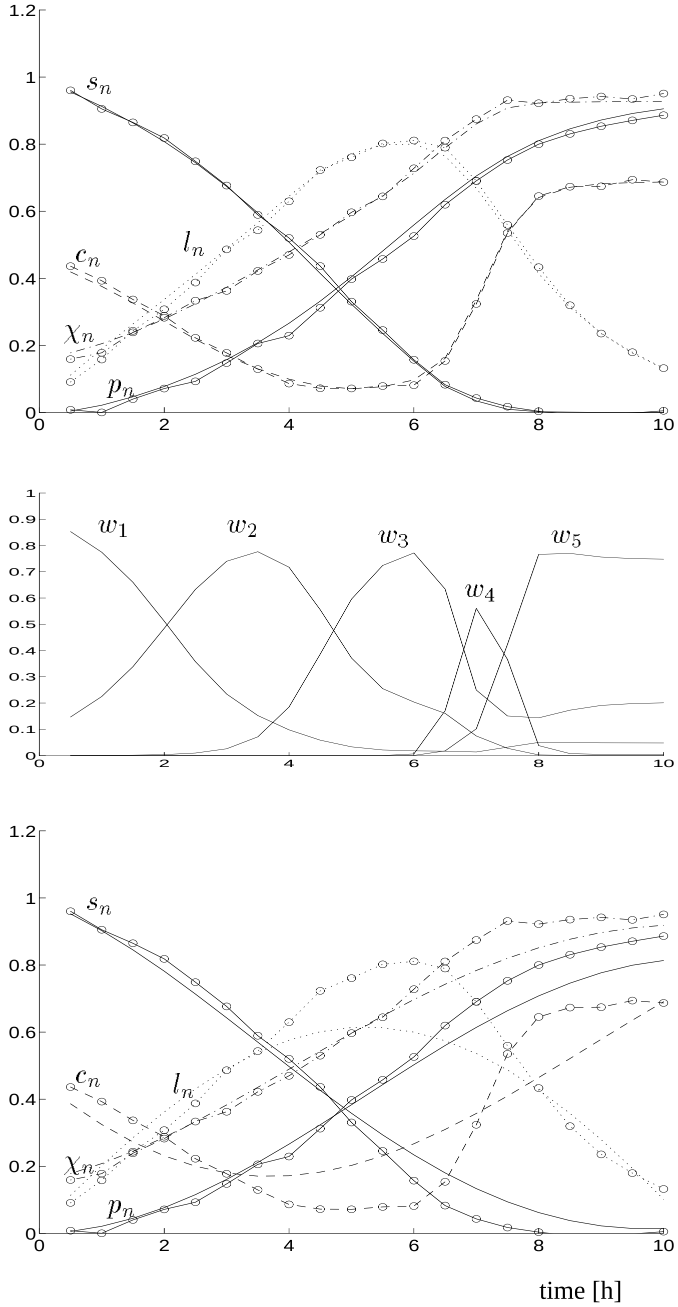


Figure 6.6: Top: Trajectories with circles are generated by the "true system", while the other trajectories are simulations of the model based on five local linear models. Middle: The relative weight of the five local linear models in the interpolation. Bottom: Simulation of an identified global linear model.

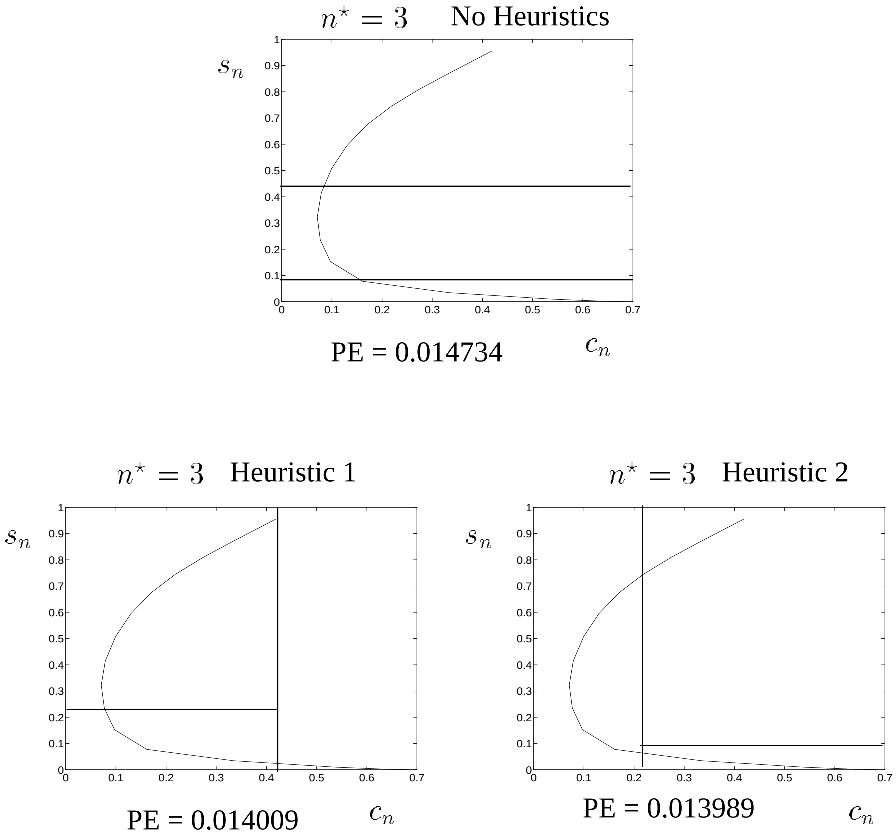


Figure 6.7: The figure shows the decompositions into three regimes found using different heuristics on the simulated fermenter data.

This only emphasizes the important fact that the success of empirical modeling is heavily dependent on the information in the empirical data, and that data deficiencies should to the highest possible extent be compensated for using prior knowledge.

**Alternative 2. Mechanistic local model structure**

A set of mass balances for this reactor can be found using knowledge of the reaction mechanism, stoichiometry, and oxygen uptake mechanism (Ghose and Ghosh 1976)

$$\left.\begin{aligned} \dot{\chi} &= \mu\chi \\ \dot{l} &= \sigma\chi - \frac{1}{Y_{l|p}} k_p l \\ \dot{p} &= k_p l \\ \dot{s} &= -\frac{1}{Y_{s|\chi}} \mu\chi - \frac{1}{Y_{l|\chi}} \sigma\chi \\ \dot{c} &= k_l a (c^\star - c) - \frac{1}{Y_1} \mu\chi - \frac{1}{Y_2} \sigma\chi \end{aligned}\right\} \tag{6.17}$$

The internal variables $\mu, \sigma$ and $k_p$ are reaction rates for the three reactions, $c^\star$ is the maximum dissolved oxygen concentration, and $k_l a$ is a mass transfer coefficient that describes the uptake of oxygen. The yield coefficients $Y_1, Y_2, Y_{l|p}, Y_{l|\chi}$, and $Y_{s|\chi}$ are determined by the stoichiometry, and assumed to be known. The model structure is not complete unless expressions for the abovementioned internal variables and parameters are known. Without any more prior knowledge, one may use empirical relations of the form

$$\left.\begin{aligned} \mu &= \mu(c,s) \\ k_p &= k_p(c,s) \\ \sigma &= \sigma(c,s) \\ c^\star &= \text{constant} \\ k_l a &= \text{constant} \end{aligned}\right\} \tag{6.18}$$

The dependence on variables like temperature, pH, and agitation speed is ignored in this example, but can be included in the same way. In summary, the model structure consists of the incomplete mechanistic model structure (6.17), and some unknown empirical functions (6.18). Let us now consider how this modeling problem can be solved within the operating regime based modeling framework. On the basis of (6.17) and (6.18), we construct a local semi-mechanistic model structures by substituting local empirical approximations

$$\begin{aligned} \mu_i(c,s) &= \theta_{i,1} + \theta_{i,2}c + \theta_{i,3}s \\ k_{p_i}(c,s) &= \theta_{i,4} \\ \sigma_i(c,s) &= \theta_{i,5} + \theta_{i,6}c + \theta_{i,7}s \end{aligned}$$

where the unknown local parameters $\theta_{i,1}, ..., \theta_{i,7}$ correspond to local model $i$. The variable $k_p$ is approximated locally by a constant, while $\sigma$ and $\mu$ are approximated locally by linear functions. The reason for this difference is that it is expected that $k_p$ will show considerably less variations than the other two variables. In addition, the model consists of two unknown global parameters

$$\begin{aligned} c^\star &= \theta_{0,1} \\ k_l a &= \theta_{0,2} \end{aligned}$$

The model is found with the aid of the structure identification algorithm based on the FPE criterion with $n^\star = 1, \gamma = 1.0$, $\kappa = 1.0$, and the parameters are estimated using the Fletcher-Reeves-Powell conjugate gradient algorithm with a line search (Press et al. 1988). The model equations are integrated numerically using an adaptive Runge-Kutta method (Press et al. 1988). We stop the algorithm when four regimes are identified, cf. Fig. 6.8. The prediction performance of the model on an independent, but typical, batch is illustrated in Fig. 6.9, from which we see that the model has captured the major non-linear effects. Since the reaction rates appear linearly in the model equations, we are able to study the identified kinetic model directly, because

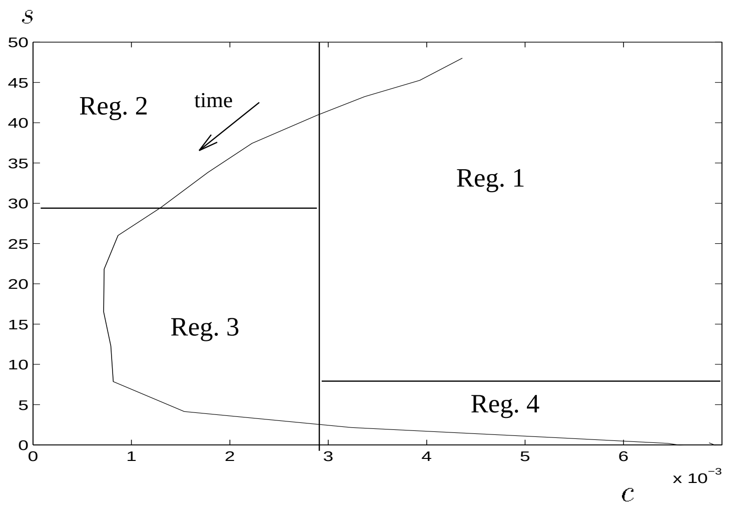


Figure 6.8: The identified decomposition into four regimes in the fermenter simulation example. The trajectory corresponds to a typical batch, projected on the operating space.

$$\hat{\mu}(c,s) = \sum_{i=1}^{4} \hat{\mu}_i(c,s) w_i(c,s)$$
$$\hat{k}_p(c,s) = \sum_{i=1}^{4} \hat{k}_{p_i}(c,s) w_i(c,s)$$
$$\hat{\sigma}(c,s) = \sum_{i=1}^{4} \hat{\sigma}_i(c,s) w_i(c,s)$$

These functions are illustrated in Fig. 6.10, which also contains the "true kinetic model" used in the simulator. Comparing the identified functions with the "true ones" clearly shows that the identified model is a close approximation to the "true model" in the parts of the operating space that are densely populated with empirical data, while there may be some mismatch elsewhere, cf. Fig. 6.8. Moreover, the identified kinetic model suggests a possible interpretation of the 4 regimes. It is evident that neither oxygen nor glucose is rate-limiting in the initial phase of the batch corresponding to Regimes 1 and 2, cf. Fig. 6.8. In these regimes it is only the biomass concentration that limits the reaction rates. On the other hand, oxygen is rate-limiting through Regime 3, and at the end of this regime and through Regime 4, glucose is rate-limiting. It should be mentioned that a similar approach was taken in (Psichogios and Ungar 1992, Aoyama and Venkatasubramanian 1993), where the kinetic model was represented by a neural network. The general properties of these models are quite similar to the one presented here.

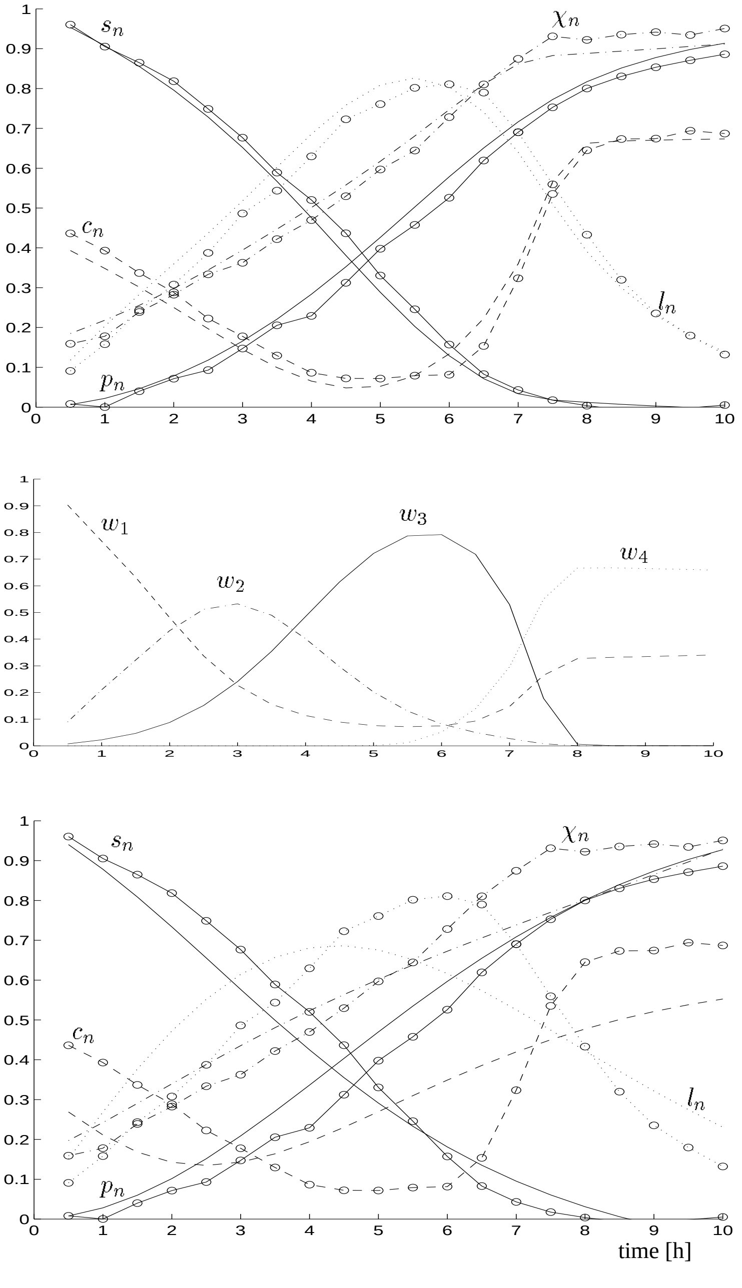


Figure 6.9: Top: Trajectories with circles are generated by the "true system", while the others trajectories are simulations of the identified model with four operating regimes and mechanistic local models. Middle: The relative weight of the various local models in the interpolation. Bottom: A simulation with an identified model with only one operating regime (for the purpose of comparison).

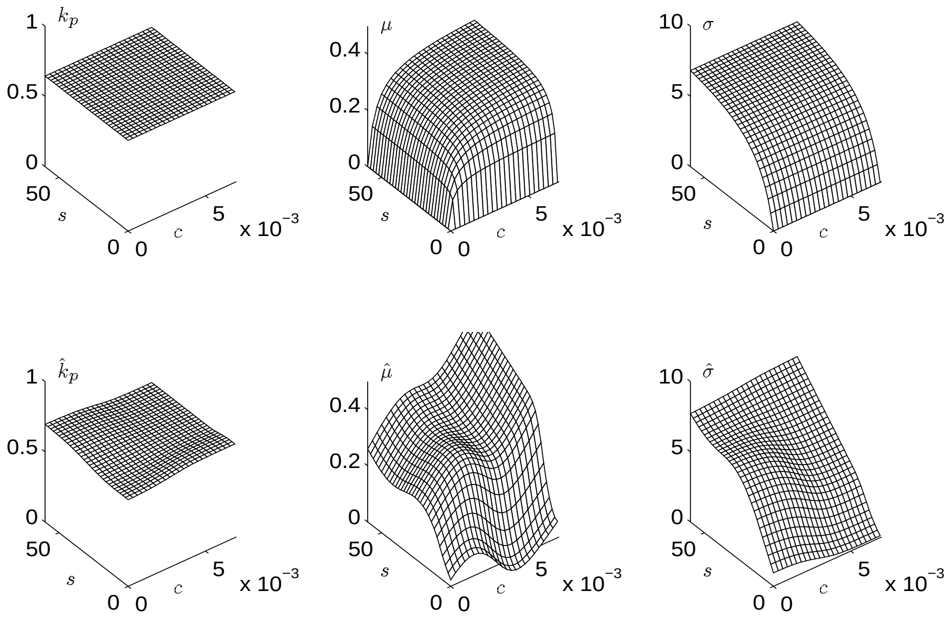


Figure 6.10: Top: The true kinetic model used in simulator. Bottom: Identified kinetic model.

### 6.4.2 Simulation Example: A pH-neutralization Tank

Consider a pH-neutralization tank, where there are three influent streams and one effluent stream:

- Influent acid stream $Q_1$ ($HNO_3$)
- Influent buffer stream $Q_2$ ($NaHCO_3$)
- Influent base stream $Q_3$ (NaOH and traces of $NaHCO_3$)
- Effluent stream $Q_4$

For our simulation study we use the model of Hall and Seborg (1989) to simulate the "true system". It is based on the assumptions of perfect mixing, constant density, fast reactions and completely soluble ions. Only the following chemical reactions are modeled

$$\begin{aligned} H_2O &\rightleftarrows OH^- + H^+ \\ H_2CO_3 &\rightleftarrows HCO_3^- + H^+ \\ HCO_3^- &\rightleftarrows CO_3^{2-} + H^+ \end{aligned}$$

because $HNO_3$ is a strong acid and NaOH is a strong base. Chemical reaction invariants for the process are (Gustafsson and Waller 1983)

$$\begin{aligned} W_a &= [H^+] - [OH^-] - [HCO_3^-] - 2[CO_3^{2-}] \\ W_b &= [H_2CO_3] + [HCO_3^-] + [CO_3^{2-}] \end{aligned}$$

Using the equilibrium equations

$$\begin{aligned} K_{a1} &= [HCO_3^-][H^+][H_2CO_3]^{-1} \\ K_{a2} &= [CO_3^{2-}][H^+][HCO_3^-]^{-1} \\ K_w &= [H^+][OH^-] \end{aligned}$$

an implicit equation for $[H^+]$ is found

$$W_a = [H^+] - \frac{K_w}{[H^+]} - W_b \frac{K_{a1}/[H^+] + 2K_{a1}K_{a2}/[H^+]^2}{1 + K_{a1}/[H^+] + K_{a1}K_{a2}/[H^+]^2}$$

Solving this equation for $[H^+]$, we can find pH $= -\log_{10}[H^+]$. A total mass balance for the tank gives

$$A\dot{h} = Q_1 + Q_2 + Q_3 - c\sqrt{h - h_0}$$

where $c$ is a valve constant, $A$ is the tank cross-section area, $h$ is the freely varying tank level, and $h_0$ the vertical distance from the bottom of the tank to the outlet. Component balances gives

$$\begin{aligned} hA\dot{W}_a &= Q_1(W_{a1} - W_a) + Q_2(W_{a2} - W_a) + Q_3(W_{a3} - W_a) \\ hA\dot{W}_b &= Q_1(W_{b1} - W_b) + Q_2(W_{b2} - W_b) + Q_3(W_{b3} - W_b) \end{aligned}$$

where $W_{ai}$ and $W_{bi}$ are chemical reaction invariants of the $i$-th stream. The variables are defined in Table 6.1.

The streams $Q_1$ and $Q_2$ are fixed, while $Q_3$ is controlled. The data used for model identification is shown in Fig. 6.11a, while the data sequence used for validating the model is shown in Fig. 6.11b. Both data sequences are noise-free, and the sampling interval is 15 $s$. Clearly, the validation data covers a significantly wider operating range than the identification data, which is typical for many applications.

### Alternative 1. A Simple Mechanistic Model

Let us first consider a very simple model of the tank, based on neglecting the buffer stream in the model. Hence, the streams are

- Influent acid stream $Q_1$ ($HNO_3$)
- Influent base stream $Q_3$ (NaOH)
- Effluent stream $Q_4$

The only reaction considered in this model is

$$H_2O \rightleftharpoons OH^- + H^+$$

A reaction invariant is $W = [H^+] - [OH^-]$, and an implicit equation for $[H^+]$ is

$$W = [H^+] - K_w/[H^+] \tag{6.19}$$

Table 6.1: Symbols, constants and variables used in models.

| Symbol | Variable | Nominal value |
|---|---|---|
| $A$ | Tank area | 207 $cm^2$ |
| $h$ | Tank level | 14 $cm$ |
| $h_0$ | Tank outlet level | 5 $cm$ |
| $Q_1$ | Acid flow-rate | 16.6 $ml/s$ |
| $Q_2$ | Buffer flow-rate | 0.55 $ml/s$ |
| $Q_3$ | Base flow-rate | 15.6 $ml/s$ |
| $[HNO_3]_1$ | Acid concentration in acid stream | 0.003 $mol/l$ |
| $[NaHCO_3]_3$ | Buffer concentration in base stream | 0.00005 $mol/l$ |
| $[NaOH]_3$ | Base concentration in base stream | 0.003 $mol/l$ |
| $[NaHCO_3]_2$ | Buffer concentration in buffer stream | 0.03 $mol/l$ |
| $c$ | Valve constant | 8 $ml/s\sqrt{cm}$ |
| $pK_{a1}$ | $-\log_{10}K_{a1}$ | 6.35 |
| $pK_{a2}$ | $-\log_{10}K_{a2}$ | 10.33 |
| $pK_w$ | $-\log_{10}K_w$ | 14.00 |
| $W_{a1}$ | $[HNO_3]_1$ | 0.003 $mol/l$ |
| $W_{a2}$ | $-[NaHCO_3]_2$ | -0.03 $mol/l$ |
| $W_{a3}$ | $-[NaHCO_3]_3 - [NaOH]_3$ | -0.00305 $mol/l$ |
| $W_{b1}$ | | 0 |
| $W_{b2}$ | $[NaHCO_3]_2$ | 0.03 $mol/l$ |
| $W_{b3}$ | $[NaHCO_3]_3$ | 0.00005 $mol/l$ |
| $W_1$ | $[HNO_3]_1$ | 0.003 $mol/l$ |
| $W_2$ | | 0 |
| $W_3$ | $-[NaOH]_3$ | -0.003 $mol/l$ |

The mass balances are

$$A\dot{h} = Q_1 + Q_3 - c\sqrt{h - h_0} \tag{6.20}$$
$$hA\dot{W} = Q_1(W_1 - W) + Q_3(W_3 - W) \tag{6.21}$$

This model is validated against the validation data by a simulation of the response to the validation data input sequence in Fig. 6.12a. We observe that the model is accurate for high pH values and low pH values, but inaccurate for intermediate values of pH. In fact, this model is a reasonable model of almost any neutralization process at high and low pH values. Of course, what is high and low in this context will depend on the particular process.

**Alternative 2. An ARX Model**

Next, the ARX model

$$pH(t+1) = 7.06 + 0.78(pH(t) - 7) + 0.10(Q_3(t) - 15)$$

is identified using the least squares method and the identification data. A simulation of the model's response to the validation data input sequence is shown in Fig.

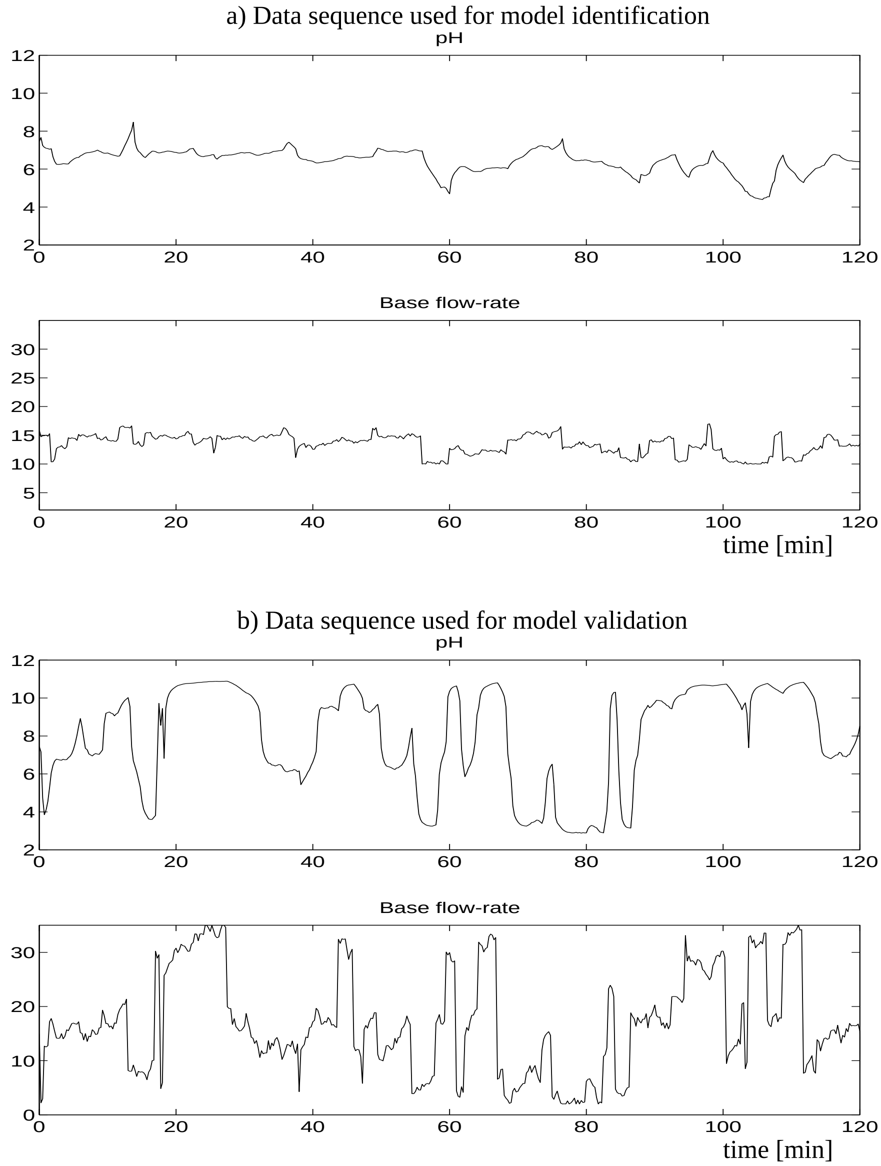


Figure 6.11: Simulated data sequences.

6.12b. The prediction error is clearly smallest for intermediate pH values, while it is completely wrong for large pH values. An NARX model based on a number of local ARX models, and identified using the structure identification algorithm gives better fit for intermediate pH values, but even less convincing predictions when extrapolated to regimes with large pH value.

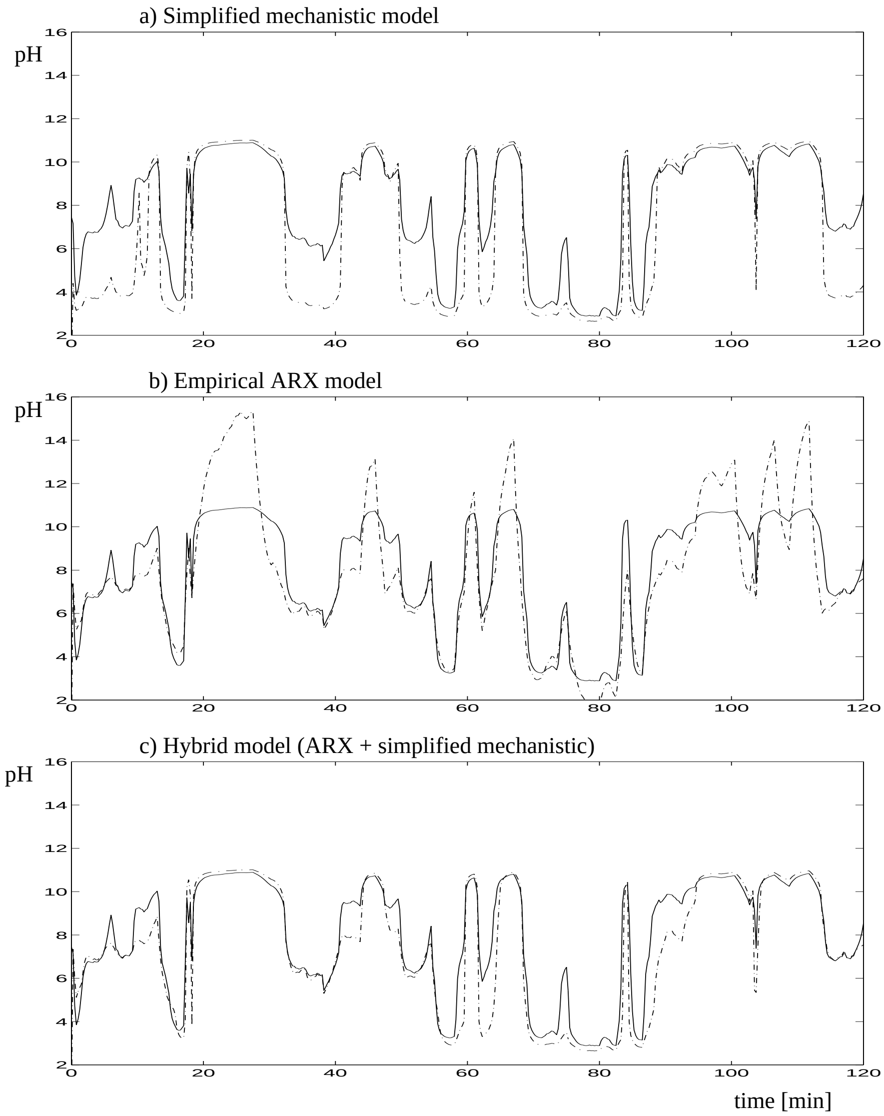


Figure 6.12: Simulations of the three models' responses to the validation data input sequences. Solid curves ( — ) are the "true system", while dashed-dotted curves ($- \cdot - \cdot -$) are model simulations.

### Alternative 3. A Hybrid Model

Finally, consider the case when the structure identification algorithm is allowed to choose between the local ARX model structure, and the simplified local mechanistic model structure. The hybrid model is a discrete-time model, and the operating point is chosen as $z(t) = pH(t)$. The continuous-time mechanistic sub-model is integrated numerically using the explicit Euler method over the 15 $s$ sampling interval (unit time-step of the ARX model). The problems discussed in section 3.1.1 regarding transition between different state-spaces in the different regimes arises here. It is possible to compute the model state $W$ directly from (6.19) and the pH value, i.e. there exists a one-to-one mapping from the output to this state variable. The state $h$ appears to be somewhat more difficult to deal with, since there exists no such mapping for this state. Fortunately, $h$ is globally observable from the inlet and outlet flow-rates, and by keep integrating (6.20) even if the mechanistic model is not relevant at the current operating point, the problem is resolved.

We use the structure identification algorithm based on $\overline{J}_{BOOT}$, where $p_{\mathrm{future}}(pH)$ is chosen to reflect a uniform distribution of pH values between 2 and 12, and $p_{\mathrm{past}}(pH)$ is a normal distribution estimated using the identification data, cf. Chapter 4. Notice that we have introduced a simplification by restricting the probability distribution to depend on the operating point only. The algorithm identifies a hybrid model that consists of three local models. The first is the simple mechanistic model that corresponds to an operating regime with high pH value, i.e. above pH equal to 9.5. The second local model is an empirical ARX model, that is valid at intermediate pH values, i.e. between 4.5 and 9.5. The third local model is the mechanistic model that corresponds to an operating regime with low pH value. The relative weight in the interpolation is shown in Fig. 6.13, and a simulation of the model's response to the validation data input sequence is shown in Fig. 6.12c. We see that the prediction accuracy has improved significantly for most operating conditions, but we also see that there exists room for improvement by tuning the regimes and local model validity functions.

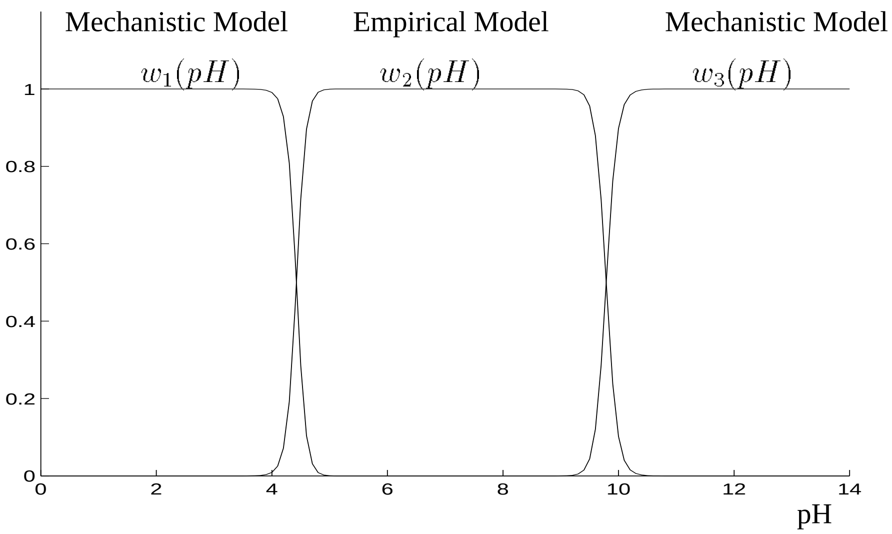


Figure 6.13: Interpolation functions for the hybrid model.

### 6.4.3 Experimental Results: Hydraulic Manipulator

A data sequence with 8000 samples, logged from a hydraulic TR4000 robot (Kavli 1990) from ABB Trallfa Robotics A/S, was used to find a model describing the inverse dynamics

$$\tau(t) = f(q(t), \dot{q}(t), \ddot{q}(t)) + e(t)$$

of a joint of this robot, where $\tau(t)$ is the control signal to the servo valve, $q(t)$ is joint position, and $e(t)$ is equation error. The joint position was logged at a sampling rate of 100 Hz while the robot was moving along a randomly generated trajectory. The joint velocity and acceleration was estimated by low-pass filtering and numerical differentiations. The predictions of an estimated linear model was subtracted from the data to emphasize the non-linearities. According to Kavli (1990), the non-linearities are mainly due to variations in the momentum arm of the hydraulic cylinder, non-linear damping, and non-linear pressure gain characteristics due to varying flow-rates in the servo valve. In addition, 1000 independent samples was used for validating the model. A number of models were identified, based on the structure identification algorithm, a least squares parameter estimation algorithm, local linear model structure, and a Gaussian basis-function with $\gamma = 1$. The results are summarized and compared to the spline models of Kavli (1993), the algorithm for local model identification of Murray-Smith (1994*a*), the MARS algorithm (Friedman 1991), and some neural network models reported by Carlin, Kavli and Lillekjendlie (1994) in Table 6.2. The table shows that the structure identification algorithm is able to find an adequate model with a small number of parameters while maintaining the high accuracy of the models found by the other empirical modeling algorithms. In all cases, only the parameters corresponding to local model parameters or basis-function coefficients are counted in the table. Notice that according to the FPE criterion with $\kappa = 1$, the data sequence allows more degrees of freedom to be added to the identified model structure. This was not pursued due to the computational complexity. The operating point was chosen as $z = (q, \dot{q}, \ddot{q})$ although complementary identification experiments showed that $z = (\dot{q}, \ddot{q})$ was sufficient to capture most of the non-linearities.

### 6.4.4 Discussion of Examples

In the fermenter example, we have identified operating regime based models using 1) a semi-empirical local linear model structure with some structural zeros, based on a linearized mass balance, and 2) a mechanistic local model structure based on a mass balance with simplified empirical equations for the unknown reaction rates. The results where qualitatively similar, with roughly the same prediction performance and interpretation of the identified regimes. With mechanistic local model structure it was possible to find a global approximation to the reaction rates directly from the identified local reaction rate equations. A study of these equations gives significant understanding of the reaction mechanisms and can guide the development of a mechanistic kinetic model. In other words, the hybrid model may be a useful intermediate step on the search for a mechanistic model. Notice

| Model | Comments | Num. Param. | NRMSE |
|---|---|---|---|
| ASMOD‡ | Quadratic Spline Basis | 561 | 15 % |
| MARS | | 155 | 16 % |
| Local | Local linear models | | |
| | $\gamma = 1$, $n^\star = 2$, Heuristics 1, 2, and 3 | 80 | 17 % |
| Local | Local linear models | | |
| | $\gamma = 1$, $n^\star = 1$, Heuristic 3 | 80 | 17 % |
| Local* | Local linear models | NA | 17 % |
| MARS | | 53 | 17 % |
| Local | Local linear models | | |
| | $\gamma = 1, n^\star = 2$, Heuristics 1, 2 and 3 | 40 | 18 % |
| Local | Local linear models | | |
| | $\gamma = 1, n^\star = 1$, Heuristic 3 | 40 | 19 % |
| RBF+ | Gaussian radial basis-functions | 112 | 19 % |
| MARS | | 16 | 20 % |
| ASMOD‡ | Quadratic Spline Basis | 48 | 20 % |
| Local | Local linear models | | |
| | $\gamma = 1, n^\star = 2$, Heuristics 1, 2 and 3 | 20 | 23 % |
| NN+ | Sigmoidal Neural Network (3-20-1) | 101 | 23 % |
| Local | Local linear models | | |
| | $\gamma = 1, n^\star = 1$, Heuristic 3 | 20 | 26 % |
| NN+ | Sigmoidal Neural Network (3-5-1) | 26 | 26 % |

Table 6.2: Results of applying various identification algorithms on the hydraulic manipulator joint data. The result marked with $*$ is with the identification algorithm of Murray-Smith (1994a). Results marked with $+$ and ‡ are from (Carlin, Kavli, Lillekjendlie, 1994) and (Kavli, 1993), respectively. The NRMSE criterion is defined as the square root of the ratio of the average squared one-step-ahead prediction error to the variance of the output, using the independent test data.

that in this example, the form of the models are equivalent to a mechanistic model, and does therefore not impose any serious restrictions on the applicability. The underlying modeling principle is quite general. First ones derives a set of balance equations with a number of internal variables that are hard to model. Such variables are typically related to reaction kinetics, thermodynamics, fluid flow, and mass- and heat-transfer. Since such variables must often be modeled empirically anyway, we suggest either simple local approximations to these relations, or simple local approximations of the balance equations themselves. If a sufficiently large amount of empirical data is available, this procedure may give a transparent, reliable and inexpensive model.

The pH neutralization example shows that the proposed model structure identification algorithm is able to identify a sensible hybrid model structure when being allowed to choose between an oversimplified local mechanistic model structure and a local linear empirical model structure. The algorithm utilizes the a priori information that the model will be applied under a significantly wider range of operating conditions than reflected in the data used for model identification. The identified model consists of local mechanistic models in the regimes where there are no data, and local empirical models in regimes that contain sufficiently large amounts of data. The hybrid model is therefore able to fit the data well, as an empirical model, and provides reliable extrapolation, as a mechanistic model. This example clearly illustrates the power of the operating regime based modeling framework and the structure identification algorithm: In addition to a data sequence, only rather elementary process knowledge is applied. Constructing a complete mechanistic model that take into consideration all possible chemical components in the tank can be a resource demanding task, in more complicated examples than this. On the other hand, a good empirical model would require a larger amount of empirical data that would cover a significantly wider range of operating conditions. In some cases, providing such additional data could require extra resources. Clearly, the structure identification algorithm is not necessary for such a simple problem as here, and regimes chosen "by hand" leads to a similar model (Foss and Johansen 1993).

In the hydraulic manipulator example, the large number of regimes makes the interpretation of the empirical model more difficult than in the previous example. The model should be viewed as a black box. Hence, this example mainly serves as a benchmark that shows that the accuracy with the operating regime based modeling approach is comparable to some of the most popular empirical modeling algorithm from the literature.

There are more examples available in the literature (both simulated and experimental) that illustrate the use of various system identification algorithm and local empirical models, typically local linear regression or local ARX models. These include river flow modeling (Sugeno and Kang 1988), environmental modeling (Nakamori and Ryoke 1994), multi-layer incinerator modeling (Sugeno and Kang 1986, Nakamori et al. 1992), a converter in a steel-making process (Takagi and Sugeno 1985), aluminum roll mill (Murray-Smith 1994*a*), and in chemometrics (Næs and Isaksson 1991, Næs 1991). In most of these examples it is justified that the model performs better than simpler empirical models, but the models are generally not compared to other complex empirical or mechanistic models. Of course, the same kind of argument can be used against some of the examples in this thesis.

## 6.5 Discussion

### 6.5.1 A Priori Knowledge: What is required, and what can be incorporated?

The system knowledge required with the proposed approach is quite reasonable. First of all, an operating point space $Z$ is required. In many cases, it is possible to choose $Z$ equal to a subspace or sub-manifold of the input space, as discussed in Chapters 2 and 3. The design of $Z$ need not be based solely on a priori knowledge, but can in addition consider the distribution of the data $\mathcal{D}_l$. Quite often, there are collinearities or correlations in the data, so that $\mathcal{D}_l$ can be embedded in a subspace or sub-manifold of considerably lower dimension than the input space. In that case, $z$ need not be of higher dimension than this embedding. As we have seen, some system knowledge will often make it possible to reduce the dimension of $z$ considerably. This is important, since it may reduce the complexity of the model, improve its transparency, and also reduce the computational complexity for the structure identification algorithm considerably.

A set of local model structure candidates must be specified. If no a priori knowledge exists to support one choice over the other, one will typically choose local linear model structures of various order and possibly with structural zeros as default, since linear models are well understood and possible to interpret. Moreover, a linear model will always be a sufficiently good approximation locally, provided the system is smooth, and the regimes are small enough. On the other hand, if there is substantial a priori knowledge available in terms of mechanistic local model structures, these can be included. Such local model structures may for example be based on simplified mass- and energy-balances.

### 6.5.2 A Posteriori Knowledge: What can be extracted from the model?

The purpose of a model can be diverse, e.g. system analysis, design, optimization, prediction, control, or diagnosis. In many applications, it is important that the model can be easily interpreted and understood in terms of the system mechanisms. With empirical models, which are often based on black-box model representations, this is often a hard or impossible task. However, the approach presented here gives transparent empirical models because

- local models are simple enough to be interpreted,
- the operating regimes are a qualitative high-level description of the system that is close to engineers and operators knowledge representation.

Finally, suppose the local model structures are all linear. We can examine the resulting operating regimes, and observe which variables cause non-linear behavior. Furthermore, by examining the parameter variations between local models corresponding to neighboring regimes, or examining the relative size of the regimes, we may get an indication on how strong the non-linearities are in the different regimes.

### 6.5.3 Related Work

The proposed identification algorithm has a number of relatives. Probably, the most well known algorithm is the MARS algorithm of Friedman (1991). This is a local search algorithm ($n^\star = 1$) that searches for a set of natural spline basis-functions. The MARS algorithm also contains a model reduction part that reduces the model structure to an optimal size. A similar algorithm is the ASMOD algorithm of Kavli (1993), with tensor product B-spline basis-functions. The structure identification part of this algorithm is based on stepwise modification of the basis-functions using one out of three competing strategies. A fundamental difference from our approach is that the result of one modification may be more than one additional basis-function, or a reduction in the number of basis-functions. The CART algorithm of Breiman et al. (1984) searches for a decomposition into regimes and builds a piecewise constant model. The structure identification part first builds a too complex model, which is subsequently reduced. Similar basis-function trees are constructed by the algorithms of Sanger (1991), and Omohundro (1987). Common to all these algorithms is that they consider non-linear regression problems, and try to describe the data using as few dimensions of the input space as possible, increasing the dimension of the domain of the basis function only when necessary. This is motivated by the curse of dimensionality. The same problem also apply to the present approach, but to a smaller extent, since the local models are not constants, but functions of the generalized input. This may allow the operating point space to be of a smaller dimension than the input space, which reduces the curse of dimensionality.

Local linear models are also applied by Jones and co-workers (1991) and by Stokbro et al. (1990) together with a clustering algorithm to determine the location of the local models. Jacobs et al. (1991) and Jordan and Jacobs (1993) use a parameterized regime description and a hierarchical estimator to estimate the regime parameters simultaneously with the local model parameters. An algorithm based on local linear models and decomposition of regimes where the system appears to be more complex than the model, has been suggested by Murray-Smith and Gollee (1994). Pottmann et al. (1993) have proposed a model representation based on local polynomial models and smooth interpolation. The structure identification algorithm is based on a orthogonal stepwise regression procedure (Kortmann and Unbehauen 1988) that sequentially adds or discards terms in the local polynomial model according to their significance.

When the interpolation functions are chosen as the characteristic functions of the regime-sets $Z_i$, a piecewise linear model results. The resulting model will not be smooth, and may not even be continuous, which may be a requirement in some applications. Also, we have experienced that smooth interpolation between local linear models usually gives better model fit compared to a local constant or piecewise linear model with the same number of parameters. The local linear modeling approach combined with a fuzzy set representation of the regimes also leads to a model representation with interpolation between the local linear models (Takagi and Sugeno 1985). In that case, it is the fuzzy inference mechanism that implicitly gives an interpolation. Structure identification algorithms based on clustering (Bezdek et al. 1981*a*, Bezdek et al. 1981*b*, Yoshinari et al. 1993, Yager and Filev 1993, Nakamori and Ryoke 1994) and a heuristic local search algorithm (Sugeno and Kang 1988) have been proposed in this context. In this case, the $b$- and $\rho_i$-functions are interpreted as membership functions for fuzzy sets. A statistical pattern recognition approach with multiple models leads to a similar representation based on a piecewise linear model and discriminant functions to represent the regime boundaries, (Skeppstedt et al. 1992). Finally, Sørheim (1990) has suggested a model representation with neural nets as local models and a structure identification algorithm based on pattern recognition. The pattern recognition algorithm will detect parts of the input space in which the model fit is inadequate, in some sense, and refine the model locally.

With this large body of literature in mind, one may ask: What are the contributions and improvements represented by the present approach? We have attempted to take the most attractive features from the algorithms in the literature and combined these into one algorithm. The algorithm of Sugeno and Kang (1988) is the closest relative to the present algorithm. The main difference is the extra flexibility and effort applied to find a closer to optimal model with the present algorithm. In addition, we have emphasized interpretability of the resulting model, flexibility with respect to incorporation of prior knowledge, and a transparent modeling and identification process that is close to engineering practise. The price we have to pay is a computer intensive algorithm. Some may also argue that the algorithm is too flexible and not completely automatic, and as a result it may be difficult to apply for inexperienced users. However, it is our view that real world applications require perhaps even more flexibility and a less automated approach.

### 6.5.4 Limitations and Possible Improvements

One major limitation of the proposed algorithm is the restriction that the regimes must be $d$-dimensional boxes with orthogonal edges. This implies that regimes that more naturally could have been described with a more complex regime boundary, must be represented as two or more separate regimes, with separate local models. Hence, more local models than necessary are required. On the other hand, the introduction of complex descriptions of the regime boundaries will increase the number of parameters needed to represent these boundaries or local model validity functions and lead to a more complex identification problem. However, the work of e.g. Jordan and Jacobs (1993) shows that this is feasible. What is the "optimal" regime representation, is currently not clear to the authors. We believe, however, that the proposed representation is a fair tradeoff between flexibility and simplicity. Its major strength is that the simple regime description combined with simple local models gives a transparent model. An optimization of the $\zeta_i$- and $\lambda_i$-parameters after the model structure and local model parameters have been fixed has been implemented. For the test-cases, the optimization gave only marginal improvements over the heuristic choices. One may conclude that the algorithm is robust with respect to the location of the splitting hyper-planes and the ad.hoc. choice of local model validity function parameters.

In the presented strategies and heuristics, no concerns has been made about which model structure candidates to evaluate first. In particular when $n^\star$ is large, it is important with a best-first search (Pearl 1984) since one cannot hope to evaluate all possible candidates in most practical cases. First locally optimal models will be found, and subsequent improvements will by found until eventually the globally optimal model will be found. The power of a best-first algorithm is that it can be interrupted at any time with a high probability of returning a good result, which is not generally true if the algorithm does not examine the most promising candidates first. Again, which are the best candidates is not known a priori, so some heuristic must be applied to find the most promising candidates at each step.

An alternative strategy could be based on first to construct a model by stepwise decomposition, and then to reduce the model using stepwise merging. The idea is that some of the locally optimal, but globally suboptimal, decomposition steps can be undone during the reduction part of the algorithm. The efficiency of this modification is strongly restricted by the fact that the regimes must be merged pairwise, and the selected regime representation will seriously restrict which regime pairs that can be merged. Furthermore, since the forward step leads to a parsimonious model, the backward part of the algorithm will often find that very little or no model reduction is possible, and the strategy gives no improvements over the proposed algorithm. This problem can be overcome by allowing the forward part of the algorithm to generate a too complex model. However, one can not guarantee that such a strategy is an improvement compared to the proposed algorithm. Alternatively, one can at each step let decomposition and merging compete when deciding what the next model structure shall be (Kavli 1993).

Another class of feasible algorithms for the structure optimization problem are stochastic optimization algorithms like simulated annealing (Kirkpatrick, Gelatt Jr. and Vecchi 1983) and genetic algorithms (Goldberg 1989). Such algorithms can handle both integer and real variables, and will under weak conditions converge to a global minimum. It is the stochastic nature of the algorithms that allows escape from local minima. Unfortunately, these algorithms converges very slowly. Yet another class of algorithms is dynamic programming, (Bellman and Dreyfus 1962, Bellman 1961*b*). The advantage of dynamic programming is that the fact that the model structure set is not a tree, allows the computational complexity to be reduced. However, dynamic programming suffers from an unreasonable demand for computer memory.

### 6.5.5 The Fundamental Assumptions

In general, the fundamental assumptions behind empirical modeling is

- the empirical data is not too contaminated by noise and other unpredictable phenomena, and
- the data set is complete in the sense that it contains a sufficient amount of observations from all interesting operating conditions and system variables.

Unfortunately, these assumptions are often not met in practical applications. The proposed algorithm should therefore be applied with care, and as a part of a computer aided modeling environment that allows flexible incorporation of prior knowledge, and not as an automatic modeling algorithm. Moreover, one should undertake a detailed study of the robustness of the algorithm with respect to contaminated, sparse, and incomplete data, in particularly for high dimensional and otherwise complex modeling problems.

# Chapter 7

# Identification of Non-linear Systems using Prior Knowledge and Empirical Data - An Optimization Approach

Identifying a model from a finite sample of observations without any prior knowledge about the system is an ill-posed problem, in the sense that a unique model may not exist, or it may not depend continuously on the observations (Tikhonov and Arsenin 1977). Indeed, without any prior knowledge except the observations, nothing is known about the system behavior in states between the observations. Hence, even if the data are not corrupted by noise, the model can behave arbitrarily between the observations, and it should be clear that such a model will not be unique, nor will it depend continuously on the data. The problem is essentially that the finite amount of data does not constrain the model set sufficiently. Fortunately, a minimum of prior knowledge will in general provide the necessary constraints, and there exists a theory that will aid reformulation of the problem into a well-posed one, namely regularization theory (Tikhonov and Arsenin 1977). The prior knowledge or assumption that is often used, is that the system has certain smoothness properties. This will constrain the system behavior in a neighborhood of each observation from changing abruptly, and some interpolation and extrapolation of the observations can be justified. Such an assumption is reasonable for a large class of real world systems, but certainly not for all. The smoothness assumption can be incorporated explicitly by the use of a possibly over-parameterized or nonparametric model, and a penalty on non-smoothness in the identification criterion (Bertero, De Mol and Pike 1985, O'Sullivan 1986, Dyn 1989, Madych and Nelson 1990, Wahba 1990). The penalty will reduce the effective number of parameters in the model set. However, the smoothness assumption is usually made implicitly through the choice of a parameterized model structure that is simple in the sense that it has few unknown parameters compared to the number of observations. In this case the dimension of the model space is reduced a priori such that the data set can determine the unique model in the model set.

It is usually desirable to effectively employ all the available prior knowledge and empirical data. Increased amounts of prior knowledge will make the identification problem better conditioned, since more constraints on the model set are imposed. If the prior knowledge is correct, this will in general lead to a better model that is robust against a deficient or incomplete data set. Herein lies the motivation for our work, and we will show that the regularization framework is a convenient tool for incorporation of prior knowledge into semi-empirical models. In particular, we will allow prior knowledge, empirical data, and desired properties of the model in the form of

- Smoothness of the model behavior.
- Partially or completely known models.
- Constraints on model structure, variables and behavior, like stability and linearity.
- Empirical data (steady-state, time series).

The various pieces of knowledge or properties can be locally valid under certain operating conditions only, or globally valid. Moreover, varying levels of accuracy, completeness, and reliability of the knowledge can be handled.

The unifying framework we propose here is an optimization formulation of the modeling and identification problem, where the different pieces of knowledge will enter as penalty terms or constraints in the optimization formulation. This formulation of the problem has found its inspiration in (Thompson and Kramer 1994), where it is discussed how different kinds of knowledge can be used to structure the model and optimization criterion, but in a significantly less general framework than here. Tulleken (1993) has also suggested the use of constraints derived from system knowledge to improve the model, in the context of linear system identification. Another source of inspiration is (Girosi, Jones and Poggio 1994), which analyze how different non-smoothness penalties lead to different basis-functions in a series expansion of the model. The present work can be viewed as a unification and extension of these approaches. Contrary to the static model formulation in (Girosi et al. 1994, Thompson and Kramer 1994), we formulate the problem for dynamical systems described by MISO NARMAX (non-linear ARMAX) models.

The chapter is a reprint of (Johansen 1994*c*), and organized as follows. In section 7.1 the problem is formulated mathematically and some simple motivating examples are given. Next, in section 7.2, the exact solution to a simplified modeling problem is found explicitly using function space methods, and this solution is discussed in detail, before numerical procedures that solve the general problem are discussed in section 7.3. In section 7.4, we discuss how to tune the parameters in the optimization criterion and present alternative criteria and procedures for this. Section 7.5 contains a fairly complete semi-realistic simulation example that illustrates the power of the approach, and also the effect of different kinds of prior knowledge and empirical data on the model. The chapter ends up in a discussion and some conclusions.

## 7.1 Problem Formulation

In this work we study the identification of NARMAX models (Leontaritis and Billings 1985) of the form

$$y(t) = f(\psi(t-1)) + e(t) \tag{7.1}$$

where $\psi(t-1) = (y(t-1), ..., y(t-n_y), u^T(t-1), ..., u^T(t-n_u), e(t-1), ..., e(t-n_e))^T$ is the information vector. For simplicity we assume the system output $y(t)$ is a scalar, while the system input $u(t) \in R^r$ may be $r$-dimensional. Let $\psi(t) \in \Psi \subset R^n$, where $\Psi$ is a compact subset in which we are interested in modeling the system's input/output behavior. Furthermore, let $\mathcal{F}$ be an inner product space containing smooth functions on $\Psi$

$$\mathcal{F} = \{f \in L_2(\Psi) \mid f \text{ is smooth}\}$$

The smoothness in the definition of $\mathcal{F}$ should be interpreted as existence and continuity of sufficiently high order derivatives of the functions in $\mathcal{F}$. For the purpose of model identification, a sequence of $l$ input/output observations

$$((u(1), y(1)), (u(2), y(2)), ..., (u(l), y(l)))$$

may be available. Moreover, we allow a possibly inaccurate default model $M_a$ to be available a priori, together with soft equality constraints $Qf \approx q$, where $Q : \mathcal{F} \to \mathcal{Q}$ is an operator, $\mathcal{Q}$ is an inner product space, and $q \in \mathcal{Q}$. Furthermore, we allow hard constraints $Hf \leq 0$ and $Pf = 0$, where $H : \mathcal{F} \to \mathcal{H}$ and $P : \mathcal{F} \to \mathcal{P}$ are operators, and $\mathcal{H}$ and $\mathcal{P}$ are inner product spaces. In addition, an operator $S : \mathcal{F} \to \mathcal{S}$ may be given, where $\mathcal{S}$ is an inner product space and $S$ indicates the non-smoothness of $f$. For example, $S$ may be a differential operator.

We are now in position to informally state the problem addressed in this work, namely to find the function $f \in \mathcal{F}$ that defines the NARMAX model (7.1) that is most consistent with the a priori knowledge and empirical data, in a well defined sense. For simplicity, we will assume the integers $n_y, n_u$ and $n_e$ are given.

Finding the best $f \in \mathcal{F}$ can be formulated as an optimization problem in the inner product space $\mathcal{F}$. The prior knowledge and empirical data are used to formulate an optimization criterion that penalizes

1. Mismatch between model prediction and the empirical data.
2. Non-smoothness of the model.
3. Mismatch between model and the default model.
4. Violation of the soft constraints.

In addition, we allow hard equality and inequality constraints to the optimization problem. We choose the criterion

$$J(f) = \frac{1}{l}\sum_{t=1}^{l} \varepsilon^2(t; f) + \lambda ||Sf||^2 + \gamma \mathcal{D}^2(f, M_a) + \beta ||Qf - q||^2 \tag{7.2}$$

subject to the constraints

$$Hf \leq 0 \quad \text{and} \quad Pf = 0$$

where $\lambda, \gamma$, and $\beta$ are non-negative constants, and the norms are induced by possibly weighted inner products an the various spaces. $\mathcal{D}(f, M_a)$ measures the distance between the model (7.1) defined by $f$ and the default model $M_a$, and $\varepsilon(t; f)$ can for example be the one-step-ahead prediction error defined by

$$\begin{aligned} \varepsilon(t; f) \;=\; & y(t) - f(y(t-1), ..., y(t-n_y), \\ & u(t-1), ..., u(t-n_u), \varepsilon(t-1; f), ..., \varepsilon(t-n_e; f)) \end{aligned}$$

or a multi-step-ahead prediction error. It must be assumed that this model on innovation form is asymptotically stable. Obviously, the prior knowledge and the choice of $\lambda, \gamma$ and $\beta$ will have great influence on the model. In particular, the problem of choosing these constants will be discussed in detail in section 7.4.

Before we proceed with the solution to this optimization problem, it will be illustrative to motivate this optimization formulation with examples of some prior knowledge and desired properties of the model that can be specified within this framework. A more complete example is provided in section 7.5.

**Example 1. Default model.**

Consider a default model of the form

$$\left.\begin{aligned} \dot{x} &= f_a(x, u) + v \\ y &= g_a(x) + w \end{aligned}\right\} \; M_a$$

The purpose of the default model is to provide a basic model that at least can be used during operating conditions when there are no empirical data or other more relevant prior knowledge available. The model $M_a$ can be used for one-step-ahead predictions, and a predictor can be formulated using an extended Kalman-filter, e.g. (Söderström and Stoica 1988)

$$\begin{aligned} \hat{x}(t|t) \;&=\; \hat{x}(t|t-1) + K(t)\left(y(t) - g_a(\hat{x}(t|t-1))\right) \\ \hat{x}(t+1|t) \;&=\; \int_{\tau=t}^{t+1} f_a(\hat{x}(\tau|t), u(\tau))d\tau \\ \hat{y}(t+1|M_a, t) \;&=\; g_a(\hat{x}(t+1|t)) \end{aligned}$$

where the Kalman-filter gain $K(t)$ is computed using the assumption that $v(t)$ and $w(t)$ are zero-mean white noise processes with known variance. A multi-step-ahead predictor can be formulated in a similar manner.

The distance $\mathcal{D}(f, M_a)$ can be defined in several ways. Since the two models are based on fundamentally different representations (NARMAX and continuous-time state-space), there is no direct way of comparing these two models in this case. Hence, the natural approach is to compare them indirectly by comparing their prediction performance when both are used to predict the response of the system to different inputs from different initial states. Hence, we suggest

$$\mathcal{D}^2(f, M_a) \;=\; \int_{x(t)}\int_{u(t)} \left(\hat{y}(t+1|f,t) - \hat{y}(t+1|M_a,t)\right)^2 \rho(t)dx(t)du(t)$$

where $\rho(t)$ is a weight-function that penalizes mismatch between the two models as a function of the operating condition at each time instant $t$.

□

**Example 2. Linear noise model.**

Suppose we want to restrict the model (7.1) to the set of models of the form

$$\begin{aligned} y(t) \ = \ & f'(y(t-1), ..., y(t-n_y), u(t-1), ..., u(t-n_u)) \\ & +c_1 e(t-1) + ... + c_{n_e} e(t-n_e) + e(t) \end{aligned}$$

where $c_1, c_2, ..., c_{n_e}$ are unknown constants. This gives a linear noise model that may be desirable because it simplifies estimation, analysis and application of the model. This restriction can be represented as a constraint

$$\frac{\partial^2}{\partial e(t-1)^2} f \ = \ \cdots \ = \ \frac{\partial^2}{\partial e(t-n_e)^2} f \ = 0$$

which can be written as a linear operator equation $Pf = 0$, by defining $\mathcal{P} = L_2^{n_e}(R^n)$ and

$$P = \begin{pmatrix} \frac{\partial^2}{\partial e(t-1)^2} \\ \vdots \\ \frac{\partial^2}{\partial e(t-n_e)^2} \end{pmatrix}$$

□

**Example 3. Known linear model.**

Next, suppose the origin is an equilibrium point for the system. A linearized model is given by $y(t) = \nabla_{\psi}^T f(0)\psi(t-1) + e(t)$. Now, suppose a linear (ARMAX) model

$$A(z^{-1})y(t) \ = \ B(z^{-1})u(t) + C(z^{-1})e(t) \tag{7.3}$$

of the system behavior near the origin is known. Penalty on deviations of the linearization of the non-linear model from this linear model can be formulated as a soft constraint of the form $Qf \approx q$. The distance between the two linear models can be measured using different metrics, for example in the frequency domain, time domain, or some model space. Perhaps the simplest is to measure their difference in the parameter space. Then we choose $\mathcal{Q} = R^n$, $Qf = \nabla_{\psi} f(0)$, and $q$ is the Riesz representation of (7.3) (written on the form $y(t) = q^T \psi(t-1) + e(t)$). A more robust model space metric would involve comparison of poles, zeros, and gain.

Notice that this formulation is very different from a default model $f_a(\psi(t-1)) = q^T \psi(t-1)$ where deviations from the default model will in general be penalized globally over the full range of operating conditions, while in this example the penalty is defined only in a point. Of course, the non-smoothness penalty will indirectly make this penalty effective in a neighborhood of this point.

If parts of a linear model is known, like an approximate pole valid in some operating regime, this knowledge can be incorporated in the same way.

□

**Example 4. Steady-state mass balance.**

Consider a system which has two input flows with rates $u_1(t)$ and $u_2(t)$, and one output flow with rate $y(t)$. If we keep the input flow rates fixed at some values $u_1(t) = \overline{u}_1$ and $u_2(t) = \overline{u}_2$, and the system reaches a steady state $\overline{y}$, then we may conclude that the following steady state mass balance hold:

$$\overline{u}_1 + \overline{u}_2 - \overline{y} = 0$$

Clearly, it may be desirable that the model have the same property. Let $Pf$ be the steady-state output of the system defined by $f$, then $Pf \in \mathcal{P} = L_2(R^2)$ is a function of the steady-state inputs $\overline{u}_1$ and $\overline{u}_2$. $P$ will in general be a non-linear operator that can be implemented by solving the difference equation exactly or approximated by a recursion that will be stopped when sufficiently close to convergence.

□

**Example 5. Stability.**

Another useful kind of prior knowledge is that the system is open-loop stable, or more precisely, that certain equilibrium points are asymptotically stable. If we seek a model of the form

$$y(t) = f(y(t-1), u(t-1)) + e(t)$$

then local asymptotic stability of an equilibrium point $(\overline{y}, \overline{u})$ is guaranteed provided the following inequality constraints are satisfied

$$-1 < \frac{\partial f}{\partial y(t-1)}(\overline{y}, \overline{u}) < 1$$

□

The choice of inner product on the different inner product spaces is of great importance. For example, the region of validity or relevance of a default model $M_a$ of the NARMAX form (7.1) defined by a function $f_a \in L_2(\Psi)$ can be represented by a weighted $L_2$ inner product

$$\langle f_1, f_2 \rangle_\rho = \int_{\psi \in \Psi} f_1(\psi) f_2(\psi) \rho(\psi) d\psi$$

where $\rho$ is a strictly positive weight function that integrates to one, and can be interpreted as our a priori knowledge about the validity of the default model as a function of the operating point, since we may define

$$\mathcal{D}^2(f, M_a) = ||f - f_a||_\rho^2 = \int_{\psi \in \Psi} (f(\psi) - f_a(\psi))^2 \rho(\psi) d\psi$$

A similar weighting may also be useful for the constraints, since they may be more or less relevant under different operating conditions. One may also represent knowledge about the smoothness of the system behavior as a function of the operating point by a weighted inner product. Alternatively, the weight can be interpreted as the relative desired accuracy of the model as a function of the operating point, if the weights are similar for all inner products. A useful class of function spaces are Sobolev spaces (Adams 1975), which are subspaces of $L_2(\Psi)$ with inner products like

$$< f_1, f_2 > = \int_{\psi\in\Psi} f_1(\psi) f_2(\psi) d\psi + \int_{\psi\in\Psi} \langle \nabla_\psi f_1(\psi), \nabla_\psi f_2(\psi) \rangle \, d\psi$$

if an objective is to make the Jacobian of the model to be a close approximation to the Jacobian of the system. This may be of particular interest if the model is used for control system design, where it is often the accuracy of the Jacobian of $f$ that is most important.

## 7.2 Optimization – Function Space Methods

In this section we will study the solution to the problem of minimizing the functional (7.2) in the function space $\mathcal{F}$, using calculus of variations. For simplicity, we will assume that there are no soft or hard constraints, the inner products are not weighted, $n_e = 0$ (we want to find an NARX model), and the default model is of the same form as the model we are seeking (i.e. an NARX model). Hence, the default model is defined by a function $f_a \in L_2(\Psi)$, which allows the simple distance measure $\mathcal{D}^2(f, M_a) = ||f - f_a||^2$. In other words, we consider a criterion

$$J(f) = \frac{1}{l} \sum_{t=1}^{l} \varepsilon^2(t; f) + \lambda ||Sf||^2 + \gamma ||f - f_a||^2$$

Like Girosi et al. (1994), we choose to represent the penalty on non-smoothness in the spatial frequency domain using the Fourier transform

$$\tilde{f}(\nu) = \int_{\psi\in\Psi} f(\psi) e^{j\psi^T\nu} d\psi$$

where $\tilde{f}$ is the Fourier transform of $f \in \mathcal{F}$, and $\nu$ is an $n$-dimensional vector that is interpreted as spatial frequency (frequency is here the inverse of distance in the information space, not time). Since $f$ is continuous, and $\Psi$ has compact support, this transform is well defined, and the inverse transform

$$f(\psi) = \frac{1}{(2\pi)^n} \int_{\nu\in R^n} \tilde{f}(\nu) e^{-j\psi^T\nu} d\nu$$

will be unique. On $L_2(\Psi)$ we apply the standard inner product

$$< f_1, f_2 > = \int_{\nu\in R^n} \tilde{f}_1(\nu) \tilde{f}_2^\star(\nu) d\nu$$

where the superscript in $\tilde{f}^{\star}(\nu)$ denotes complex conjugate. We restrict our attention to operators $S$ of the form (Girosi et al. 1994)

$$(Sf)(\psi) = \frac{1}{(2\pi)^n} \int_{\nu \in R^n} \frac{1}{\tilde{G}'(\nu)} \tilde{f}(\nu) e^{-j\psi^T \nu} d\nu \tag{7.4}$$

In other words, $\tilde{f}$ is filtered by $\left(\tilde{G}'(\nu)\right)^{-1}$, which will be chosen as a symmetric filter that amplifies high-frequency energy, since this gives penalty on high-frequency energy in $f$, which is an intuitive measure of non-smoothness. The induced norm is

$$||Sf||^2 = \int_{\nu \in R^n} \frac{\left|\tilde{f}(\nu)\right|^2}{\tilde{G}(\nu)} d\nu$$

where $\tilde{G}(\nu) = \left|\tilde{G}'(\nu)\right|^2$.

The case when $\gamma = 0$ is analyzed in detail by Girosi et al. (1994). Applying the mathematical tools in (Wahba 1990, Madych and Nelson 1990, Dyn 1989), one can show that the global minimum of $J$ can be represented as

$$f(\psi) = \sum_{t=1}^{l} \alpha_t G(\psi - \psi(t-1)) + \sum_{j=1}^{d_G} \beta_j \xi_j^G(\psi) \tag{7.5}$$

where $\alpha_1, \alpha_2, ..., \alpha_l$ and $\beta_1, \beta_2, ..., \beta_{d_G}$ are real constants, and the set of functions $\{\xi_1^G, \xi_2^G, ..., \xi_{d_G}^G\}$ satisfy

$$\frac{1}{\tilde{G}(\nu)} \tilde{\xi}_j^G(\nu) = 0 \tag{7.6}$$

The constant parameters $\alpha_1, \alpha_2, ..., \alpha_l$ and $\beta_1, \beta_2, ..., \beta_{d_G}$ can be found explicitly by substituting (7.5) back into the functional $J$ and minimizing this as a function of this finite number of parameters, cf. (Wahba 1990).

Observe that the $\xi_j^G$-terms in the model appear because certain functions will be "invisible" to the non-smoothness penalty, and such terms can be added to the solution without imposing any cost. For example, if $\tilde{G}'(\nu) = (j\nu)^{-2}$ then $\tilde{G}(\nu) = (-j\nu)^{-4} = \nu^{-4}$, and $(\tilde{G}(\nu))^{-1}$ is an operator that differentiates four times. Hence, $\xi_1^G, ..., \xi_{d_G}^G$ is a basis for the space of polynomials with order no greater than three. As a curiosity, the resulting basis functions are in this case cubic splines (Wahba 1990).

Next, we show that the case $\gamma > 0$ can be treated using the same technique. First, we rewrite the criterion $J(f)$ in terms of $\tilde{f}$ as

$$\begin{aligned} J(f) = {} & \frac{1}{l} \sum_{t=1}^{l} \left( y(t) - \frac{1}{(2\pi)^n} \int_{\nu \in R^n} \tilde{f}(\nu) e^{-j\nu^T \psi(t-1)} d\nu \right)^2 \\ & + \lambda \int_{\nu \in R^n} \frac{\tilde{f}(\nu) \tilde{f}^{\star}(\nu)}{\tilde{G}(\nu)} d\nu + \gamma \int_{\nu \in R^n} (\tilde{f}(\nu) - \tilde{f}_a(\nu))(\tilde{f}^{\star}(\nu) - \tilde{f}_a^{\star}(\nu)) d\nu \end{aligned}$$

$J$ is a strictly convex functional that must have a unique global minimum. Using calculus of variations (e.g. (Luenberger 1969)), we first find the Gateaux variation

$$\begin{aligned}\delta J(f;\Delta f) \;=\; & -2\frac{1}{l}\sum_{t=1}^{l}(y(t)-f(\psi(t-1)))\int_{\nu\in R^n}\Delta\tilde{f}(\nu)e^{-j\nu^T\psi(t-1)}d\nu \\ & +\lambda\int_{\nu\in R^n}\frac{\tilde{f}(\nu)\Delta\tilde{f}^\star(\nu)+\tilde{f}^\star(\nu)\Delta\tilde{f}(\nu)}{\tilde{G}(\nu)}d\nu \\ & +\gamma\int_{\nu\in R^n}\left(\Delta\tilde{f}(\nu)\left(\tilde{f}^\star(\nu)-\tilde{f}_a^\star(\nu)\right)+\Delta\tilde{f}^\star(\nu)\left(\tilde{f}(\nu)-\tilde{f}_a(\nu)\right)\right)d\nu\end{aligned}$$

for an arbitrary perturbation $\Delta f \in \mathcal{F}$. It is straightforward to show that the two last terms are real, hence

$$\begin{aligned}\delta J(f;\Delta f) \;=\; 2\int_{\nu\in R^n}\Delta\tilde{f}(\nu) & \left(-\frac{1}{l}\sum_{t=1}^{l}(y(t)-f(\psi(t-1)))e^{-j\nu^T\psi(t-1)}\right. \\ & \left.+\lambda\frac{\tilde{f}^\star(\nu)}{\tilde{G}(\nu)}+\gamma\left(\tilde{f}^\star(\nu)-\tilde{f}_a^\star(\nu)\right)\right)d\nu\end{aligned}$$

It follows that a sufficient condition for the Gateaux variation to be zero for all $\Delta f \in \mathcal{F}$ is that $f$ satisfy the Euler-Lagrange equation

$$\frac{1}{l}\sum_{t=1}^{l}(y(t)-f(\psi(t-1)))e^{-j\nu^T\psi(t-1)} \;=\; \lambda\frac{\tilde{f}^\star(\nu)}{\tilde{G}(\nu)}+\gamma\left(\tilde{f}^\star(\nu)-\tilde{f}_a^\star(\nu)\right) \tag{7.7}$$

for all $\nu \in R^n$. Defining $\alpha_t = (y(t)-f(\psi(t-1)))/l\gamma$, and

$$\tilde{K}(\nu) \;=\; \tilde{G}(\nu)\left(\tilde{G}(\nu)+\lambda/\gamma\right)^{-1}$$

we get

$$\tilde{f}(\nu) \;=\; \tilde{K}(\nu)\tilde{f}_a(\nu)+\sum_{t=1}^{l}\alpha_t\tilde{K}(\nu)e^{j\nu^T\psi(t-1)} \tag{7.8}$$

If the set of functions $\{\xi_1^K,\xi_2^K,...,\xi_{d_K}^K\}$ satisfy

$$\frac{1}{\tilde{K}(\nu)}\tilde{\xi}_j^K(\nu) \;=\; 0 \tag{7.9}$$

we may add to the solution $f$ any term in span$\{\xi_1^K,\xi_2^K,...,\xi_{d_K}^K\}$ without violating the Euler-Lagrange equation (7.7). Hence, by transforming back into the original coordinates, we get the representation

$$f(\psi) \;=\; (K*f_a)(\psi)+\sum_{t=1}^{l}\alpha_t K(\psi-\psi(t-1))+\sum_{j=1}^{d_K}\beta_j\xi_j^K(\psi) \tag{7.10}$$

In the next paragraph, we will argue that $\tilde{K}(\nu)$ will always be a low-pass filter. Hence, $K * f_a$ is a smoothed variant of $f_a$. The fundamental differences between the cases $\gamma = 0$ and $\gamma > 0$ is interesting. When $\gamma > 0$, the model (7.10) is the smoothed default model $f_a$ in addition to some smooth terms that compensate for mismatch between the smoothed default model and the observed data. Another major difference is that with $\gamma = 0$, the basis-functions are translations of $G(\psi)$, while they are translations $K(\psi)$ for $\gamma > 0$. As one should expect, $K \rightarrow G$ as $\gamma \rightarrow 0$. We will in the following illustrate this difference with some examples.

Let us look at some typical choices for $\tilde{G}'(\nu)$, and look at the corresponding basis-functions $G(\psi)$ and $K(\psi)$ in the one-dimensional case ($n = 1$). As discussed, we must choose $\tilde{G}'(\nu)$ such that $(\tilde{G}(\nu))^{-1}$ is a symmetric filter that amplifies high-frequency energy.

1. Let us first choose $\tilde{G}'(\nu) = (j\nu)^{-1}$

$$\tilde{G}(\nu) = \frac{1}{\nu^2}, \quad G(\psi) = -\frac{1}{2}|\psi|$$
$$\tilde{K}(\nu) = \frac{\sigma^2}{1+\sigma^2\nu^2}, \quad K(\psi) = \frac{1}{2\sigma}e^{-|\psi|/\sigma}$$

where $K(\psi)$ is characterized by the parameter $\sigma = \sqrt{\lambda/\gamma}$. A large penalty on non-smoothness (large $\lambda$) will give a soft basis-function $K$ that will extrapolate the data points widely, while a small $\lambda$ will give basis-functions that are sharp spikes. It is interesting to observe that while $K(\psi)$ is a local basis-function (a bump centered at the origin), $G(\psi)$ is a global basis-function. The reason for this will be discussed below. A model representation that utilizes the basis-function $G$ is the hinging hyper-plane model structure of Breiman (1993). Also notice that $S$ is the differential operator, and $\tilde{G}(\nu) = -(j\nu)^{-2}$ differentiates twice. Then $d_G = 2$, and the basis $\xi_1^G(\nu) = 1$ and $\xi_2^G(\nu) = \psi$ is a possible choice.

2. Next, we impose infinite penalty on high frequencies

$$\tilde{G}(\nu) = \begin{cases} 0, & |\nu| > \overline{\nu} \\ 1, & |\nu| \leq \overline{\nu} \end{cases}, \quad G(\psi) = \frac{\pi}{\overline{\nu}}\mathrm{sinc}(\overline{\nu}\psi)$$
$$\tilde{K}(\nu) = \begin{cases} 0, & |\nu| > \overline{\nu} \\ \frac{\sigma^2}{1+\sigma^2}, & |\nu| \leq \overline{\nu} \end{cases}, \quad K(\psi) = \frac{\pi\sigma^2}{\overline{\nu}(1+\sigma^2)}\,\mathrm{sinc}(\overline{\nu}\psi)$$

which gives basis-functions $K(\psi)$ and $G(\psi)$ which are equal, except for a factor, and are local oscillating bumps centered at the origin. Notice that (7.5) and (7.9) are satisfied for all functions that have Fourier transform equal to zero for frequencies above $\overline{\nu}$. This is an infinite-dimensional space of band-limited function, and $d_K$ and $d_G$ are infinite. The basis-function $G$ is a reproducing kernel for the Hilbert space of band-limited functions, and such kernels have been suggested as basis-functions in empirical modeling by Mosca (1970).

From the definition of $\tilde{K}(\nu)$ we immediately get the approximation

$$\tilde{K}(\nu) \approx \begin{cases} \tilde{G}(\nu), & \text{if } \left(\tilde{G}(\nu)\right)^{-1} > 1 \\ 1, & \text{if } \left(\tilde{G}(\nu)\right)^{-1} < 1 \end{cases} \tag{7.11}$$

The first example illustrates the case when $\left(\tilde{G}(\nu)\right)^{-1} \to \infty$ as $\nu \to \infty$ and $\left(\tilde{G}(\nu)\right)^{-1} \to 0$ as $\nu \to 0$. While the filter $\tilde{G}(\nu)$ amplifies low frequency energy and suppresses high-frequency energy, $\tilde{K}(\nu)$ will let low frequency energy pass while it suppresses high frequency energy. The amplification of low frequency energy is essentially why the impulse response $G(\psi)$ is a global basis-function, while $K(\psi)$ is a local basis-functions.

In the second example, $\left(\tilde{G}(\nu)\right)^{-1} \to \infty$ as $\nu \to \infty$, while $\left(\tilde{G}(\nu)\right)^{-1} \to 1$ as $\nu \to 0$. This leads to $\tilde{K}(\nu) \approx \tilde{G}(\nu)$ for all $\nu$, and since $\tilde{G}(\nu)$ does not amplify low frequency energy, both $K(\psi)$ and $G(\psi)$ will be local basis-functions. We conclude that it is the low-frequency characteristic of $\left(\tilde{G}(\nu)\right)^{-1}$ that determines whether $G(\psi)$ is local or global. From (7.11) it follows that as long as $\left(\tilde{G}(\nu)\right)^{-1}$ does not amplify low frequency energy, $K(\psi)$ will be a local basis-function. For all sensible choices of $\tilde{G}'(\nu)$, this will be the case. This emphasizes the interpretation that the data-dependent part of the model will compensate for the mismatch between the observed data and the smoothed default model, and this data dependent part will only have influence on the total model locally in a neighborhood of each observation. Intuitively, this makes sense, and a similar hybrid parametric model structure has been suggested on the basis of engineering considerations by Kramer et al. (1992), see also (Thompson and Kramer 1994). The main difference is that the default model is not smoothed, and that the basis functions are chosen in an ad. hoc. manner.

## 7.3 Optimization – Parameterized Approximation

Solving the modeling problem directly in the function space gives a mathematically very elegant solution. However, a simple solution can only be found for rather simple problems not involving complicated constraints, non-linear operators, complicated default model, complicated inner products, or a complicated multidimensional geometry of $\Psi$. For the purpose of solving practical modeling problems we will therefore study solutions based on approximating the optimal model closely with a finite dimensional approximation. With a suitable parameterization of $f$, the problem of directly minimizing (7.2) with respect to these parameters will be computationally feasible in many cases. First, we must choose a rich representation for $f(\psi;\theta)$. This representation can be almost arbitrary since it is of minor importance with the augmented optimization criterion approach. What is required is only that the optimal model can be approximated closely. As we shall see later, the effective number of parameters in this model will be controlled through the choice of $\lambda, \gamma$, and $\beta$. Hence, the number of parameters in the representation of $f$ can be arbitrarily large. Of course, if prior knowledge is available and suggest a particular parameterized structure for $f$, this should indeed be used. On the other hand, it is tempting to choose a linearly parameterized representation $f(\psi;\theta) = \theta^T\varphi(\psi)$, like a look-up table with interpolation. The reason is simply that this will make the criterion quadratic for a significant number of problems. However, in cases when the dimension of $\psi$ is high, such representations may not be well suited because the required number of parameters will typically grow exponentially with the dimension $n$ if the accuracy of the approximation is required to be uniform. This is known as the curse of dimensionality. Hence, non-linearly parameterized representations based on low-dimensional projections, like neural nets, may be more effective in these cases.

With a parameterized model structure it is straightforward to reformulate the infinite-dimensional optimization problem (7.2) into the finite dimensional problem

$$\hat{J}(\theta) = \frac{1}{l}\sum_{t=1}^{l}\varepsilon^2(t;f(\cdot;\theta)) + \lambda||Sf(\cdot;\theta)||^2 + \gamma\mathcal{D}^2(f(\cdot;\theta),M_a) + \beta||Qf(\cdot;\theta)-q||^2$$

subject to

$$Hf(\cdot;\theta) \leq 0, \quad Pf(\cdot;\theta) = 0$$

If all operators operating on $f$ are redefined as operators operating on the parameter vector $\theta$, we get the optimization problem

$$\hat{J}(\theta) = \frac{1}{l}\sum_{t=1}^{l}\varepsilon^2(t;\theta) + \lambda||\hat{S}(\theta)||^2 + \gamma\mathcal{D}^2(f(\cdot;\theta),M_a) + \beta||\hat{Q}(\theta)-q||^2 \qquad (7.12)$$

subject to

$$\hat{H}(\theta) \leq 0, \quad \hat{P}(\theta) = 0$$

Notice that the range of the operators $\hat{H}, \hat{P}, \hat{S}$ and $\hat{Q}$ are still of possibly infinite dimension. A taxonomy for such problems, together with possible solution methods, is

- Quadratic criterion, and no constraints. This gives a set of linear algebraic equations that can easily be solved.
- Quadratic criterion, linear finite-dimensional constraints. Solved using quadratic programming.
- Non-quadratic criterion, no constraints. This is a non-linear programming problem that can be solved iteratively.
- Non-quadratic criterion, finite-dimensional constraints. Constrained non-linear programming problem.
- Infinite-dimensional constraints gives a semi-infinite programming problem.

An in-depth discussion of various practical optimization algorithms that take advantage of the particular structure of the problem is outside the scope of this work. Instead, we refer to the vast literature on general-purpose algorithms, e.g. (Luenberger 1984, Gill et al. 1981, Tanaka, Fukushima and Ibaraki 1988), and software tools like the MATLAB Optimization Toolbox (Grace 1990), and only discuss in depth the simple case when $f$ is linearly parameterized, i.e. $f \in \text{span}\{\varphi_n\}_{n=1}^N$, where $\{\varphi_n\}_{n=1}^N$ is a set of linearly independent smooth basis-functions defined on the domain $\Psi$. Furthermore, we assume there are no constraints, and the default model is an NARMAX model defined by the function $f_a \in L_2(\Psi)$. We use the distance measure $\mathcal{D}^2(f, M_a) = ||f - f_a||^2$. While this is perhaps the simplest possible special case, it contains enough complexity to support a discussion of the various problems involved in solving the general optimization problem.

In this case we can easily compute the solution to the optimization problem explicitly. We define the vectors and matrices

$$
\begin{aligned}
Y = (Y_t),&\quad Y_t = y(t), \quad t = 1, ..., l \\
\Phi = (\Phi_{ti}),&\quad \Phi_{ti} = \varphi_j(\psi(t)), \quad t = 0, ..., l-1, \quad i = 1, ..., N \\
\overline{V} = (\overline{V}_i),&\quad \overline{V}_i =< \varphi_i, f_a >, \quad i = 1, ..., N \\
\overline{W} = (\overline{W}_i),&\quad \overline{W}_i =< \varphi_i, q >, \quad i = 1, ..., N \\
\overline{S} = (\overline{S}_{ij}),&\quad \overline{S}_{ij} =< S\varphi_i, S\varphi_j >, \quad i = 1, ..., N, \quad j = 1, ..., N \\
\overline{R} = (\overline{R}_{ij}),&\quad \overline{R}_{ij} =< \varphi_i, \varphi_j >, \quad i = 1, ..., N, \quad j = 1, ..., N \\
\overline{Q} = (\overline{Q}_{ij}),&\quad \overline{Q}_{ij} =< Q\varphi_i, Q\varphi_j >, \quad i = 1, ..., N, \quad j = 1, ..., N
\end{aligned}
$$

Suppose $S$ and $Q$ are linear operator. Then we can reformulate the criterion as

$$
\begin{aligned}
\hat{J}(\theta) &= \frac{1}{l} < \Phi\theta - Y, \Phi\theta - Y > + \lambda \left\langle \sum_{n=1}^N \theta_n S\varphi_n, \sum_{n=1}^N \theta_n S\varphi_n \right\rangle \\
&\quad + \gamma \left\langle \sum_{n=1}^N \theta_n \varphi_n - f_a, \sum_{n=1}^N \theta_n \varphi_n - f_a \right\rangle \\
&\quad + \beta \left\langle \sum_{n=1}^N \theta_n Q\varphi_n - q, \sum_{n=1}^N \theta_n Q\varphi_n - q \right\rangle \\
&= \frac{1}{l}\theta^T \Phi^T \Phi\theta - 2\frac{1}{l}\theta^T \Phi^T Y + \frac{1}{l} Y^T Y + \lambda\theta^T \overline{S}\theta \\
&\quad + \gamma \left(\theta^T \overline{R}\theta - 2\theta^T \overline{V} + < f_a, f_a >\right) + \beta \left(\theta^T \overline{Q}\theta - 2\theta^T \overline{W} + < q, q >\right)
\end{aligned}
$$

It follows that $\theta = (\theta_1, ..., \theta_N)^T$ is a global minimum of $\hat{J}$, provided

$$
\left(\frac{1}{l}\Phi^T \Phi + \lambda\overline{S} + \gamma\overline{R} + \beta\overline{Q}\right)\theta = \frac{1}{l}\Phi^T Y + \gamma\overline{V} + \beta\overline{W} \tag{7.13}
$$

This set of equations will never be over-determined, so a solution will always exist. Moreover, it will be unique under very reasonable conditions, since all terms on the left-hand side of (7.13) are positive semi-definite, and uniqueness follows if at least one of them is positive definite. One possibility is that $\text{rank}(\Phi^T \Phi) = N$, which is related to the classical condition that one needs at least as many observations as unknown parameters. However, the introduction of prior knowledge allows us to relax this condition and use an over-parameterized approximate representation of the model. For example, a default model will always make the problem well posed. In addition, there should be some restrictions on the choice of basis-functions. These restrictions are related to the fact that the operators $S$ and $Q$ may have null-spaces that may make certain functions "disappear" from the equations that determine the parameters, giving a singular set of equations. A typical example is a differential operator that will map polynomials of sufficiently low order to the zero function.

It is evident that the computational complexity may be a considerable problem even for solving (7.13). The major computational cost involved is computing the inner products. In general, the basis-functions or operators may not always be so nice that this can be done analytically, and numerical integration suffers from the curse of dimensionality. Hence, the computational complexity may grow exponentially with the dimension of the information space. Of course, with non-linear parameterization, operators, or constraints, the problem is more apparent, and the computational complexity will reduce the applicability of the approach.

## 7.4 Tuning the Criterion

In section 7.1 we formulated the semi-empirical modeling and identification problem as an optimization problem, and in sections 7.2 and 7.3 we discussed the solution to this problem. However, it is the choice of criterion that is the most challenging and interesting aspect in this optimization approach, and tuning of the criterion parameters $(\lambda, \gamma, \beta)$ is the topic of this section. Viewed in another way, the problem we discuss in this section is basically to decide on how much we trust the empirical data and the various pieces of prior knowledge relative to each other. Obviously, a lot can be said about this general problem, but we do not have space to give such an in-depth discussion of this problem as it deserves. We will focus on the case when we always require a smooth model, and give precedence to the empirical data relative to the prior knowledge for operating conditions when there are conflicts between the empirical data and the prior knowledge. Moreover, we always rely on the prior knowledge under operating conditions when data is lacking.

If the available empirical data sequence is viewed as the realization of some stochastic process, and a model is identified on the basis of this data, then the expected squared prediction error can be decomposed into a systematic component (bias) and a random component (variance). The systematic component is due to the fact that the prior knowledge and empirical data set can be incorrect or incomplete. The random component is due to unpredictable noise, but most importantly caused by the fact that a finite number of observations is in general not sufficient to identify the best model in the model set. The best model has a complexity that gives the best balance between bias and variance, cf. Fig. 4.1. It is clear that the model complexity should depend on the information contents in the empirical data sequence, and the prior knowledge available. We will later see that the complexity of a model in some cases can be explicitly related to the effective number of parameters in the model, a quantity that depends on $(\lambda, \gamma, \beta)$. Hence, it makes sense to let $(\lambda, \gamma, \beta)$ depend on the empirical data. Assume that we want to find criterion parameters $(\lambda, \gamma, \beta)$ such that the model minimizing $J$ will be the one that has best expected squared prediction error. It is easy too see that the performance of the model will indeed depend on these parameters. For example, a too small $\lambda, \gamma$ or $\beta$ may lead to over-fitting and poor performance when extrapolating, since too much weight will be given to the data. Likewise, a too large $\lambda, \gamma$ or $\beta$ may give too little weight on the empirical data, and a model that is biased under operating conditions where the prior knowledge is incorrect or incomplete.

This suggests a hierarchical optimization approach where the criterion parameters $(\lambda, \gamma, \beta)$ are optimized in an outer loop using a criterion $V(\lambda, \gamma, \beta)$ that reflects the expected squared prediction error on future data, cf. Fig 7.1. The main problem is to find such a criterion $V(\lambda, \gamma, \beta)$. In a stochastic framework, exact computation of this criterion requires the joint probability distribution for all stochastic variables in the model to be known. Clearly, this is unrealistic and we therefore consider some practical alternatives based on empirical data, and mainly assumes ergodicity of the stochastic processes.

Suppose another independent data sequence

$$((u(l+1), y(l+1)), (u(l+2), y(l+2)), ..., (u(l+L), y(l+L)))$$

with $L$ input/output pairs is available. Then an estimate of $V(\lambda, \gamma, \beta)$ is given by

$$V_S(\lambda, \gamma, \beta) = \frac{1}{L} \sum_{t=l+1}^{L} \varepsilon^2(t; \theta(\lambda, \gamma, \beta))$$

where the prediction error $\varepsilon(t; \theta(\lambda, \gamma, \beta))$ is with a parameter vector $\theta(\lambda, \gamma, \beta)$ that solves the problem (7.12). In other words, an independent data sequence is used to test the prediction performance of the identified model, which is reasonable if this data sequence reflects the underlying distribution of the future data. This method is simple, but suffers from the major drawback that a larger amount of empirical data is required.

To overcome this drawback, one can re-use the original data sequence using cross-validation. The cross-validation criterion is an estimate of $V$ given by

$$V_{CV}(\lambda, \gamma, \beta) = \frac{1}{l} \sum_{t=1}^{l} \varepsilon^2(t; \theta^{-t}(\lambda, \gamma, \beta))$$

where $\theta^{-t}(\lambda, \gamma, \beta)$ is the parameter vector that minimizes the criterion (7.12) modified such that there is no weight on the residual at time $t$. This means that $l$ models are fitted by removing one residual at a time from the criterion, and the prediction performance of each model is tested one the single observation that was removed from the criterion. More details can be found in e.g. (Wahba 1990, Stoica et al. 1986).

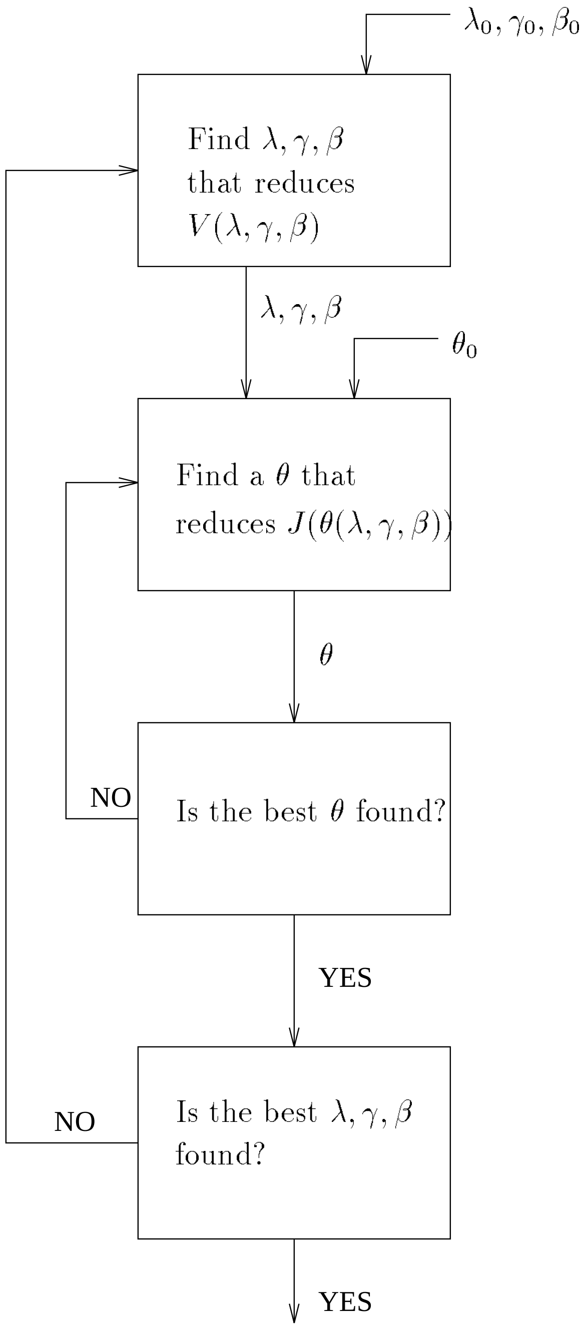


Figure 7.1: Hierarchical optimization approach. The inner loop can be viewed as parameter identification, while the outer loop can be viewed as a kind of structure identification as it estimates the optimal model complexity.

When $\varepsilon(t;\theta)$ depends linearly on $\theta$, a suitable criterion is a modified version of the Final Prediction Error criterion of Akaike (1969):

$$V_{FPE}(\lambda,\gamma,\beta) = \frac{1+d_1(\lambda,\gamma,\beta)/l}{1-d_1(\lambda,\gamma,\beta)/l}\ V_0(\theta(\lambda,\gamma,\beta))$$

where $d_1(\lambda,\gamma,\beta) = \mathrm{trace}(A(\lambda,\gamma,\beta))$ is interpreted as the effective number of parameters in the model (degrees of freedom), and

$$\begin{aligned} A(\lambda,\gamma,\beta) &= \Phi^T\Phi\left(\Phi^T\Phi + \lambda l\overline{S} + \gamma l\overline{R} + \beta l\overline{Q}\right)^{-1} \\ V_0(\theta) &= \frac{1}{l}\sum_{t=1}^{l}\varepsilon^2(t;\theta) \end{aligned} \tag{7.14}$$

When the prediction error depends non-linearly on $\theta$, the FPE criterion for regularized models of Larsen (1994) can be applied

$$V_{FPER}(\lambda,\gamma,\beta) = \frac{1+d_2(\lambda,\gamma,\beta)/l}{1-2d_3(\lambda,\gamma,\beta)/l + d_2(\lambda,\gamma,\beta)/l}\ V_0(\theta(\lambda,\gamma,\beta))$$

where

$$\begin{aligned} d_2(\lambda,\gamma,\beta) &= \mathrm{trace}\left(\mathcal{A}(\lambda,\gamma,\beta)\right) \\ d_3(\lambda,\gamma,\beta) &= \mathrm{trace}\left(\mathcal{A}(\lambda,\gamma,\beta)\mathcal{A}(\lambda,\gamma,\beta)\right) \\ \mathcal{A}(\lambda,\gamma,\beta) &= \frac{\partial^2}{\partial\theta^2}V_0(\theta(\lambda,\gamma,\beta))\left(\frac{\partial^2}{\partial\theta^2}V_0(\theta(\lambda,\gamma,\beta)) + \frac{\partial^2}{\partial\theta^2}\hat{\Omega}(\theta(\lambda,\gamma,\beta))\right)^{-1} \\ \hat{\Omega}(\theta) &= \lambda\|\hat{S}(\theta)\|^2 + \gamma\mathcal{D}^2(f(\cdot;\theta),M_a) + \beta\|\hat{Q}(\theta)-q\|^2 \end{aligned}$$

Next, consider the case when the parameters are constrained. Let the parameter estimate that solves (7.12) be a regular point denoted $\hat{\theta}$, and let $c$ be the number of linearly independent equality and active inequality constraints at $\hat{\theta}$. Let us decompose $\theta$ into a $c$-dimensional sub-vector $\theta_2'$ and an $(N-c)$-dimensional sub-vector $\theta_1'$ such that $\theta = (\theta_1', \theta_2')$ and in a neighborhood of $\hat{\theta}$ there exists a function $g$ such that the constraints are satisfied on the manifold defined by $\theta_2' = g(\theta_1')$. The existence of such a decomposition and function $g$ is guaranteed by the Implicit Function Theorem, e.g. (Luenberger 1984). In other words, in a neighborhood of $\hat{\theta}$, the optimization problem can be reformulated as an unconstrained problem on an $(N-c)$-dimensional subspace of the parameter space. When regularization is not applied, it follows directly that the effective number of parameters equals $N-c$. However, in the general case, one must explicitly know the "implicit function" $g$ in order to substitute $\theta = (\theta_1', g(\theta_1'))$ into $\hat{J}(\theta)$. It is easy to see that a linear approximation to $g$ that is exact in the point $\hat{\theta}$, will be sufficient. Fortunately, such a linearization can usually be found from the equations for the constraints.

Clearly, there will be situations when the automatic procedure for choosing $(\lambda,\gamma,\beta)$ suggested in this section will lead to spurious results. A trivial example is when the given default model is significantly wrong over the whole range of operating conditions where there are empirical data available, but adequate for other operating conditions. Then it is optimal to choose $\gamma = 0$, unless the inner product is chosen to give zero weight on the model near all the data points, in which case $\gamma$ can be arbitrary. Hence, one reason why the algorithm may fail is in general that the future data distribution will be different from the distribution of the empirical data used for identification. As discussed before, the automatic choice of $\lambda, \gamma$, and $\beta$ is based on the assumption that the empirical data are more reliable than the prior knowledge. Such an assumption must be made in order to resolve conflicts between the different pieces of knowledge and data, and is controversial both in the philosophy of science, and in engineering practise. Still, even if the algorithm may fail when searching for $\gamma$ or $\beta$, the choice of $\lambda$ may still be sensible, since smoothness may be a uniform property of the system, i.e. uniformly true for all operating conditions, while the default model and constraints may be incorrect under some operating conditions. In summary, we do not recommend a completely automated choice of tuning parameters, in particularly not $\gamma$ and $\beta$.

The estimated relative weights on the different terms in the criterion may be interpreted as indicators on the relative correctness or relevance of the corresponding prior knowledge or empirical data. As we have discussed above, such interpretations should be made carefully, as it is possible that knowledge that is incorrect or irrelevant (or both) for all the operating conditions we have empirical data available for, will be adequate for other operating conditions. One must always have the information contents in the data set in mind. However, if there are large amounts of data available, and the data covers all operating conditions, such interpretations may be valuable.

Finally, it should be mentioned that it is possible is include the order parameters $n_y$, $n_u$, and $n_e$ together with $\lambda, \gamma$, and $\beta$ in the search.

## 7.5 Simulation Example: pH-neutralization

The purpose of this simulation example is to illustrate the effect of imprecise, incomplete, and incorrect prior knowledge and empirical data on the identified model, and how this framework can handle the different kinds of prior knowledge and empirical data.

The system we will study is a pH neutralization tank. The system has one input (the base flow rate $u(t)$), one output (the measured pH $y(t)$), and no disturbances or noise. There are three influent streams and one effluent stream:

- Influent acid stream $Q_1$ ($HNO_3$)
- Influent buffer stream $Q_2$ ($NaHCO_3$)
- Influent base stream $Q_3$ (NaOH and traces of $NaHCO_3$)
- Effluent stream $Q_4$

We use a model developed by Hall and Seborg (1989) to simulate the "true system", see section 6.4.2.

The available prior knowledge and empirical data is supposed to be as follows:

1. **Data sequence** from a limited region of operation, cf. Fig 7.2a. The sampling interval is 15 s, and the sequence contains 500 observations.

2. **Steady-state data** from 5 operating points, pH $\in$ { 5.03, 5.78, 6.31, 7.03, 9.15 }.

3. **Simplified mass-balance model** where the buffer stream is neglected, cf. section 6.4.2.

4. **pH range**, in other words the fact that $0 \leq \text{pH} \leq 14$.

5. **Smoothness** of the system behavior.

6. **Open-loop stability** of the system.

We will study the problem of identifying a model of the NARX form

$$y(t) = f(y(t-1), u(t-1)) + e(t) \tag{7.15}$$

The interesting operating range is defined by $0 \leq y \leq 14$ and $0 \leq u \leq 40$ ml/s. Hence, $\Psi = [0, 14] \times [0, 40]$. Notice that the available data sequence only covers a small part of $\Psi$, cf. Fig. 7.3a, while the model will be partially validated using a data sequence that covers a much larger part of $\Psi$, cf. Figs. 7.2b and 7.3b. The unknown non-linear function $f$ is represented by a 2-dimensional lookup-table $(\theta_{ij})$ with linear B-spline interpolation on this domain. The choice of representation is of minor importance, and the lookup-table is chosen because of its simplicity. The 390 entries in the lookup-table are considered as unknown parameters.

We choose a non-smoothness penalty of the form

$$||Sf||^2 = \theta^T \overline{S} \theta \approx \int_{\psi \in \Psi} \left|\left| \frac{\partial^2}{\partial \psi^2} f(\psi) \right|\right|^2 d\psi$$

which is based on an easy-to-compute finite-difference approximation of the Hessian

$$\frac{\partial^2 f(\psi_{ij})}{\partial \psi^2} \approx \begin{pmatrix} \frac{\theta_{i+1,j} - 2\theta_{i,j} + \theta_{i-1,j}}{\Delta_1^2} & \frac{\theta_{i+1,j+1} + \theta_{i,j} - \theta_{i+1,j} - \theta_{i,j+1}}{\Delta_1 \Delta_2} \\ \frac{\theta_{i-1,j-1} + \theta_{i,j} - \theta_{i-1,j} - \theta_{i,j-1}}{\Delta_1 \Delta_2} & \frac{\theta_{i,j+1} - 2\theta_{i,j} + \theta_{i,j-1}}{\Delta_2^2} \end{pmatrix}$$

where $f(\psi_{ij}) = \theta_{ij}$. Furthermore, $\Delta_1 = 1.0$ and $\Delta_2 = 1.6$ are the uniform discretization intervals in the lookup-table. One motivation behind this choice is that penalty on the Hessian can be interpreted as penalty on deviations from a linear model structure. This is reasonable, since a linear model structure is perhaps the simplest possible that may be adequate. Another motivation is that small curvature intuitively is equivalent to smoothness.

In the following we will identify five models based on different combinations of the available knowledge and data, and as an alternative model structure we try a neural network for comparison.

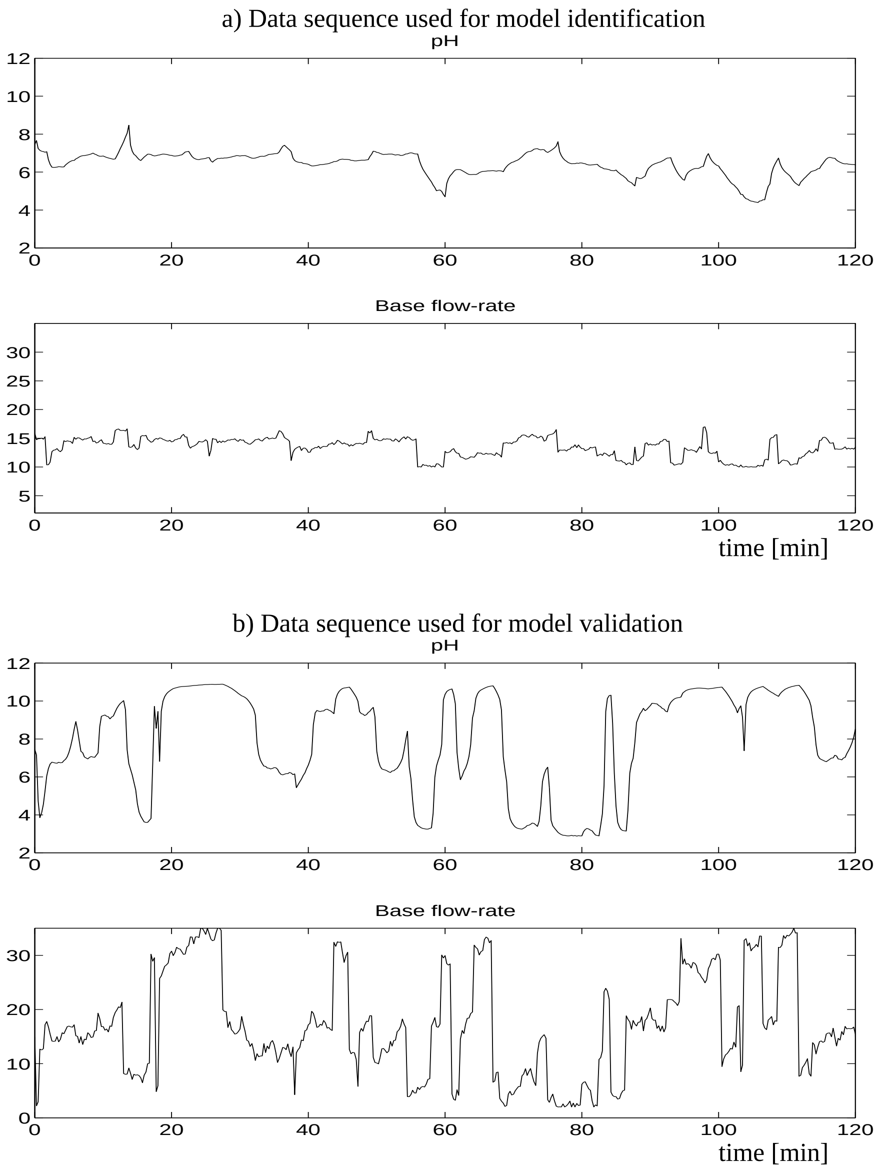


Figure 7.2: Simulated data sequences.

## Model 1. Knowledge: Data sequence, Smoothness, Stability

Open loop stability is represented as a constraint in the problem

$$\hat{J}(\theta) = \frac{1}{l}\sum_{t=1}^{l}(y(t) - \psi(t-1)^T\theta)^2 + \lambda\theta^T\overline{S}\theta$$

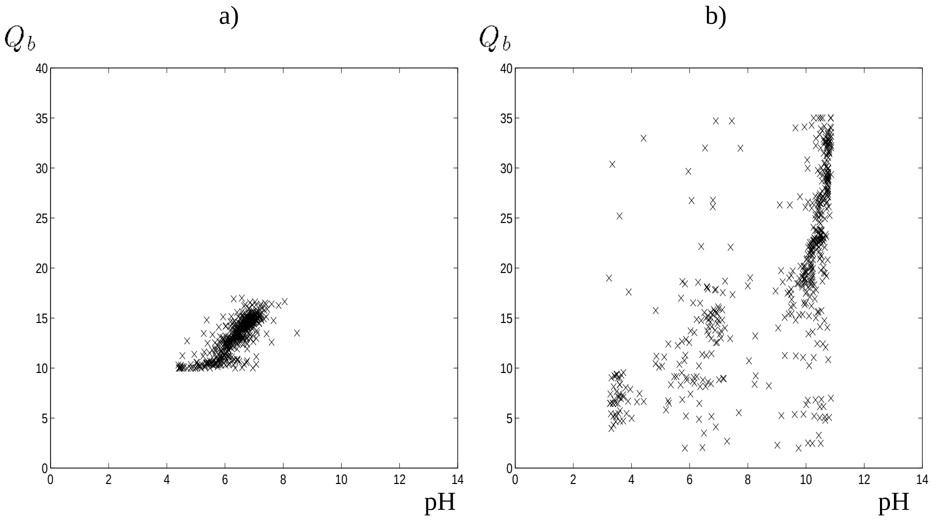


Figure 7.3: Empirical distribution of a) identification data, and b) validation data.

subject to

$$|\theta_{i,j} - \theta_{i-1,j}| \leq \Delta_1, \qquad \text{for all} \;\; i, j$$

This optimization problem is solved using quadratic programming (MATLAB function **qp**). Notice that the use of a lookup-table with linear interpolation allows us to reduce the stability constraint to a finite number of constraints on the parameters. The generalized FPE criterion with the effective number of parameters corrected for the active constraints, is minimized using a line search algorithm (MATLAB function **fmin**) to find the smoothing parameter $\lambda = 3.8 \cdot 10^{-4}$ corresponding to $d_1 = 17.1$ effective parameters. Notice that the prior knowledge about open loop stability is of great importance, since otherwise the model will be unstable for a wide range of $\lambda$ values.

## Model 2. Knowledge: Data sequence, pH range, Smoothness

Notice that bounded pH range automatically implies BIBO-stability of the model. Here we solve the minimization problem

$$\hat{J}(\theta) = \frac{1}{l}\sum_{t=1}^{l}(y(t) - \psi(t-1)^T\theta)^2 + \lambda\theta^T\overline{S}\theta$$

subject to

$$0 \leq \theta_{ij} \leq 14, \qquad \text{for all} \;\; i, j$$

This quadratic programming problem is finite-dimensional, and solved using the MATLAB function **qp**. Similar to Case 1, a minimization of the generalized FPE criterion gives $\lambda = 5.6 \cdot 10^{-5}$, which gives $d_1 = 21.7$ effective parameters.

## Model 3. Knowledge: Simplified mass-balance

No identification is needed, cf. the model in section 6.4.2. Notice that this model contains very limited knowledge about the chemistry in the tank, since the buffer stream is neglected. Hence, the prior knowledge assumed is significantly less than the "true system", and the model is considerably simpler.

## Model 4. Knowledge: Data sequence, Simplified mass-balance, Smoothness

The default model is stable, so the stability constraint is not required. The criterion is now

$$\hat{J}(\theta) = \frac{1}{l}\sum_{t=1}^{l}(y(t) - \psi(t-1)^T\theta)^2 + \lambda\theta^T\overline{S}\theta + \gamma D(f(\cdot;\theta), M_a)$$

where

$$D(f(\cdot;\theta), M_a) = \frac{1}{1000}\sum_{t=1}^{1000}(y_a(t) - \psi_a^T(t-1)\theta)^2\rho(t)$$

we have defined $\psi_a(t) = (y_a(t), u_a(t))^T$, and the data sequence

$$((u_a(1), y_a(1)), (u_a(2), y_a(2)), ..., (u_a(1000), y_a(1000)))$$

is generated by simulating the model $M_a$. The input sequence is chosen to cover as much of the operating range of the process as possible. The weighting function is defined as $\rho(t) = 1$. In this case a search for $\lambda$ and $\gamma$ using the generalized FPE criterion fails. The reason is simply that the default model is significantly wrong for all the operating conditions captured by the sequence of empirical data available. Hence, the optimization gives $\gamma \approx 0$. A more intelligent choice of weighting function $\rho(t)$ that applies the prior knowledge that the default model is poor for intermediate pH values, does not resolve the problem. Instead, we have chosen $\lambda = 10^{-3}$ and $\gamma = 0.05$, which gives $d_1 = 13.5$. This choice gives a smooth transition to the default model outside the range there are data available from.

## Model 5. Knowledge: Steady-state data, Simplified mass-balance, Smoothness

Here we solve the minimization problem

$$\hat{J}(\theta) = \lambda\theta^T\overline{S}\theta + \gamma D(f(\cdot;\theta), M_a) + \beta\sum_{i=1}^{5}(y_{\text{steady-state}}(u_i) - y_i)^2$$

where $D(f(\cdot;\theta), M_a)$ is the same as above. Since there is data sequence available, we make the somewhat arbitrary choice $\lambda = 0.005, \gamma = 0.05, \beta = 0.05$. The steady-state model output is computed by simulating the model until a steady state is reached. The model's steady-state output is a non-linear function of the steady-state input, and the steady-state data are represented as five soft constraints. This optimization problem is solved using the MATLAB function **fminu**.

## Model 6 and 7. Sigmoidal Neural Network

For comparison, we have also identified a feed-forward neural network model of the form (7.15). Compared to the optimization approach, this corresponds to the a priori choice of a particular parameterized model structure. The parameters are identified using the MATLAB Neural Network Toolbox (Demuth and Beale 1993), i.e. the back-propagation algorithm with momentum term using the identification data sequence and a least squares criterion. Depending on the number of hidden nodes, the identification algorithm tuning-parameters, and initial parameter estimates, the results varies considerably. We have chosen to present two identified models with the same structure (10 hidden nodes with sigmoidal non-linearities, corresponding to 41 parameters), the same identification algorithm tuning-parameters, but different (small random) initial parameter estimates.

## Results

The prediction performance of the models is illustrated by simulations using the validation data input sequence in Fig. 7.4. Moreover, the major non-linearity in this process is the steady-state response, or titration curve, which is shown for the different models in Fig. 7.5.

These curves clearly show that different aspects of the system behavior are captured, depending on the prior knowledge applied. The performance of Model 1 is poor at at high and low pH values, since no empirical data and only weak prior knowledge like smoothness and stability is relevant under these conditions. Restricting the pH range gives some improvement, cf. Model 2. Likewise, the simplified mass balance that makes up Model 3 gives reasonable predictions only at high and low pH values. Using this model as a default model combined with the empirical data sequence from the intermediate pH range gives a model with reasonable predictions over the full pH range, cf. Model 4. It is also interesting to observe that the default model combined with a small amount of steady-state data from the intermediate pH range, but no dynamic data, does not give poorer prediction, cf. Model 5.

A comparison with the neural network models (Models 6 and 7) is also quite interesting. The two neural network models perform equally well on predicting the identification data, but perform considerably different on the validation data and have quite different titration curves. It should be mentioned that these differences are typical, and more pronounced differences (on the validation data) appear when the model structure and identification algorithm tuning-parameters varies as well. Notice that it is most fair to compare Models 6 and 7 with Model 1, since they are both based on weak prior knowledge.

It seems reasonable to conclude that this example illustrates that the incorporation of increasing amounts of prior knowledge may improve the quality of the identified model. Furthermore, the optimization approach seems more adequate than the application of a black-box model structure like a neural network in this application, since the empricial data does not cover the operating range sufficiently. This

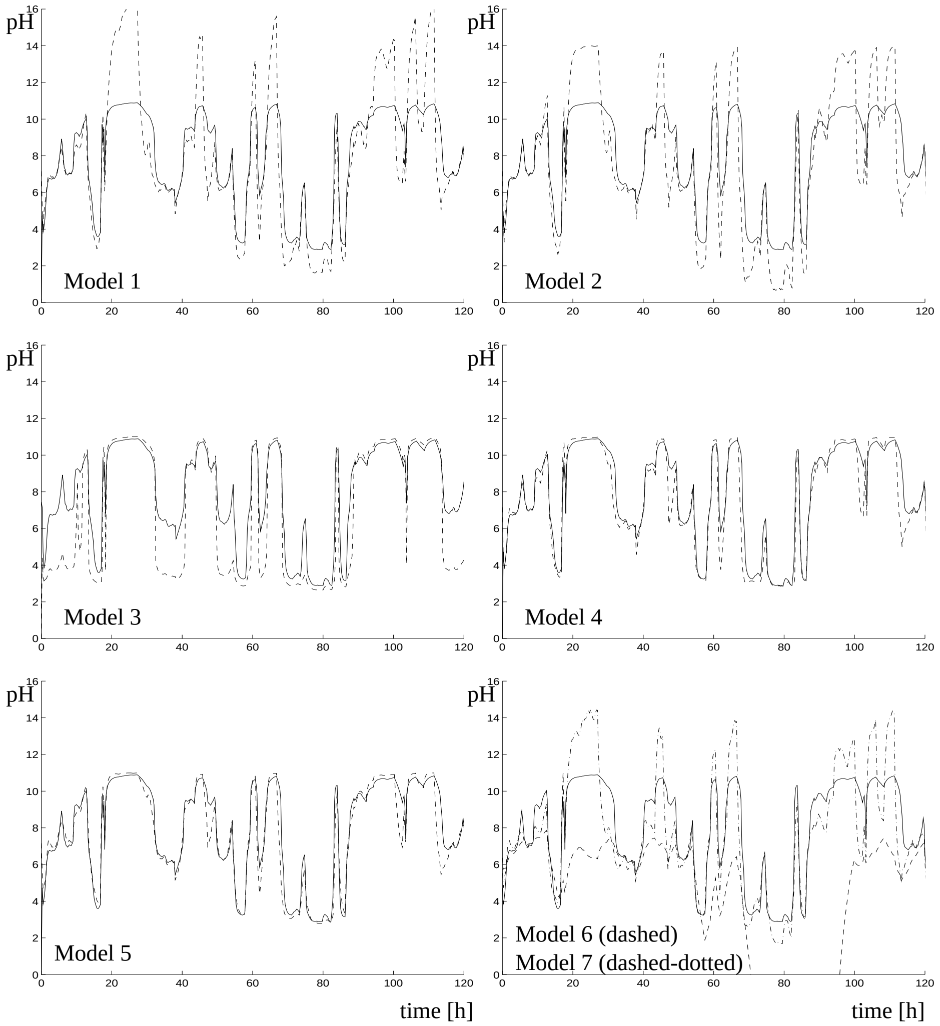


Figure 7.4: Simulation of the 7 models when the validation data input sequence is applied to the models. The solid curves are the validation data pH, while the dashed curves are simulated pH using the models. For models 6 and 7, the dashed and dashed-dotted curves are with different initial parameter estimates.

supports our initial claim that the choice of model structure in empirical modeling is a difficult and critical problem.

This example also shows that the automatic choice of criterion tuning parameters $(\lambda, \gamma, \beta)$ is feasible only if the available data covers the interesting operating conditions sufficiently well. In particular, the algorithm gives reasonable choices for Models 1 and 2, but failes for Model 4.

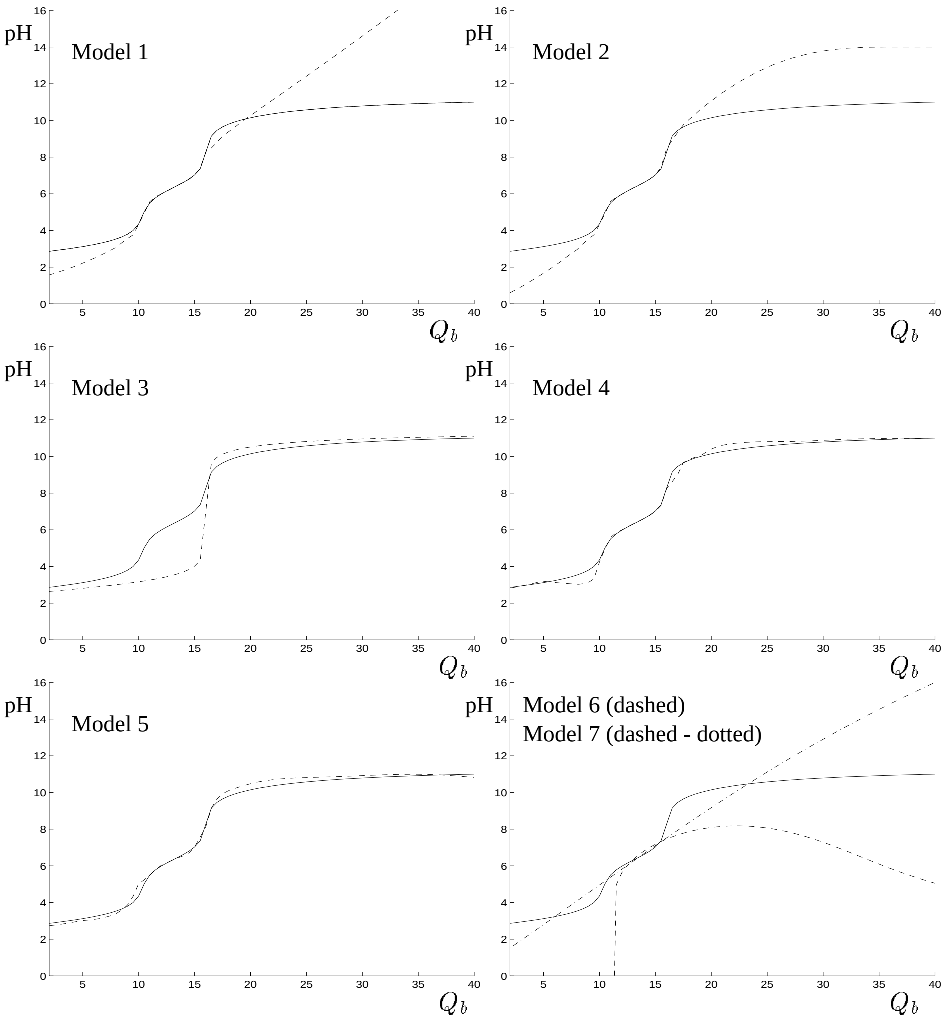


Figure 7.5: Titration curves (steady state response) of the 7 models. The solid curves are the system's true titration curve, and the dashed curves are the models. For models 6 and 7, the dashed and dashed-dotted curves are with different initial parameter estimates.

## 7.6 Discussion

We have formulated the semi-empirical modeling problem as an optimization problem on a function space. In addition to the standard penalty on mismatch between model prediction and the empirical data, the criterion also penalizes mismatch between the model and various pieces of prior knowledge. Furthermore, prior knowledge can enter as constraints in the optimization problem. We have solved a simplified optimization problem in the function space, and discussed finite-dimensional approximation to the general solution. The approach is interpreted in the context of regularization, and it is shown how different pieces of prior knowledge will reduce the effective number of parameters in the model. The problem of weighting the various pieces of knowledge and data is approached similarly to the problem of choosing a regularization parameter, and some criteria are suggested.

Our work provides a unified framework for incorporation of prior knowledge in non-linear system identification. It is primarily based on (Thompson and Kramer 1994, Girosi et al. 1994). In particular, the framework provides an attractive alternative to the direct specification of a parametric model structure, since the formulation of an optimization criterion is clearly a more high-level and transparent procedure that requires less guesswork.

However, there are at least two major drawbacks with this approach:

1. The computational complexity may limit its applicability. This problem will in particular be apparent when the dimension of the information space is high, or there are non-linear operators or parameterizations involved. The problem may be reduced with the aid of tailored optimization algorithms, solving the problem analytically as far as possible, and trying to simplify the various elements involved. A library of flexible and computationally effective building blocks for representing various pieces of prior knowledge and information would be useful.
2. The resulting model need not be transparent, in the sense that it cannot easily be interpreted and analyzed, since it is non-parametric. The only possibility is to interpret the model in terms of the prior knowledge and empirical data applied. However, this may be of limited help for many applications. We would like to stress that all black-box model structures suffer from the same drawback, perhaps to an even larger extent, since the prior knowledge is often applied in a less transparent manner.

The major advantage of the approach is its flexibility with respect to incorporation of prior knowledge in various forms. This flexibility can possibly be extended and improved in several directions:

1. The extension to the MIMO-case is conceptually straightforward.
2. We believe that extensions to other model representations than NARMAX models (like state-space models) is possible.
3. A more accurate model (and better interpretation of estimated criterion tuning-parameters) may be found if the terms in the criterion are split into different terms that are valid under different operating conditions.
4. As mentioned before, a library of building blocks would be useful. Such a library might contain various model space metrics, non-smoothness operators, operators for representing steady-state data, stability etc.
5. More work is needed to develop practical engineering procedures for the coding of prior knowledge into the criterion and constraints. One can imagine a high-level expert system based user interface that automatically codes prior knowledge like "We have good confidence in the default model when the pH is either high or low, but no confidence in it for intermediate pH values" into the optimization criterion. This elevates the modeling problem to an even higher level where the optimization formulation is hidden for the user.

It is commonly stated that the main benefit of "modern" non-linear empirical modeling approaches like radial basis-functions, wavelets, and neural networks is that essentially no prior knowledge is required. This is to some extent true, but one must remember that the price to pay may be uncertainty about what set of basis-functions to apply, and a more or less ill-conditioned identification problem, which will typically lead to poor extrapolation capabilities of the model, cf. the simulation example in section 7.5. Unfortunately, it has also been common practise that the lack of transparency of such black-box approaches has led to neglection of the available prior knowledge, and the resulting ill-conditioning has been handled implicitly using regularization methods like stopped training (Sjöberg and Ljung 1992, Ljung, Sjöberg and McKelvey 1993). However, simple explicit regularization methods like default models (Thompson and Kramer 1994, Su et al. 1992, Kramer et al. 1992, Johansen and Foss 1992*c*), constraints (Joerding and Meador 1991, Thompson and Kramer 1994), penalty on parameter magnitude (Weigend, Huberman and Rumelhart 1990), and smoothness regularization (Bishop 1991, Girosi et al. 1994), has also been suggested. The optimization framework presented here will be useful for regularizing such complex parameter identification problems, and is useful as a complementary technique applied together with other approaches such as neural networks, in order to reduce the sensitivity with respect to the a priori choice of model structure.

# Chapter 8

# Operating Regime based Model Predictive Control

A major application area for non-linear dynamic models is dynamic optimization, or model predictive control. In this chapter we will present a semi-realistic simulation example, where the operating regime based modeling framework is used to develop a model that is used in a model predictive controller that is applied to a batch fermentation reactor. The performance with this model is compared to two other models.

The continuation of this chapter is structured as follows: First, we present a brief survey of the batch process control and dynamic optimization literature, with emphasis on the intersection of these fields. Second, the predictive control problem is defined. This is, essentially, nonlinear state-space MPC utilizing the operating regime based modeling framework. The method is applied to a simulated batch fermentation process. Model development is emphasized. MPC based on local modeling is compared to the use of a conventional nonlinear state-space model and a linear model. Some discussion finalize the chapter, which is based on (Foss, Johansen and Sørensen 1994).

## 8.1 Batch Processes and Dynamic Optimization

Dynamic optimization has for decades been a basis for control. During the 60's and 70's, closed-form solutions to the optimization problem was the driving force for much research activity. The linear-quadratic controller seemed to provide a solution to many multi-variable control problems. Experience has shown that this approach has serious shortcomings: First, the solution assumes no constraints on the states and control inputs. Second, the linear-quadratic controller need not be robust when estimator dynamics are included for state estimation, as shown by Doyle (1978).

Within the domain of optimization-based process control research, the interest and successful industrial applications have the last decade focused on model-based predictive control (MPC); see (Garcia, Prett and Morari 1989) and (Rawlings, Meadows and Muske 1994) for comprehensive surveys of this field. The idea is to solve the optimization problem at a given time instant by utilizing the most recent process measurements. The optimization problem is defined on some horizon, and a control trajectory is computed on this horizon. Only the first part of the control trajectory is applied to the process, and the entire optimization is repeated at the next sampling instant, again utilizing process measurements up until this new time instant. This methodology was first presented by Cutler and Ramaker (1979), and they minimize a quadratic criterion weighting the control errors and changes in the control inputs. They use a linear moving-average model for prediction. More important, this optimization-based controller can handle constraints in both the control inputs and the controlled variables.

The large majority of MPC is based on linear models. An important reason for this is that linear MPC, like other linear controllers, can handle processes with weak non-linearities. Although the use of MPC within the process industries has been extensive, it has been mainly limited to continuous processes. Such processes are often characterized by small variations in operating conditions. This is not the case for batch and fed-batch processes which are also widely used in industry. The difference between batch and fed-batch processes is that no feed is added and no product is removed in a batch process, while this is done in a fed-batch process. Batch processes will typically exhibit large variations in the operating conditions during a batch. Moreover, the product specifications may differ among batches, thereby changing the operating conditions significantly between batches. Johnson (1987) compares control of continuous processes and batch-type processes by stating that the optimization problem of batch processes is a dynamic problem involving highly nonlinear process models. As a contrast, continuous processes can often be optimized by a static formulation. Control of a batch reactor is usually carried out by a two step procedure. Time-varying trajectories for the important variables are first derived. This is either done in a heuristic manner based on process insight and experience from earlier operation, or by open-loop optimization based on a model of the batch reactor. Second, the tracking of the variables is accomplished by set-point controllers. Reviewing the batch reactor control literature, emphasizing fermentation reactors, shows that four questions are focused; the generation of optimal trajectories, controller design for setpoint control, computing on-line estimates of reactor states, and the issue of reactor modeling, (Rippin 1989), (Johnson 1987) and (Jörgensen and Jensen 1989).

The work on optimal trajectories is usually based on some non-linear mechanistic model of the process in question. The cost criterion typically include productivity and input costs, and the optimization problem is solved off-line. Examples of this can be found in (Impe, Nicola, Vanrolleghem, Spriet, Moor and Vandewalle 1994) and (Sargantanis, Karim, Murphy and Ryoo 1993). There are, however, also examples on the use of on-line optimization, i.e. use of MPC, (Lim and Lee 1991).

There is a quite extensive literature on setpoint control of batch-type reactors, in particular the use of adaptive control and feedback linearization. Bastin and

Dochain (1988) and Pomerleau, Perrier and Dochain (1989) base their controller design on mechanistic models, while recent work by Proll and Karim (1994) and Keulers (1993) use empirical models for nonlinear control.

A major problem when implementing advanced control in biotechnical processes is the lack of good measurements. Hence, research has also focused on-line estimation of reactor states, particularly substrate and product concentrations. Examples of this is found in (Hengjie, Jianzhong, Shuqing and Jicheng 1989, Hilaly, Karim and Linden 1994, Keulers 1993).

Reactor modeling for the purpose of control, span from linear and nonlinear mechanistic models to linear and nonlinear empirical models. Of particular interest to our work is (Zhang et al. 1994). They utilize the fact that different phenomena dominate during different parts of a batch cycle, and construct a set of local models that that are valid during different parts of a batch cycle. In addition, they specify a method to select the appropriate model at a given time. The advantage of this concept is that the individual local models are simpler than a global model that can represent the whole batch cycle. A similar approach is proposed by Konstantinov and Yoshida (1989).

This work investigates the use of MPC on batch processes using a non-linear model in the controller. There are some reports on this in the literature. Lim and Lee (1991) describe the use of MPC using on-line parameter estimation. The control trajectory is computed by simultaneous parameter estimation and re-optimization. Garcia (1984) extends the method introduced by Cutler and Ramaker (1979) by using a nonlinear model for output prediction. This controller is tested on a polymerization reactor model. A similar approach is presented by Peterson, Hernandez, Arkun and Schork (1989).

## 8.2 Model Predictive Control

The present work rests on two assumptions. The first assumption is that the performance of MPC depends critically on the predictive capabilities of the underlying process model. The wide operating range of a batch makes the use of a non-linear prediction model particularly interesting. The second assumption is that nonlinear model building is a cumbersome task. Hence, semi-empirical modeling techniques are interesting. This is also motivated by the observation that practically all predictive control loops implemented in industry are based on empirical models.

In this chapter we apply a state-space formulation of the model, cf. Chapter 3. This leads to a nonlinear state-space MPC problem, cf. (Balchen, Ljungquist and Strand 1992). For batch and semi-batch processes, this nonlinear state-space MPC problem can be formulated as

$$\max_{u \in U} \left( m(x(T_s)) + \int_t^{T_s} l(x(\tau), y(\tau), u(\tau)) d\tau \right)$$

subject to

$$\begin{aligned} \dot{x} &= f(x,u,v), \quad x(t) \text{ given} \\ y &= g(x,w) \\ h(x,y,u) &\leq 0 \end{aligned}$$

where typically $T_s = \min(t+T, T_f)$. The optimization is defined on some horizon $T$, starting at the present time $t$. Time $t = 0$ defines the start of a batch and $t = T_f$ defines the end of a batch. The end time $T_f$ need not be fixed, and this variable is often optimized, too.

Both equality and inequality constraints can be defined. Soft constraints may be defined as an integrated part of the optimization criterion. Measurements are explicitly mentioned in the formulation to emphasize the fact that it is sometimes natural to optimize with respect to these variables.

To reduce the complexity of the optimization problem, the set of possible control input trajectories $U$ is restricted to a finite-dimensional space. The control input is here parameterized as a piecewise constant function:

$$u(\tau) = \begin{cases} \mu_1, & \tau \in [t, t+\Delta T) \\ \mu_2, & \tau \in [t+\Delta T, t+2\Delta T) \\ \vdots & \end{cases}$$

where $\Delta T$ is the sampling interval. The optimization problem is solved by the use of a non-linear programming algorithm at time instants $t \in \{0, \Delta T, 2\Delta T, ..., T_f - \Delta T\}$ using the most recent process measurements. Only, the first part of the optimal trajectory, $\mu_1$, is applied as the control input.

A major problem with the above formulation is its dependence on the initial states $x(t)$. In practice, these are not readily available. Hence, some estimate of the states must be computed. This may be accomplished by state estimation or an observer.

Since most batch-type processes are highly non-linear, there are two potential advantages in applying nonlinear MPC for batch processes, compared to linear MPC. First, the predictive capability on the optimization horizon may improve by utilizing a nonlinear as opposed to a linear model. Second, the states $x(t)$ may be estimated with improved accuracy by the use of a nonlinear model.

## 8.3 Simulation Example: Batch Fermentation Reactor

A semi-realistic simulation study of a batch fermentation process illustrates the ideas presented in this chapter. In this study, five controllers are formulated, based on the above formulation. All controllers utilize the same performance criterion and constraints, equal control input parameterization, and identical optimization algorithms. The controllers differ in the following way:

1. The 1st MPC uses an ideal process model, i.e. the model and the "true system" are identical. Provided the initial values $x(0)$ are correct, this controller gives an upper limit to the performance of MPC.
2. The 2nd MPC uses a nonlinear operating regime based state-space model for both prediction and state estimation.
3. The 3rd MPC uses a global linear state-space model for both prediction and state estimation.
4. The 4th controller is an open-loop optimal controller (OLOC)using the same non-linear model as the 2nd MPC.
5. The 5th controller is also an open-loop optimal controller (OLOC), using the global linear model for prediction.

### 8.3.1 System Description

The simulated "true system" model is adapted from (Ghose and Ghosh 1976) and (Rai and Constantindes 1973), and describes the fermentation of glucose to gluconic acid by the micro-organism *Pseudomonas ovalis* in a well-stirred batch reactor. The main overall reaction mechanism can be described by

$$\begin{aligned}
\text{Cells} + \text{Glucose} + O_2 &\rightarrow \text{More Cells} \\
\text{Glucose} + O_2 &\overset{\text{Cells}}{\rightarrow} \text{Gluconolactone} \\
\text{Gluconolactone} + H_2O &\rightarrow \text{Gluconic Acid}
\end{aligned}$$

The first reaction is the reproduction of cells, using the substrate glucose and oxygen. The second reaction is the production of gluconolactone, again using glucose and oxygen. This reaction is enzyme-catalyzed by the cells, while the final product, gluconic acid, is formed by the last reaction. The following state-space model is used to simulate the "true system":

$$\begin{aligned}
\dot{x}_1 &= \mu_m \frac{x_1 x_4 x_5}{K_s x_5 + K_0 x_4 + x_4 x_5} \\
\dot{x}_2 &= v_L \frac{x_1 x_4}{K_L + x_4} - 0.9082 K_p x_2 \\
\dot{x}_3 &= K_p x_2 \\
\dot{x}_4 &= -\frac{1}{Y_s} \mu_m \frac{x_1 x_4 x_5}{K_s x_5 + K_0 x_4 + x_4 x_5} - 1.011 v_L \frac{x_1 x_4}{K_L + x_4} \\
\dot{x}_5 &= k_l a (x_5^* - x_5) - 0.09 v_L \frac{x_1 x_4}{K_L + x_4} - \frac{1}{Y_0} \mu_m \frac{x_1 x_4 x_5}{K_s x_5 + K_0 x_4 + x_4 x_5}
\end{aligned}$$

where $x_1$ is the cell concentration, $x_2$ is gluconolactone concentration, $x_3$ is gluconic acid concentration, $x_4$ is glucose concentration and $x_5$ is dissolved oxygen concentration. The parameters $\mu_m$, $K_L$, $v_L$, and $K_p$ depend on temperature and pH. This dependency is given by an interpolated lookup table based on the experimental data in (Rai and Constantindes 1973). The remaining parameters can be found in (Rai and Constantindes 1973) and (Ghose and Ghosh 1976). Initial values for the batch are $x_1(0) = x_{10}, x_2(0) = 0, x_3(0) = 0, x_4(0) = x_{40}$, and $x_5(0) = x_5^\star$.

The setpoints to the temperature and pH basis-control loops are used as control inputs by the predictive controller. The basis-control loops are assumed to be perfect, which is realistic, since the system dynamics are slow compared to the typical bandwidth of these loops.

Three perfect on-line measurements are available at 0.5 $h$ intervals during the batch: Dissolved oxygen concentration, biomass concentration and gluconic acid concentration. There are no noise or disturbances in the simulations.

### 8.3.2 Modeling and Identification

All the local models are chosen to have the same linear structure

$$x(t+1) = a_i + A_i x(t) + B_i u(t) + v(t) \tag{8.1}$$

where $x = (x_1, ..., x_5)$, $u = (\text{pH}, \text{temp})$, $a_i$ is a vector of unknown parameters, $B_i$ is a $5 \times 2$-matrix of unknown parameters, and $A_i$ has the structure

$$A_i = \begin{pmatrix} A_{11}^i & 0 & 0 & A_{14}^i & A_{15}^i \\ A_{21}^i & A_{22}^i & 0 & A_{24}^i & 0 \\ 0 & A_{32}^i & 1 & 0 & 0 \\ A_{41}^i & 0 & 0 & A_{44}^i & A_{45}^i \\ A_{51}^i & 0 & 0 & A_{54}^i & A_{55}^i \end{pmatrix}$$

The structural zeros follow from a simple mass-balance based on the reaction mechanism and the assumption that the reaction rates only depends on $x_4$ and $x_5$, in addition to $u$, which is a quite natural assumption to make, since these are the rate-limiting components.

By examining the main reaction mechanisms, four operating regimes can be identified, see also section 3.3.2. At the beginning of the batch, the production of gluconolactone is small due to the small concentration of cells. Hence, the production of gluconic acid is small due to the low concentration of gluconolactone. This regime is characterized by a relatively high concentration of both dissolved oxygen and glucose. In the intermediate stages of the batch, the production of cells and gluconolactone proceeds at a high rate, and some gluconic acid is produced. There is a relatively low concentration of dissolved oxygen, and the concentration of glucose is decreasing. Depending on whether the dissolved oxygen concentration is so low that the transfer of oxygen to the cells is rate-limiting or not, the dynamic behavior of the process is different. This gives two regimes that are characterized by a medium concentration of glucose, and either low or medium concentration of oxygen. During the final stages of the batch, the production of cells and gluconolactone is reduced due to shortage of glucose. The only significant reaction is the production of gluconic acid from gluconolactone. This regime is characterized by low substrate concentration, and high dissolved oxygen concentration.

These four regimes can all be characterized by the concentration of dissolved oxygen and glucose, and these two variables are chosen to be the variables that defines the operating point. The four regimes were chosen on the basis of the discussion above, and their interpolation functions are shown in Fig. 8.1. Since the dependencies on pH and temperature are highly nonlinear, the local model within each of these four regimes should therefore depend non-linearly on temperature and pH. The chosen local model structure (8.1) does not, so each of these four regimes is therefore further decomposed into four new regimes along the temperature and pH axes, as shown in Fig. 8.2. Hence, the model is based on a total of 16 local models. The model validity functions $\rho_i$ were chosen to be Gaussian functions, with some suitable overlap.

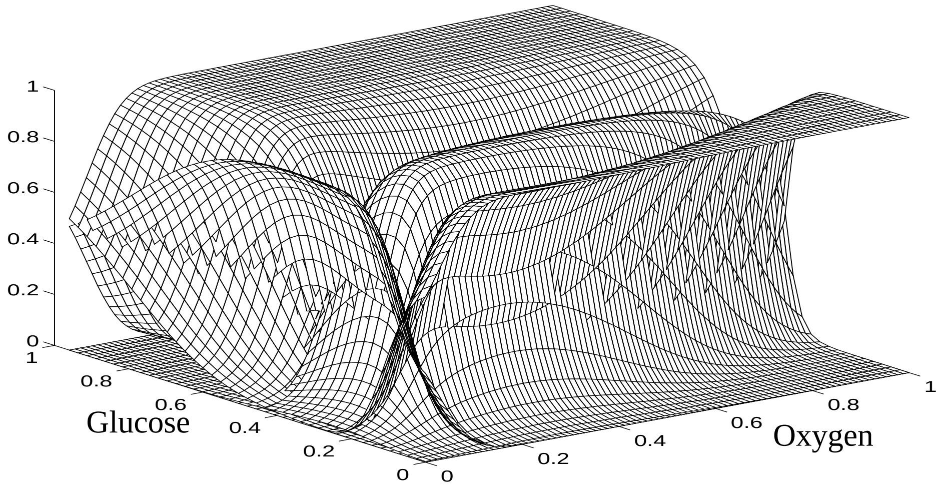


Figure 8.1: Interpolation functions for the four regimes in the plane spanned by oxygen concentration and glucose concentration. Notice that the axes are scaled.

The 448 unknown model parameters are estimated using the least-squares method, and simulated data from 600 batches, each run for 10 $h$, and all states "measured" every 0.5 $h$. For every batch, the initial states $x_{10}$ and $x_{40}$ were randomly chosen from the intervals $[0.4, 0.5]$ and $[40, 50]$, respectively. The control input trajectories were designed by randomly selecting between 0 and 2 step changes, within the allowable ranges of both temperature and pH, during the batch. A global linear model was also found using the same estimation method and identification data.

Both models were visually "validated" on a number of independent batches not used for identification. In these batches, the pH and temperature were randomly changed every 0.5 $h$. A typical ballistic prediction is shown in Fig. 8.3, and indicates that the prediction accuracy of the non-linear model is satisfactory, while the linear model has poorer prediction capabilities on the full batch length.

Due to the inaccuracy in the models used for MPC, and because not all states are measured, state estimators are implemented, using a time-varying extended Kalman filter for the non-linear model, and a time-varying Kalman filter for the linear model. The initial state-estimates of the filters equals the initial states of

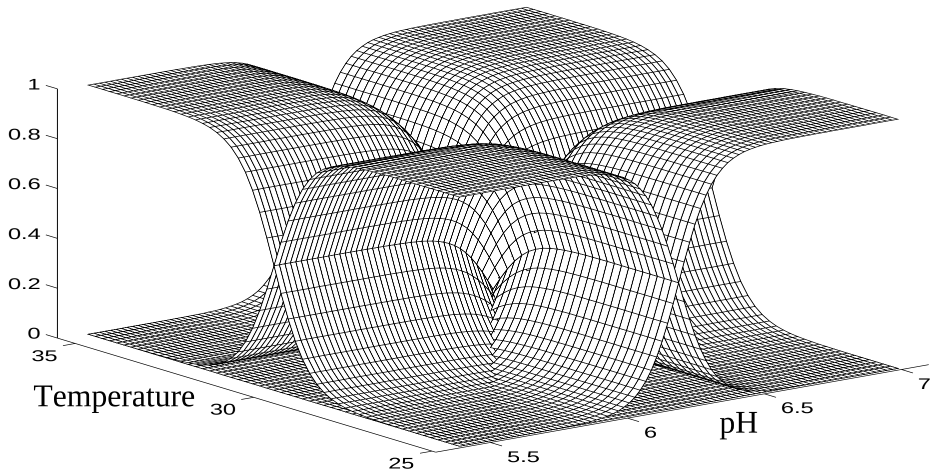


Figure 8.2: Interpolation functions for the four regimes in the plane spanned by temperature and pH.

the "true system". The covariance matrices were tuned to make the estimator loop fast compared to the system dynamics.

### 8.3.3 Model Predictive Control

The objective of the MPC is to maximize the average production rate of gluconic acid, neglecting the costs of substrate, cells, and separation. The time $T_c$ from finishing one batch to starting the next, due to emptying, cleaning and initializing the reactor, is $T_c = 1\ h$.

This optimization problem is formulated as

$$\max_{(u,T_f)\in(U,\mathcal{T})} \frac{x_3(T_f)}{T_f + T_c} \tag{8.2}$$

subject to the model equations, and the restrictions $5.4 \leq u_1 \leq 7.0$ , $25.0 \leq u_2 \leq 35.4$, and $x_1, ..., x_5 \geq 0$ at all time. The trajectories are optimized from time $t$ to the batch end time $T_f$. The batch end time is restricted to $\mathcal{T} = \{t, t + \Delta T, t + 2\Delta T, ...\}$, and the sampling interval is $\Delta T = 0.5\ h$. In the optimization, the current state is estimated using the extended Kalman filter, and the model is used to compute ballistic predictions from this initial value. The criterion (8.2) is maximized using a sequential quadratic programming algorithm with line search (MATLAB function **constr**, (Grace 1990)). The initial values to the search algorithm are constant input trajectories corresponding to pH = 5.6 and temp = 28.3.

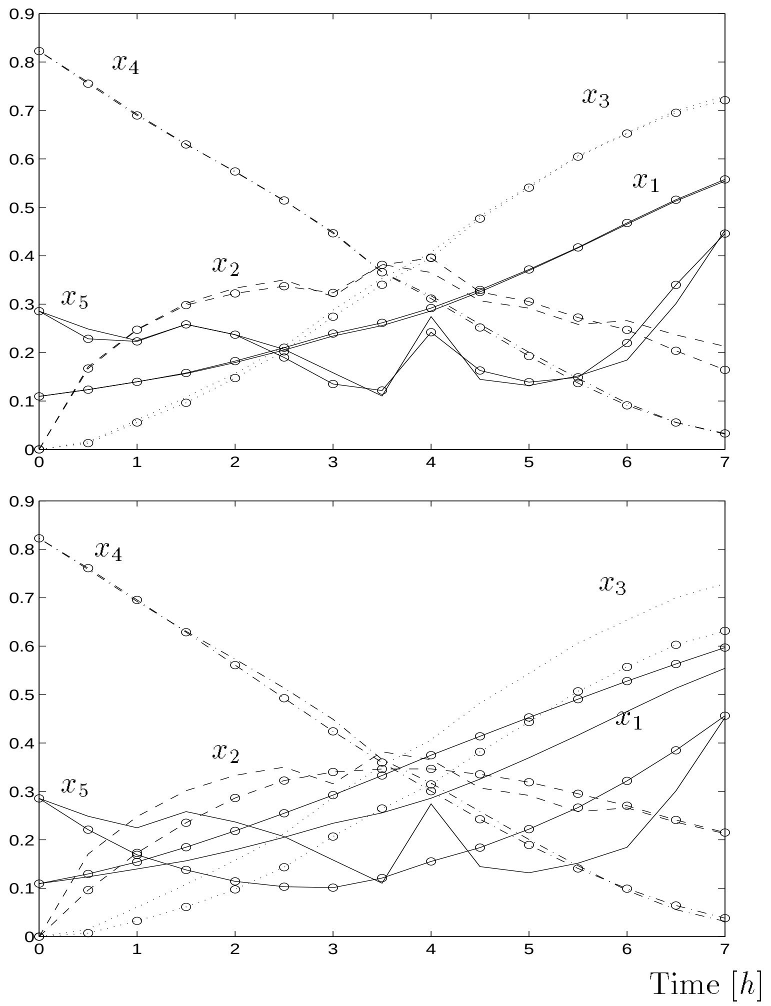


Figure 8.3: Simulation (ballistic prediction) from correct initial values for a typical batch, marked with circles (∘), and "true system" trajectories. The upper part is with the non-linear model based on 16 local models, while the lower part is with the linear model. Notice that the variables are scaled.

### 8.3.4 Results

The results of simulations using the five controllers described at the beginning of this section are summarized in Table 8.1. The results are averages computed over seven representative initial states. The temperature and pH trajectories for one typical initial state for these five cases are shown in Fig. 8.4. The corresponding state trajectories for the three MPC simulations are shown in Fig. 8.5.

Table 8.1: Summary of results.

| | Average prod. rate $\dot{p}$ $[g/lh]$ | Average end time $T_f$ $[h]$ |
|---|---|---|
| MPC, Ideal model | 6.03 | 5.5 |
| MPC, Local modeling | 5.90 | 5.5 |
| MPC, Linear model | 5.47 | 6.9 |
| OLOC, Local modeling | 5.88 | 5.1 |
| OLOC, Linear model | 5.51 | 6.4 |

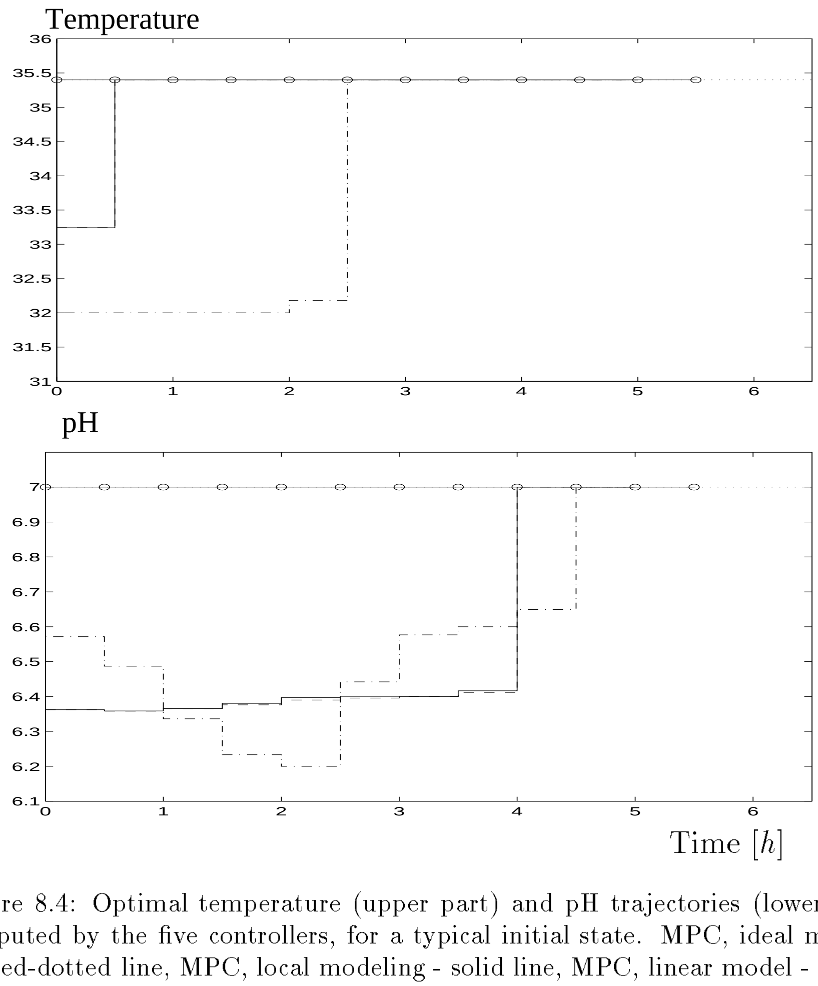


Figure 8.4: Optimal temperature (upper part) and pH trajectories (lower part) computed by the five controllers, for a typical initial state. MPC, ideal model - dashed-dotted line, MPC, local modeling - solid line, MPC, linear model - dotted line, OLOC, local modeling - dashed line, and OLOC, linear model - solid line with circles. Notice that the different trajectories have different end times.

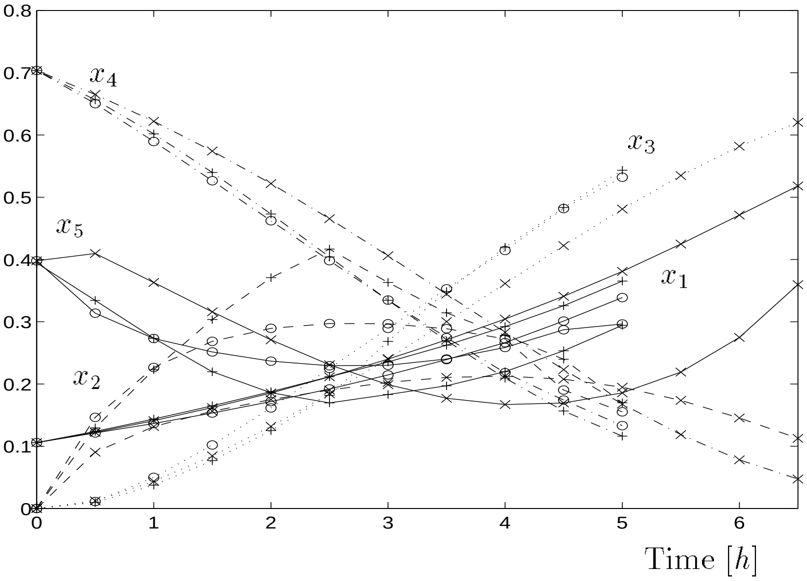


Figure 8.5: System trajectories for the three MPC simulations and a typical initial state. Notice that the variables are scaled. Lines marked $+$ are with the ideal model, lines marked $\circ$ are with operating regime based modeling, and lines marked $\times$ are with the linear model.

### 8.3.5 Discussion

The results show, as might be expected, significantly improved performance by moving from a linear to a non-linear model as the basis for MPC for this type of process. The improvement is somewhat limited by the fact that the control inputs for all five controllers are limited by hard upper constraints during significant parts of the batch. Furthermore, the results show that re-optimization during a batch by MPC may be advantageous compared to open-loop optimization. However, this is not true for the linear model, since the poor prediction capabilities make the optimization unreliable.

The experience from the modeling and identification suggests that with the operating regime based modeling method, it is both sufficient and necessary to have some rather elementary process knowledge to develop the model structure, and a set of informative empirical data for parameter estimation. In particular the decomposition into operating regimes is a critical part of the modeling, where it important to use process knowledge to get a sound model structure. However, the applied process knowledge is significantly less than what would be needed for developing a mechanistic model based on mass balances. On the other hand, the amount of data is significantly larger than what would be needed for the identification of a mechanistic model. It should be mentioned that less empirical data may be sufficient to identify an accurate model. While this aspect is obviously important from a practical point of view, we have not investigated it here.

# Chapter 9

# Operating Regime based Adaptive Control

The application of operating regime based models for adaptive control may be attractive because the local nature of the model representation will automatically focus the attention of the on-line parameter identification algorithm to the parameters of only the local model(s) that are relevant at the current operating conditions. Hence, drift phenomena related to poor global excitation of certain parameters will be avoided, simply because we do not update the parameters the model is not sensitive with respect to. Notice that drift and bursting related to poor local excitation may still be present.

In this chapter, we will first present some preliminaries and briefly review the model representation, in sections 9.1 and 9.2, respectively. Then, in section 9.3, we describe and analyze a MIMO non-linear decoupler based on local models. Then we will analyze the properties of an on-line parameter identification algorithm that is modified according to the guidelines above, in section 9.4. In section 9.5, closed loop stability and robustness of an adaptive non-linear decoupler is analyzed. The theory is illustrated with a semi-realistic simulation example, in section 9.6, and the chapter ends up in a discussion.

The material in the chapter is from (Johansen 1994*a*, Johansen 1994*b*).

## 9.1 Preliminaries and Notation

The norm of a vector $x \in R^n$ is $||x|| = \sqrt{x^T x}$. The norm of a discrete-time transfer function $H(q^{-1})$ is defined as $||H(q^{-1})||_\infty = \text{ess}\sup_{\omega \in [0,2\pi)} |H(e^{-j\omega})|$, where $q^{-1}$ is the delay operator. We will use the mixed time/frequency-domain notation $y(t) = H(q^{-1})u(t)$ instead of the convolution $y(t) = h(t) * u(t)$, where $h(t)$ is the impulse response of $H(q^{-1})$. We will in this chapter neglect those initial values that do not affect the stability arguments. The analysis is based on input/output stability theory. Background material can be found in (Desoer and Vidyasagar 1975, Kreisselmeier and Anderson 1986, Datta 1993), and we will make use of the following lemmas frequently:

**Lemma 1** *Let $y(t) = H(q^{-1})u(t)$, where $H(q^{-1})$ is a causal transfer function that has all roots inside the unit circle. Suppose $\sum_{\tau=0}^{t} u^2(\tau) < \infty$ for all $t \geq 0$. Then for all $t \geq 0$*

$$\sum_{\tau=0}^{t} y^2(\tau) \leq ||H(q^{-1})||_\infty^2 \sum_{\tau=0}^{t} u^2(\tau)$$

**Proof:** See e.g. (Datta 1993).

□

**Lemma 2** *Let $s_1$ be a positive sequence, i.e. $s_1(t) \geq 0$ for all $t$. Let the transfer function $H(q^{-1})$ be defined by $H(q^{-1}) = K(1 - \sigma q^{-1})^{-1}$ for some $K \geq 0$ and $\sigma \in [0, 1)$. Suppose the sequence $s_2$ satisfies*

$$s_2(t) \leq H(q^{-1}) \left(\gamma_0 s_1(t) + \gamma_1 s_1(t-1) + ... + \gamma_n s_1(t-n)\right)$$

*for some non-negative integer $n$ and constants $\gamma_0, ..., \gamma_n \geq 0$. Then,*

$$s_2(t) \leq H(q^{-1}) \left(\gamma_0 + \gamma_1 \sigma^{-1} + ... + \gamma_n \sigma^{-n}\right) s_1(t)$$

*except for an exponentially decaying term due to possibly non-zero initial conditions.*

**Proof:**

$$s_2(t) \leq K\sigma^t s_2(0) + K \sum_{\tau=0}^{t} \sigma^{t-\tau} \left(\gamma_0 s_1(\tau) + ... + \gamma_n s_1(\tau - n)\right) \tag{9.1}$$

Clearly, for any $k \in \{0, ..., n\}$

$$\begin{aligned} \sum_{\tau=0}^{t} \sigma^{t-\tau} \gamma_k s_1(\tau - k) &= \sum_{\tau=-k}^{t-k} \sigma^{t-\tau-k} \gamma_k s_1(\tau) \\ &= \sigma^{-k} \sum_{\tau=0}^{t-k} \sigma^{t-\tau} \gamma_k s_1(\tau) + \text{exp. decaying term} \\ &\leq \sigma^{-k} \sum_{\tau=0}^{t} \sigma^{t-\tau} \gamma_k s_1(\tau) + \text{exp. decaying term} \end{aligned}$$

where the last inequality follows because $s_1$ and $\gamma_k$ are positive. Substituting into (9.1) gives the desired result.

□

## 9.2 Model Representation

The multi-input, multi-output NARX model based on local ARX models

$$\begin{aligned} y(t+d) &= A_{0,i}y(t) + ... + A_{n_y,i}y(t-n_y) + \\ &\qquad B_{0,i}u(t) + ... + B_{n_u,i}u(t-n_u) + C_i + \nu(t+d) \end{aligned}$$

and operating regimes introduced in Chapter 2 can be written in the form (see also (Priestley 1988))

$$\begin{aligned} y(t+d) &= A_0(z(t))y(t) + ... + A_{n_y}(z(t))y(t-n_y) + \\ &\qquad B_0(z(t))u(t) + ... + B_{n_u}(z(t))u(t-n_u) + C(z(t)) + \nu(t+d) \end{aligned} \quad (9.2)$$

The input- and output-vectors are defined by $y(t) = (y_1(t), ..., y_m(t))^T$ and $u(t) = (u_1(t), ..., u_r(t))^T$, and $m$ and $r$ are the number of outputs and inputs, respectively. It is assumed that $r \geq m$. The vector of positive integers $d = (d_1, ..., d_m)^T$ is the system's time-delay. Alternatively, $d$ can be viewed as the system's relative degree (Monaco and Normand-Cyrot 1987). For later use, we define $d^\star = \max(d_1, ..., d_m)$, and for convenience, we introduce the notation $y(t+d) = (y_1(t+d_1), ..., y_m(t+d_m))^T$ and $\nu(t+d) = (\nu_1(t+d_1), ..., \nu_m(t+d_m))^T$. The $m \times m$-matrices $A_0, ..., A_{n_y}$, the $m \times r$-matrices $B_0, ..., B_{n_u}$ and the $m$-vector $C$ depend on the operating point $z(t) \in Z$ in the following way

$$\left.\begin{aligned} A_0(z(t)) &= \textstyle\sum_{i=1}^N A_{0,i} w_i(z(t)) \\ &\vdots \\ A_{n_y}(z(t)) &= \textstyle\sum_{i=1}^N A_{n_y,i} w_i(z(t)) \\ B_0(z(t)) &= \textstyle\sum_{i=1}^N B_{0,i} w_i(z(t)) \\ &\vdots \\ B_{n_u}(z(t)) &= \textstyle\sum_{i=1}^N B_{n_u,i} w_i(z(t)) \\ C(z(t)) &= \textstyle\sum_{i=1}^N C_i w_i(z(t)) \end{aligned}\right\} \quad (9.3)$$

where the $w_i$-functions are interpolation functions, cf. Chapter 2. The operating set $Z$ is an Euclidean space, or a subset of one. The sequence $\nu$ contains unknown, unstructured uncertainty, such as measurement noise, disturbances, unmodeled dynamics, modeling error, effects caused by sampling a continuous-time system etc. The "true system" may for example be an infinite-dimensional continuous-time system, approximated by the above finite-dimensional discrete-time model representation, and in addition to external phenomena like noise and disturbances, the approximation error contributes to the unstructured uncertainty. The information vector is defined by $\psi(t) = (y^T(t), ..., y^T(t-n_y), u^T(t), ..., u^T(t-n_u))^T$.

## 9.3 Non-linear Decoupling Control

The control structure we will study in this chapter is a discrete-time non-linear decoupler (Monaco and Normand-Cyrot 1987). Such controllers are often called feedback linearizing controllers (Isidori 1989) because a non-linear feedback and a state-transform are designed in such a way that the nominal system, with a new set of input variables, is rendered linear. With the NARX model representation, the model is already in the appropriate controlable canonical form, so the state-transform in not needed. We will choose the feedback such that the nominal system is not only linearized, but also decoupled.

The control objective is to track a given, bounded reference sequence $y^\star(t) = (y_1^\star(t), ..., y_m^\star(t))^T$ while rejecting the impact of disturbances and noise (contained in $\nu$). Consider the feedback

$$\begin{aligned} v(t) \;=\; & A_0(z(t))y(t) + ... + A_{n_y}(z(t))y(t-n_y) + \\ & B_0(z(t))u(t) + ... + B_{n_u}(z(t))u(t-n_u) + C(z(t)) \end{aligned} \tag{9.4}$$

where $v(t) = (v_1(t) \cdots v_m(t))^T$ is an external input. Eq. (9.4) is implicit in $u(t)$, but can easily be solved if $z(t)$ does not explicitly depend on $u(t)$, since $B_0(z(t))$ is non-singular by the definition of the relative degree $d$ (Monaco and Normand-Cyrot 1987). If, however, $z(t)$ is a function of $u(t)$, this function must be invertible, in the sense that (9.4) has at least one solution for $u(t)$ at any time $t \geq 0$. Assuming a solution $u(t)$ can be found at all $t \geq 0$, the feedback (9.4) gives the closed loop input/output behavior

$$y_j(t + d_j) \;=\; v_j(t) + \nu_j(t + d_j) \tag{9.5}$$

for $j \in \{1, ..., m\}$. The nominal system seen from $v$ to $y$ is transformed into a known linear and decoupled system. The problem is reduced to one of controlling the simple system (9.5) using $v(t)$ as the new control input. However, the unmodeled dynamics introduce couplings. We must therefore design a controller for (9.5) that is robust with respect to unmodeled dynamics, and able to reject the disturbances while tracking the reference trajectory. If the dead-times $d_1, ..., d_m$ are small, then $m$ decentralized PI or I controllers tuned with some suitable stability margins is a good choice. However, a large dead-time will significantly limit the performance attainable with a PI or I controller, and more advanced strategies like a Smith-predictor may be considered. In the following, suppose the external inputs are chosen as

$$v_j(t) \;=\; y_j^\star(t + d_j) - G_j(q^{-1})\tilde{y}_j(t) \tag{9.6}$$

where $G_j(q^{-1}) = P_j(q^{-1})/Q_j(q^{-1})$ is a discrete-time transfer function, and the tracking error is defined as $\tilde{y}(t) = (\tilde{y}_1(t), ..., \tilde{y}_m(t))^T = y(t) - y^\star(t)$. Combining the model equation (9.2) with the linearizing feedback (9.4) and the feedback (9.6), the closed loop satisfies

$$\tilde{y}_j(t) \;=\; M_j(q^{-1})\nu_j(t) \tag{9.7}$$

for $j \in \{1, ..., m\}$, where

$$M_j(q^{-1}) \;=\; \frac{1}{1 + q^{-d_j}G_j(q^{-1})} \;=\; \frac{Q_j(q^{-1})}{Q_j(q^{-1}) + q^{-d_j}P_j(q^{-1})}$$

Design techniques for robust decentralized control may be applied for designing decentralized controllers $G_j(q^{-1})$ such that the closed loop has the desired disturbance rejection and tracking properties.

Suppose the model (9.2) can be inverted, i.e. there exists a globally defined function $f$ such that

$$u(t) = f(u(t-1), ..., u(t-n_u), y(t+d), y(t), y(t-1), ..., y(t-n_y), \nu(t+d)) \quad (9.8)$$

This system should be viewed as the inverse of (9.2), with output $u$ and inputs $y$ and $\nu$. Boundedness of the control input requires a restriction of the class of nominal models to ones where the inverse system (9.8) is globally exponentially stable. Such systems are often called minimum phase system (Byrnes and Isidori 1984, Monaco and Normand-Cyrot 1987).

For the purpose of analyzing the closed loop, we introduce a relative bound on the unstructured uncertainty:

$$n(t) = \delta n(t-1) + ||\psi(t)|| + 1 \quad (9.9)$$

$$|\nu_j(t)| \leq \mathcal{V}_j n(t-d_j) \quad (9.10)$$

for $j \in \{1, ..., m\}$, where $n(0) > 0$, $\delta \in (0,1)$, and $\mathcal{V}_j \geq 0$ is a small constant. By construction, $n(t)$ will bound all other signals in the closed loop, so (9.10) requires the unstructured uncertainty to be relatively small compared to the inputs and outputs. Notice that the uncertainty formulation is sufficiently general to include structural modeling error and unmodeled dynamics (Kreisselmeier and Anderson 1986), and observe in particular that non-minimum phase effects can be hidden in this uncertainty. Of course, the requirement that the uncertainty must be relatively small will require that these non-minimum phase effects are small.

Let $\rho \in (0,1)$ and $K_\rho \geq 0$ be defined such that the impulse response coefficients of $M_1(q^{-1}), ..., M_m(q^{-1})$ are all bounded by $K_\rho \rho^t$ for all $t \geq 0$. Exponential stability of the transfer functions $M_1(q^{-1}), ..., M_m(q^{-1})$ will ensure that this is possible.

**Theorem 1** *Let the system (9.2) be controlled by (9.4) and (9.6), and suppose*

1. *The reference sequence $y^\star$ is bounded and known $d^\star$ steps ahead in time. The bound is denoted $K^\star = \sup_{t \geq 0} ||y^\star(t)||$.*
2. *The model's relative degree $d = (d_1, ..., d_m)^T$, and the order parameters $n_u$ and $n_y$ are known.*
3. *The inverse system (9.8) is globally exponentially stable, i.e. there exist constants $K_u \geq 0$ and $\sigma \in [0,1)$ such that for all $t \geq 0$*

$$||u(t)|| \leq K_u \sum_{\tau=0}^{t} (||y(\tau+d)|| + ||\nu(\tau+d)|| + 1)$$

4. *The transfer functions $M_1(q^{-1}), ..., M_m(q^{-1})$ are proper and exponentially stable.*
5. *Eq. (9.4) has at least one solution for $u(t)$ at all $t \geq 0$.*

*6. The unstructured uncertainty satisfies (9.10).*

*Let* $\mathcal{V} = \max(\mathcal{V}_1, ..., \mathcal{V}_m)$, *and*

$$\kappa = \frac{m\mathcal{V}}{1-\delta}\left(\left(\frac{K_2 K_\rho}{1-\rho}\right)\left(\frac{K_1 K_u}{1-\sigma} + \delta^{-d^\star}\right) + \frac{K_1 K_u}{1-\sigma}\right)$$

*where the constants* $K_1$ *and* $K_2$ *are defined in the proof of the theorem. If* $\kappa < 1$, *then for arbitrary initial conditions, all variables in the closed loop are bounded and the average squared tracking error satisfies for all* $t > 0$

$$\frac{1}{t}\sum_{\tau=0}^{t} ||\tilde{y}(\tau)||^2 \leq K_4 \; + \; \text{exp. decaying term} \tag{9.11}$$

*where*

$$K_4 = \left(\frac{K_3'}{1-\kappa}\right)^2 \left(\frac{K_\rho}{1-\rho}\right)^2 \sum_{j=1}^{m} \mathcal{V}_j^2$$

*and* $K_3'$ *is defined in the proof of the theorem.*

**Proof:** From the definition of the information vector, we get

$$||\psi(t)|| \leq ||u(t)|| + ... + ||u(t-n_u)|| + ||y(t)|| + ... + ||y(t-n_y)||$$

Together with the definition of the normalizing signal, and Lemma 2, this yields the inequality

$$n(t) \leq \frac{1}{1-\delta q^{-1}}\left(K_1||u(t)|| + K_2||y(t)|| + 1\right)$$

where $K_1 = 1+\delta^{-1}+...+\delta^{-n_u}$ and $K_2 = 1+\delta^{-1}+...+\delta^{-n_y}$. Using the exponential stability of the inverse system, we get

$$\begin{aligned} n(t) &\leq \frac{1}{1-\delta q^{-1}}\left(\frac{K_1 K_u}{1-\sigma q^{-1}}\left(||y(t+d)|| + ||\nu(t+d)|| + 1\right) + K_2||y(t)|| + 1\right) \\ &\leq H_1(q^{-1})||y(t+d)|| + H_2(q^{-1})||\nu(t+d)|| + K_3 \end{aligned}$$

where

$$\begin{aligned} H_1(q^{-1}) &= \frac{K_1 K_u}{(1-\delta q^{-1})(1-\sigma q^{-1})} + \frac{K_2 \delta^{-d^\star}}{1-\delta q^{-1}} \\ H_2(q^{-1}) &= \frac{K_1 K_u}{(1-\delta q^{-1})(1-\sigma q^{-1})} \\ K_3 &= \frac{1}{1-\delta} + ||H_2(q^{-1})||_\infty \end{aligned}$$

This gives

$$n(t) \leq H_1(q^{-1})||\tilde{y}(t+d)|| + H_2(q^{-1})||\nu(t+d)|| + K_3'$$

where $K_3' = K_3 + K^\star||H_1(q^{-1})||_\infty$. Next, from (9.7) we get

$$\begin{aligned} n(t) &\le \sum_{j=1}^{m} \left(H_1(q^{-1})M_j(q^{-1}) + H_2(q^{-1})\right)|\nu_j(t+d_j)| + K_3' \\ &\le \mathcal{V}R(q^{-1})n(t) + K_3' \end{aligned}$$

where

$$R(q^{-1}) \;=\; H_1(q^{-1})\sum_{j=1}^{m} M_j(q^{-1}) + mH_2(q^{-1})$$

We define the linear time-invariant system

$$\overline{n}(t) \;=\; \mathcal{V}R(q^{-1})\overline{n}(t) + K_3'$$

The initial condition is chosen as $\overline{n}(0) = n(0)$. It follows that $\overline{n}(t) \ge n(t)$ for all $t \ge 0$. By the Small Gain Theorem (Zames 1966), it follows that $\overline{n}$ is a bounded sequence, since $\mathcal{V}||R(q^{-1})||_\infty \le \kappa < 1$. Hence, $n$ is also a bounded sequence, and it follows from the definition of $n$ that $y$ and $u$ are bounded sequences. We get for $t \ge 0$

$$n(t) \;\le\; \frac{K_3'}{1 - \mathcal{V}R(q^{-1})} \;\le\; \frac{K_3'}{1-\kappa}$$

From (9.7) and Lemma 1 we get

$$\begin{aligned} \sum_{\tau=0}^{t} \tilde{y}_j^2(\tau) &\le ||M_j(q^{-1})||_\infty^2 \sum_{\tau=0}^{t} \nu_j^2(\tau) \\ \frac{1}{t}\sum_{\tau=0}^{t} ||\tilde{y}(\tau)||^2 &\le \left(\frac{K_\rho}{1-\rho}\right)^2 \left(\frac{K_3'}{1-\kappa}\right)^2 \sum_{j=1}^{m} \mathcal{V}_j^2 \end{aligned}$$

**Remarks:** Eq. (9.11) for the average tracking error indicates that the tracking error scales with the unstructured uncertainty, and in particular that $||\tilde{y}(t)|| \to 0$ exponentially when $t \to \infty$ if $\mathcal{V} = 0$. The exponentially decaying term will be of the form $K_\kappa \kappa^t$ where $K_\kappa \ge 0$ is some constant, and is due to non-zero initial conditions that have been neglected in the analyzis, as these do not affect the stability of the system. The expression for $\kappa$ is rather complex, and several of the constants involved (such as $\mathcal{V}$, $\sigma$ and $K_u$) are usually not known. The result may therefore best be interpreted in a qualitative manner. The most important observation is that $\kappa$ scales with $\mathcal{V}$, and that stability requires the unstructured uncertainty to be small.

The proof is based on input/output stability theory (Desoer and Vidyasagar 1975), which is well known to give very conservative results in many cases. This also suggests that one should emphasize the qualitative interpretation of the results. If one wants to emphasize the quantitative bounds, it must be observed that appropriate scaling of the various signals is important in order to make the results as little conservative as possible.

□

## 9.4 Parameter Estimation

The interpolation functions are assumed not to contain any unknown parameters, and all the local model structures are linear in the parameters. Hence, the global representation (9.2) is also linear in the parameters. It can be reformulated into a linear regression form

$$y_j(t+d_j) = \varphi^T(t)\theta_j^\star + \nu_j(t+d_j) \tag{9.12}$$

well suited for parameter estimation, for $j \in \{1, ...m\}$. The regressor vector is given by

$$\varphi(t) = \begin{pmatrix} w_1(z(t)) \\ \psi(t) w_1(z(t)) \\ \vdots \\ w_N(z(t)) \\ \psi(t) w_N(z(t)) \end{pmatrix}$$

and the "true parameter" vector $\theta_j^\star$ is defined as

$$\theta_j^\star = \begin{pmatrix} \theta_{j,1}^\star \\ \vdots \\ \theta_{j,N}^\star \end{pmatrix}, \quad \theta_{j,i}^\star = \begin{pmatrix} \mathrm{row}_j(c_i) \\ \mathrm{row}_j(A_{0,i}) \\ \vdots \\ \mathrm{row}_j(A_{n_y,i}) \\ \mathrm{row}_j(B_{0,i}) \\ \vdots \\ \mathrm{row}_j(B_{n_u,i}) \end{pmatrix}$$

for $j \in \{1, ..., m\}$. The vector $\mathrm{row}_j(A)$ is the $j$-th row of the matrix $A$. It is assumed that a closed and convex sub-set $\Theta_j$ of the parameter space is known, and that $\theta_j^\star \in \Theta_j$.

In the following, we will investigate the properties of recursive parameter estimation algorithms for the linearly parameterized model (9.12). The parameter estimate at time $t$ is denoted $\hat{\theta}_j(t)$, and the parameter error vector is defined by $\tilde{\theta}_j(t) = \hat{\theta}_j(t) - \theta_j^\star$. We define the *prediction error*

$$e_j(t) = \hat{y}_j(t|t-1) - y_j(t) = \varphi^T(t-d_j)\tilde{\theta}_j(t-1) - \nu_j(t) \tag{9.13}$$

For convenience, the normalizing signal (Praly 1984) is redefined as

$$n(t) = \delta n(t-1) + ||\varphi(t)||$$

for some constant $\delta \in (0,1)$ and initial condition $n(0) > 0$. Again, assume there exists a small constant $\mathcal{V}_j \geq 0$ such that

$$|\nu_j(t)| \leq \mathcal{V}_j n(t-d_j) \tag{9.14}$$

for $j \in \{1, ..., m\}$. The *normalized prediction error* is defined by

$$\varepsilon_j(t) = e_j(t)/n(t - d_j)$$

Consider the projection algorithm with relative dead-zone (Kreisselmeier and Anderson 1986) and parameter projection:

$$\hat{\theta}'_j(t) = \hat{\theta}_j(t-1) - \Lambda_j(t)\frac{\varphi(t-d_j)}{\varphi^T(t-d_j)\varphi(t-d_j)}n(t-d_j)\mathcal{D}(\varepsilon_j(t))$$

$$\hat{\theta}_j(t) = \mathcal{P}_{\Theta_j}\left(\hat{\theta}'_j(t)\right)$$

$$\mathcal{D}(\varepsilon_j(t)) = \begin{cases} \varepsilon_j(t) + \overline{d}_j^0 & \text{if } \varepsilon_j(t) < -\overline{d}_j^0 \\ 0 & \text{if } |\varepsilon_j(t)| \leq \overline{d}_j^0 \\ \varepsilon_j(t) - \overline{d}_j^0 & \text{if } \varepsilon_j(t) > \overline{d}_j^0 \end{cases}$$

An initial estimate $\hat{\theta}_j(0) \in \Theta_j$ is assumed to be given. Observe that $||\varphi(t - d_j)||$ is bounded away from zero for all $t \geq 0$ because $\sum_{i=1}^N w_i(z(t)) = 1$ for all $z(t) \in Z$. The continuous projection operator $\mathcal{P}_{\Theta_j}$ projects $\hat{\theta}'_j(t)$ to the closest point in $\Theta_j$ (using Euclidean norm). The existence and uniqueness of a closest point is guaranteed by the closedness and convexity of $\Theta_j$. The continuous function $\mathcal{D}$ is the dead-zone function with dead-zone $\overline{d}_j^0$. The purpose of the time-varying matrix $\Lambda_j(t)$ is to select which local models to update, in addition to being a gain-matrix. Traditionally, this matrix is chosen as $\Lambda_j(t) = \lambda I$, where $\lambda \in (0, 2)$ is the scalar gain (Goodwin, Ramadge and Caines 1980). Using local models, it may be desirable to choose $\Lambda_j(t) = \text{diag}\,(\lambda_{j,1}(t)I, ..., \lambda_{j,N}(t)I)$, where

$$\lambda_{j,i}(t) = \begin{cases} \lambda, & \text{for } i \in \mathcal{I}_j(t) \\ 0, & \text{otherwise} \end{cases} \tag{9.15}$$

where $\mathcal{I}_j(t) \subset \{1, ..., N\}$ is a time-varying and operating-point dependent index set. To see the motivation behind this, suppose we use Gaussian local model validity functions

$$\rho_i(z) = \exp\left(-\frac{1}{2}(z - z_i)^T \Sigma_i^{-1}(z - z_i)\right)$$

The constant vector $z_i$ is of the same dimension as $z$, and $\Sigma_i$ is a positive definite matrix of appropriate dimension. Remember that the regressor vector $\varphi(t)$ has sub-vectors $\psi(t)w_i(z(t))$. Using the projection algorithm with $\Lambda_j(t) = \lambda I$ will involve the updating of the parameter estimates of all the local models, since $w_i(z) > 0$ for all $z \in Z$ and $i \in \{1, ..., N\}$. This may be undesirable, since it means that local models corresponding to irrelevant operating regimes will have their parameter estimates slightly updated. If the system operates within a single operating regime for a long time, this may cause drift of the parameter estimates which do not correspond to this regime. This problem can be handled in at least two ways:

1. The local model validity functions can be chosen such that they are exactly zero at distant operating points, instead of going asymptotically to zero as

the Gaussian. This will give zeros in the regressor vector and automatically prevent the updating of the parameters which correspond to irrelevant operating regimes. If the dimension of $z(t)$ is high, designing such local model validity functions is sometimes difficult, since we do not always know an a priori bound for $z(t)$, and because for any $z(t)$, at least one of the local model validity functions must be non-zero to ensure that the global model is complete. It may also be desirable that the local model validity functions are smooth, which may cause an additional design problem.

2. It may sometimes be simpler to use smooth, strictly positive local model validity functions, and ensure by modifying the estimation algorithm that only the most relevant local model(s) are updated at a given time.

Using the second approach, we need to design the gain-matrix $\Lambda_j(t)$ such that at a given time $t$, only the most relevant local model parameter estimates are updated, and the stability properties of the parameter estimation algorithm are preserved:

**Theorem 2** *Suppose the data are generated by (9.12), $\Theta_j$ is a known compact and convex sub-set of the parameter space with the property $\theta_j^\star \in \Theta_j$, and the unstructured uncertainty satisfies (9.14). Consider the projection algorithm with dead-zone $\overline{d}_j^0 = \mathcal{V}_j$ and gain (9.15), where*

$$\mathcal{I}_j(t) = \left\{ i \in \{1, ..., N\} \;\middle|\; w_i(z(t)) > \frac{|\mathcal{D}(\varepsilon_j(t))| n(t-d_j)}{4N(1 + ||\psi(t-d_j)||)\Phi_{j,i}} \right\}$$

*and $\Phi_{j,i} \geq \sup_{\theta_{j,i} \in \Theta_{j,i}} ||\theta_{j,i} - \theta_{j,i}^\star||$ for $i \in \{1, ..., N\}$. If $\lambda \in (0, \lambda_{max})$, where $\lambda_{max} = 3/2$, this algorithm has the properties*

$$||\tilde{\theta}_j(t+1)|| \leq ||\tilde{\theta}_j(t)|| \;\leq\; ||\tilde{\theta}_j(0)|| \tag{9.16}$$

$$\sum_{\tau=0}^{t} \mathcal{D}^2(\varepsilon_j(\tau)) \leq \frac{1}{\lambda(\lambda_{max} - \lambda)} ||\tilde{\theta}_j(0)||^2 \tag{9.17}$$

$$|\varepsilon_j(t)| \leq \mathcal{V}_j + |\mathcal{D}(\varepsilon_j(t))| \tag{9.18}$$

$$\sum_{\tau=0}^{t} ||\Delta\tilde{\theta}_j(\tau)||^2 \leq \frac{\lambda}{\lambda_{max} - \lambda} ||\tilde{\theta}_j(0)||^2 \tag{9.19}$$

*for all $t \geq 0$.*

**Proof:** First observe that the existence of $\Phi_{j,i}$ is guaranteed by the boundedness of $\Theta_{j,i}$. Boundedness of $||\tilde{\theta}_j(t)||$ follows directly from the boundedness of $\Theta_j$ and the parameter projection. The normalized prediction error can be written

$$\varepsilon_j(t) = \frac{\varphi^T(t-d_j)}{n(t-d_j)} \tilde{\theta}_j(t-1) - \frac{\nu_j(t)}{n(t-d_j)}$$

From the definition of the relative dead-zone and (9.14), there exists a sequence $\beta_j$ such that $0 \leq \beta_j(t) \leq 1$ and

$$\mathcal{D}(\varepsilon_j(t)) = \beta_j(t) \frac{\varphi^T(t-d_j)}{n(t-d_j)} \tilde{\theta}_j(t-1) \tag{9.20}$$

Hence,

$$\tilde{\theta}'_j(t) = \left(I - \beta_j(t)\Lambda_j(t)\frac{\varphi(t-d_j)\varphi^T(t-d_j)}{\varphi^T(t-d_j)\varphi(t-d_j)}\right)\tilde{\theta}_j(t-1) \tag{9.21}$$

We define the Lyapunov-function $V_j(t) = ||\tilde{\theta}_j(t)||^2$. Using the projection operator property $||\tilde{\theta}_j(t)|| \leq ||\tilde{\theta}'_j(t)||$, we get

$$\begin{aligned} V_j(t) - V_j(t-1) &\leq -2\beta_j(t)\tilde{\theta}_j^T(t-1)\Lambda_j(t)\frac{\varphi(t-d_j)\varphi^T(t-d_j)}{\varphi^T(t-d_j)\varphi(t-d_j)}\tilde{\theta}_j(t-1) \\ &+\beta_j^2(t)\tilde{\theta}_j^T(t-1)\frac{\varphi(t-d_j)\varphi^T(t-d_j)\Lambda_j^T(t)\Lambda_j(t)\varphi(t-d_j)\varphi^T(t-d_j)}{(\varphi^T(t-d_j)\varphi(t-d_j))^2}\tilde{\theta}_j(t-1) \end{aligned}$$

If $\beta_j(t) \neq 0$ and $\tilde{\theta}_j^T(t-1)\varphi(t-d_j) \neq 0$, then

$$V_j(t) - V_j(t-1) \leq -\gamma_j(t)\left(\frac{n^2(t-d_j)\mathcal{D}^2(\varepsilon_j(t))}{\varphi^T(t-d_j)\varphi(t-d_j)}\right) \tag{9.22}$$

where

$$\gamma_j(t) = 2\frac{\tilde{\theta}_j^T(t-1)\Lambda_j(t)\varphi(t-d_j)}{\beta_j(t)\tilde{\theta}_j^T(t-1)\varphi(t-d_j)} - \frac{\varphi^T(t-d_j)\Lambda_j^T(t)\Lambda_j(t)\varphi(t-d_j)}{\varphi^T(t-d_j)\varphi(t-d_j)} \tag{9.23}$$

By the definition of $\mathcal{I}_j(t)$, we get

$$\begin{aligned} \frac{\tilde{\theta}_j^T(t-1)\Lambda_j(t)\varphi(t-d_j)}{n(t-d_j)\mathcal{D}(\varepsilon_j(t))} &= \frac{\lambda\tilde{\theta}_j^T(t-1)\varphi(t-d_j) - \tilde{\theta}_j^T(t-1)\left(\lambda I - \Lambda_j(t)\right)\varphi(t-d_j)}{n(t-d_j)\mathcal{D}(\varepsilon_j(t))} \\ &\geq \frac{\lambda}{\beta_j(t)} - \\ &\quad \sum_{i \notin \mathcal{I}_j(t)} \frac{\lambda||\tilde{\theta}_{j,i}(t-1)|| \cdot (1 + ||\psi(t-d_j)||)w_i(z(t-d_j))}{n(t-d_j)\mathcal{D}(\varepsilon_j(t))} \\ &\geq \lambda\left(1 - \left(\frac{N - \#\mathcal{I}_j(t)}{4N}\right)\right) \geq \frac{3}{4}\lambda \end{aligned}$$

From (9.23) we get $\gamma_j(t) \geq 3\lambda/2 - \lambda^2 > 0$. If $\beta_j(t) = 0$ or $\tilde{\theta}_j^T(t-1)\varphi(t-d_j) = 0$, then it follows from (9.21) that $V_j(t) = V_j(t-1)$, and we may conclude from (9.22) that $V_j$ is a non-increasing sequence, and (9.16) follows. Furthermore, from (9.22)

$$\mathcal{D}^2(\varepsilon_j(t)) \leq \frac{1}{\lambda(3/2-\lambda)}\left(\frac{\varphi^T(t-d_j)\varphi(t-d_j)}{n^2(t-d_j)}\right)(V_j(t-1) - V_j(t))$$

if $\beta_j(t) \neq 0$, and from (9.20) it follows that $\mathcal{D}(\varepsilon_j(t)) = 0$ if $\beta_j(t) = 0$. Hence

$$\sum_{\tau=0}^{t} \mathcal{D}^2(\varepsilon_j(t)) \leq \frac{1}{\lambda(3/2-\lambda)}||\tilde{\theta}_j(0)||^2$$

and (9.17) follows. Directly from the definition of the dead-zone, it follows that $|\varepsilon_j(t)| \leq \mathcal{V}_j + |\mathcal{D}(\varepsilon_j(t))|$. Finally,

$$\begin{aligned}
\sum_{\tau=0}^{t} ||\Delta\tilde{\theta}_j(\tau)||^2 &\leq \sum_{\tau=0}^{t} ||\tilde{\theta}_j'(\tau) - \tilde{\theta}_j(\tau-1)||^2 \\
&\leq \sum_{\tau=0}^{t} \left(\beta_j(\tau)\Lambda_j(\tau)\frac{\varphi(\tau-d_j)\varphi^T(\tau-d_j)}{\varphi^T(\tau-d_j)\varphi(\tau-d_j)}\tilde{\theta}_j(\tau-1)\right)^T \cdot \\
&\qquad \left(\beta_j(\tau)\Lambda_j(\tau)\frac{\varphi(\tau-d_j)\varphi^T(\tau-d_j)}{\varphi^T(\tau-d_j)\varphi(\tau-d_j)}\tilde{\theta}_j(\tau-1)\right) \\
&\leq \frac{\lambda}{3/2-\lambda}\sum_{\tau=0}^{t} \mathcal{D}^2(\varepsilon_j(\tau))
\end{aligned}$$

**Remarks:** No assumptions on the data, such as boundedness or persistence of excitation, are made in this theorem.

Similar results can be found if the dead-zone modification is replaced by one of several other possible robustness modifications, including $\sigma$-modification and $\epsilon$-modification (Ioannou and Datta 1991, Narendra and Annaswamy 1989).

□

The assumption that $\Theta_j$ is a compact set is made because explicit bounds $\Phi_{j,i}$ on the parameter error must be known to compute $\mathcal{I}_j(t)$. If we use the parameter estimation algorithm without thresholding, i.e. $\mathcal{I}_j(t) = \{1, ..., N\}$, then $\Theta_j$ may be unbounded, and the same conclusions as in Theorem 2 with $\lambda_{max} = 2$ can be proved.

The threshold suggested in the definition of $\mathcal{I}_j(t)$ may be too close to zero for practical purposes, because of the conservative bounds applied in the derivation of this threshold. We therefore suggest that a somewhat larger threshold is applied, typically in the range 0.05 to 0.25.

## 9.5 Adaptive Control

Any unknown parameters in (9.4) can be replaced by their estimates, which gives the *certainty equivalence* feedback

$$\begin{aligned}
v(t) = \; & \hat{A}_0(z(t),t)y(t) + ... + \hat{A}_{n_y}(z(t),t)y(t-n_y) + \\
& \hat{B}_0(z(t),t)u(t) + ... + \hat{B}_{n_u}(z(t),t)u(t-n_u) + \hat{C}(z(t),t) \quad (9.24)
\end{aligned}$$

It is assumed that this implicit equation has at least one solution for $u(t)$ at every time $t \geq 0$, and that a solution can be computed. Now, the closed loop is described by

$$\begin{aligned} y(t+d) \;=\; & v(t) + \left(A_0(z(t)) - \hat{A}_0(z(t),t)\right) y(t) \\ & +...+ \left(A_{n_y}(z(t)) - \hat{A}_{n_y}(z(t),t)\right) y(t-n_y) \\ & + \left(B_0(z(t)) - \hat{B}_0(z(t),t)\right) u(t) \\ & +...+ \left(B_{n_u}(z(t)) - \hat{B}_{n_u}(z(t),t)\right) u(t-n_u) \\ & + \left(C(z(t)) - \hat{C}(z(t),t)\right) + \nu(t+d) \end{aligned} \tag{9.25}$$

Notice that in this case both the unmodeled dynamics and the errors in the parameter estimates introduce couplings. Using the same arguments as in section 9.3, the external inputs can be chosen according to (9.6). An overview of the control structure is shown in Fig. 9.1. Combining the model equation (9.2) with the certainty equivalence feedback (9.24) and the feedback (9.6), the closed loop satisfies

$$\tilde{y}_j(t) \;=\; M_j(q^{-1})\left(-\varphi^T(t-d_j)\tilde{\theta}_j(t-d_j) + \nu_j(t)\right) \tag{9.26}$$

**Theorem 3** *Suppose system (9.2) is controlled by (9.24) and (9.6), and the projection algorithm with dead-zone* $\overline{d}_j^0$ *is applied to estimate the unknown parameter vectors* $\theta_j^\star$, *for* $j \in \{1,2,...,m\}$. *Furthermore, suppose assumptions 1-4 of Theorem 1 hold, and in addition*

1. *The sets* $\Theta_j$, *for* $j \in \{1,...,m\}$, *are closed and convex. If thresholding is used,* $\Theta_j$ *must also be bounded. Furthermore, these sets are such that* $\theta_j^\star \in \Theta_j$, *and equation (9.24) has at least one solution* $u(t)$ *at all* $t \geq 0$.
2. *The parameter estimators have gain* $\lambda \in (0, \lambda_{max})$, *where* $\lambda_{max} = 3/2$ *if thresholding is used, and* $\lambda_{max} = 2$ *otherwise.*
3. *The unstructured uncertainty satisfies (9.14), and the dead-zone is chosen as* $\overline{d}_j^0 = \mathcal{V}_j$, *for* $j \in \{1,...,m\}$.

*Let* $\mathcal{V} > \max(\mathcal{V}_1,...,\mathcal{V}_m)$, *and let* $\kappa$ *be as in Theorem 1. If* $\kappa < 1$, *then for arbitrary initial conditions, all variables in the closed loop are bounded and the average squared tracking error satisfies for all* $t > 0$

$$\frac{1}{t}\sum_{\tau=0}^{t} ||\tilde{y}(\tau)||^2 \;\leq\; 4K_4 + \frac{1}{t}K_5 + \text{exp. decaying term} \tag{9.27}$$

*where*

$$\begin{aligned} K_4 \;&=\; \left(\frac{K_3''}{1-\kappa}\right)^2 \left(\frac{K_\rho}{1-\rho}\right)^2 \sum_{j=1}^{m} \mathcal{V}_j^2 \\ K_5 \;&=\; 2\left(\frac{K_3''}{1-\kappa}\right)^2 \left(\frac{K_\rho}{1-\rho}\right)^2 \sum_{j=1}^{m} \left(\frac{2}{\lambda(\lambda_{max}-\lambda)} + \frac{(d_j-1)^2\lambda}{\lambda_{max}-\lambda}\right) ||\tilde{\theta}_j(0)||^2 \end{aligned}$$

*and* $K_3''$ *is defined in the proof of the theorem.*

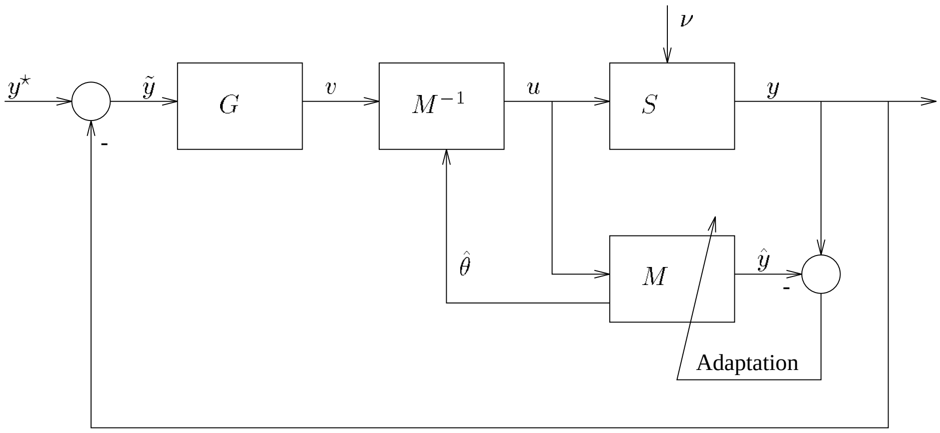


Figure 9.1: An overview of the control structure. $S$ is the system, $M$ is the model, $M^{-1}$ is the inverse model, where it is understood that the non-minimum-phase effects are not inverted, $\hat{\theta}$ is the model parameter estimate, and $G$ is the outer level controller.

**Proof:** We define

$$\tilde{Y}_j(t) = -e_j(t) - \varphi^T(t-d_j)\left(\hat{\theta}_j(t-d_j) - \hat{\theta}_j(t-1)\right) \tag{9.28}$$

and with this notation, from (9.26) we get the following equation for the tracking error

$$\tilde{y}_j(t) = M_j(q^{-1})\tilde{Y}_j(t) \tag{9.29}$$

Using the same arguments and notation as in the proof of Theorem 1, we get

$$\begin{aligned} n(t) &\leq H_1(q^{-1})||\tilde{y}(t+d)|| + H_2(q^{-1})||\nu(t+d)|| + K_3' \\ &\leq \sum_{j=1}^{m}\left(H_1(q^{-1})M_j(q^{-1})|\tilde{Y}_j(t+d_j)| + H_2(q^{-1})|\nu_j(t+d_j)|\right) + K_3' \end{aligned}$$

From (9.28) we get

$$\frac{\tilde{Y}_j(t)}{n(t-d_j)} = -\varepsilon_j(t) - \frac{\varphi^T(t-d_j)}{n(t-d_j)}\left(\hat{\theta}_j(t-d_j) - \hat{\theta}_j(t-1)\right) \tag{9.30}$$

From Theorem 2 it is evident that since $\mathcal{V} > \mathcal{V}_j = \overline{d}_j^0$, there exists a $T_j \geq 0$ such that for all $t \geq T_j$

$$\overline{d}_j^0 + |\mathcal{D}(\varepsilon_j(t))| + ||\tilde{\theta}_j(t-d_j) - \tilde{\theta}_j(t-1)|| \leq \mathcal{V}$$

because both $|\mathcal{D}(\varepsilon_j(t))| \to 0$ and $||\tilde{\theta}_j(t-d_j) - \tilde{\theta}_j(t-1)|| \to 0$ as $t \to \infty$. Hence, for $t \geq T_j$ it follows from (9.30) that $|\tilde{Y}_j(t)| \leq \mathcal{V}n(t-d_j)$. Together with (9.14), this gives for $t \geq T = \max(T_1, ..., T_m)$

$$\begin{aligned} n(t) &\leq \sum_{j=1}^{m} \left( H_1(q^{-1}) M_j(q^{-1}) \mathcal{V} n(t) + H_2(q^{-1}) \mathcal{V} n(t) \right) + K_3' \\ &\leq \mathcal{V} R(q^{-1}) n(t) + K_3' \end{aligned}$$

Using the same argument as in the proof of Theorem 1, we conclude that since

$$\mathcal{V} ||R(q^{-1})||_\infty \leq \kappa < 1$$

all signals in the closed loop are bounded, and for $t \geq T$ we have $n(t) \leq K_3'(1-\kappa)^{-1}$. Since $T$ is finite and the system does not have finite escape time, there exists a constant $K_3'' \geq K_3'$ such that $n(t) \leq K_3''/(1-\kappa)$ for all $t \geq 0$. The equation for the tracking error (9.29) together with (9.30) gives

$$\tilde{y}_j(t) = M_j(q^{-1}) \left( \varepsilon_j(t) + \frac{\varphi^T(t-d_j)}{n(t-d_j)} \left( \tilde{\theta}_j(t-d_j) - \tilde{\theta}_j(t-1) \right) \right) n(t-d_j)$$

From Lemma 1, we get

$$\sum_{\tau=0}^{t} \tilde{y}_j^2(\tau) \leq 2 ||M_j(q^{-1})||_\infty^2 \sum_{\tau=0}^{t} \left( \varepsilon_j^2(\tau) + ||\tilde{\theta}_j(\tau-d_j) - \tilde{\theta}_j(\tau-1)||^2 \right) n^2(\tau-d_j)$$

and

$$\sum_{\tau=0}^{t} ||\tilde{y}(\tau)||^2 \leq 2 \left( \frac{K_3''}{1-\kappa} \right)^2 \left( \frac{K_\rho}{1-\rho} \right)^2 \sum_{j=1}^{m} \sum_{\tau=0}^{t} \left( \varepsilon_j^2(\tau) + ||\tilde{\theta}_j(\tau-d_j) - \tilde{\theta}_j(t-1)||^2 \right)$$

Using Theorem 2, the conclusion of the theorem is proved.

**Remarks:** The bound on the average squared tracking error (9.27) consists of three terms. The first term $4K_4$ is due to the unstructured uncertainty, and will vanish as $\mathcal{V} \to 0$. The second term $K_5/t$ is due to parametric error, and will asymptotically go to zero as $t \to \infty$, since the estimator will asymptotically eliminate the effect of parametric uncertainty. The third term is due to exponentially decaying terms that have been neglected in the analysis. While $K_4$ bounds the controller's asymptotic performance, $K_5$ is a bound on the controller's transient performance. A major drawback of the analysis is that an explicit expression for $K_3''$ is not found. However, more explicit bounds may be found using the Bellman-Gronwall lemma (Desoer and Vidyasagar 1975), cf. Appendix A.

The result clearly shows that the initial parameter error is the most important cause for poor transient behavior, and while the result is global stability, it is clear from the discussion above that $||\tilde{\theta}(0)||$ must be small to avoid large transients. Persistence of excitation has not been assumed, nor has it been proved. However, we expect the transient performance to improve if the input is persistently exciting, since the effect of $||\tilde{\theta}(0)||$ will decay exponentially. Notice that a necessary condition for global persistence of excitation is that every operating regime is regularly visited, a condition that is unlikely to be fulfilled for many systems.

While the result is that the average tracking error is small, it is not possible to show that the tracking error is small at a certain instant in time. Hence, bursting phenomena may be present if the input is not locally persistently exciting.

□

An alternative controller for the decoupled and linearized system may be

$$v_j(t) = -G_j(q^{-1})\tilde{y}_j(t)$$

which does not require the future reference trajectory to be known. The analysis will be identical, except for a term $y_j^*(t)$ that will appear in the equation for the tracking error. If the reference sequence is slowly time-varying and $G_j(q^{-1})$ contains integral action, we get similar performance results in this case.

Of practical importance are methods for on-line solution of the implicit equation (9.24) for $u(t)$. The simplest case is when $z(t)$ does not explicitly depend on $u(t)$. Provided $\hat{B}_0(z(t),t)$ is non-singular, an expression for $u(t)$ can be found explicitly. Non-singularity can be enforced either through the specification of the convex parameter sets $\Theta_1, \Theta_2, ..., \Theta_m$, or by varying the gain $0 < \lambda_{min} \leq \lambda(t) \leq \lambda_{max}$ at each discrete time step $t$ such that non-singularity is ensured (Goodwin et al. 1980). An alternative to directly inverting $\hat{B}_0(z(t),t)$ when computing $u(t)$ is to use the Levenberg-Marquardt regularized inverse

$$\hat{B}_0^{-1}(z(t),t) \approx \left(\alpha I + \hat{B}_0^T(z(t),t)\hat{B}_0(z(t),t)\right)^{-1} \hat{B}_0^T(z(t),t)$$

where $\alpha > 0$ is a small regularization constant. Using this modification, it is not necessary to restrict the parameters or modify the parameter estimation algorithm to avoid singularity, at the cost of an approximate solution. In the events of $z(t)$ explicitly depending on $u(t)$, a numerical equation solver will usually be required. Should it be impossible to ensure a priori that there exists an exact solution of (9.24), an optimization approach can be taken. In that case, one can also include constraints such as actuator saturation, component failures, or constraints on the outputs into the optimization problem. It is also possible to include some weight on the control input and to handle input redundancy when $r > m$.

Usually, the operating regime based modeling approach gives a model structure that contains a significant number of unknown parameters. In general, a large number of parameters may lead to drift phenomena and bursting, since the model may be insensitive to some parameters, if the excitation is not persistent. However, this will not represent a serious problem for the proposed method. The reason is simply that the local model representation will naturally focus the attention on the few relevant parameters at any time, namely the parameters of the relevant local model(s) at the current operating point. The parameter estimation algorithm with thresholding will emphasize this focus. On the other hand, the number of unknown parameters can be reduced by using prior knowledge in terms of known local models, if available. It is a straightforward exercise to modify the algorithms presented here in such a way that local models with different structures are allowed. Also, the parameters of some local models may be fixed. This flexibility may give robust control structures with high performance, since one may choose simple but robust models in regimes when robustness is more important than performance, and more complex model structures with more parameters in regimes where performance is more important. Of course, one must take into account the available prior knowledge, and the excitation of the system.

## 9.6 Simulation Example: A $2 \times 2$ CSTR.

The theoretical analysis in the previous sections has shown that the model based control structure may give a stable and robust control system under certain assumptions. The most fundamental assumption is that the unstructured uncertainty must be small, i.e. neither the modeling error nor the disturbances must be too large. With a semi-realistic simulation example, we will illustrate the model development and justify our claim that the operating regime based modeling approach may lead to models that describe the behavior of the system sufficiently well for model based control applications.

The adaptive control structure is applied to a simulated exothermic continuous stirred tank reactor (CSTR) where a first order chemical reaction $A \rightarrow B$ takes place. The simulated system is described by the mass- and energy-balances

$$
\begin{aligned}
V \frac{d}{dt} c_A &= c_{Ai} q_i - c_A q_o - V r_A \\
\rho c_v V \frac{d}{dt} T &= \rho c_v T_i q_i - \rho c_v T q_o + Q - \Delta H_r V r_A \\
r_A &= k_0 c_A \exp\left( -\frac{E_A}{R} \left( \frac{1}{T} - \frac{1}{T_R} \right) \right)
\end{aligned}
$$

where the symbols are described in Table 9.1. In addition, the model contains dead-time caused by transportation delay and stirring dynamics.

The control objective is to track reference trajectories for the composition $c_A$ and temperature $T$, while rejecting disturbances which are due to variations in feed composition $c_{Ai}$. The temperature set-point $T^\star$ will typically vary between 360 and 385 $K$, while the composition set-point $c_A^\star$ is constant at 1.0 $mol/l$. We assume that $T$ and $c_A$ are measured on-line, and use the heat flow-rate $Q$ and feed flow rate-rate $q_i$ as control variables. The two-dimensional vectors $y(t)$ and $u(t)$ are appropriately scaled outputs and inputs. The system output is sampled at one-minute intervals, and the control input may be changed at the same rate. Furthermore, we assume ideal control of the volume, such that $V$ and $\rho$ are kept constant.

The operating point vector $z(t) = (T(t), q_i(t))^T$ is two-dimensional and contains the temperature and feed flow-rate. This choice is motivated by the well known facts that the reaction rate is non-linearly dependent on temperature, and the hold-up time depends on the flow-rate. Hence, the major non-linear phenomena are expected to be captured by this choice of operating point. The local model validity functions are illustrated in Fig. 9.2. It is a rule of thumb that an increase in temperature of about 10 $K$ will lead to an approximately doubled reaction rate

| Symbol | Value | Unit | Description |
|---|---|---|---|
| $T$ | | $K$ | Reactor temperature |
| $T_i$ | 310 | $K$ | Feed temperature |
| $T_R$ | 350 | $K$ | Reference |
| $c_A$ | | $mol/l$ | Concentration of $A$ in reactor |
| $c_{Ai}$ | | $mol/l$ | Concentration of $A$ in feed |
| $E_A$ | 70000 | $J/mol$ | Activation energy |
| $R$ | 8.314 | $J/(Kmol)$ | Gas constant |
| $k_0$ | 0.042 | $1/min$ | Frequency factor |
| $\rho$ | 1.0 | $kg/l$ | Ave. density in reactor |
| $c_v$ | 4000 | $J/(Kkg)$ | Ave. heat capacity in reactor |
| $V$ | 10000 | $l$ | Reactor volume |
| $\Delta H_r$ | -90000 | $J/mol$ | Reaction energy |
| $q_i$ | | $l/min$ | Feed flow-rate |
| $q_o$ | | $l/min$ | Outlet flow-rate |
| $Q$ | | $MW$ | Heat flow-rate (heat exchanger) |
| $\tau$ | 1.0 | $min$ | Dead-time |
| $T^\star$ | 360-385 | $K$ | Temperature set-point |
| $c_A^\star$ | 1.0 | $mol/l$ | Concentration set-point |

Table 9.1: Symbols.

for many chemical reactions (including this one) (Vogler 1986). On this basis, we choose the distance between each local model to be 20 $K$, since we expect this to give an approximation of the non-linearities that is sufficiently good for control purposes. The flow-rate $q_i$ and the hold-up time are inversely proportional. Hence, by considering the range the flow-rate is expected to vary within, it seems reasonable to distinguish between low and high flow-rates only. In other words, the operating regimes are designed using fuzzy sets representation of "high" and "low" feed flow-rate, and "high", "medium", and "low" temperature, cf. Fig. 9.2.

The model structure is based on six local ARX model structures of the form

$$\hat{y}(t+2|t) = A_{0,i}y(t) + B_{0,i}u(t) + B_{1,i}u(t-1) + C_i$$

Clearly, since the model structure is different from the simulated system, there will be modeling error. In principle, the modeling error can be reduced by a finer decomposition of the operating regimes, but this will lead to an increase in the number of parameters, which will introduce uncertainty in itself, cf. the bias/variance dilemma discussed in Chapter 6. However, we will see that although the chosen decomposition is quite rough, it is sufficient for our purpose, and it provides a good trade-off between model bias and variance. Observe that rather elementary process knowledge has been applied in order to find the model structure.

The implicit control equation is non-linear with respect to $u(t)$, and a numerical equation solver is used to solve (9.24) at each discrete time step. The control input from the previous time-step is used as the initial value in this algorithm. No convergence problems were observed. Since the system is open loop unstable ($\Delta H_r < 0$), it is important to have initial parameter estimates which give a

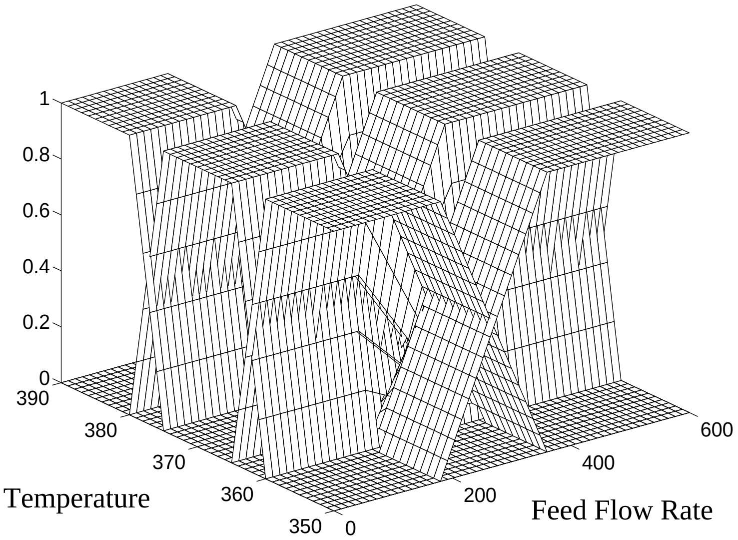


Figure 9.2: Local model validity functions used in simulation example.

controller that stabilizes the system. Initial estimates of the 84 unknown parameters in the model are found off-line using a 1000 time-steps data sequence and a least squares estimator. There are no restrictions on the parameter estimates, and we use the on-line estimator described in section 9.4. Since there is no problem with local model validity functions that asymptotically go to zero, cf. Fig. 9.2, no thresholding is applied. The gain is low, $\lambda = 0.05$, assuming there is no need for fast adaptation. On the basis of the linearized and decoupled nominal system $y_i(t) = q^{-2}v_i(t)$, we choose the outer level controller as simple integrators $G_j(q^{-1}) = 0.2(1-q^{-1})^{-1}$ which can be proved to give a fair trade-off between high bandwidth and robustness for this nominal system.

A simulation sequence with changing temperature set-point and disturbances in feed composition is shown in Fig. 9.3. For comparison, we also show results using two SISO PI-controllers

$$\begin{aligned} Q(t) &= K_Q \frac{1+\tau_i s}{\tau_i s} \left(T^\star(t) - T(t)\right) \\ q_i(t) &= K_q \frac{1+\tau_i s}{\tau_i s} \left(c_A^\star(t) - c_A(t)\right) \end{aligned}$$

where $\tau_i = 8\ min$, $K_Q = 0.25\ MW/K$, $K_q = 600\ l^2/(mol\ min)$, and $s$ is the Laplace operator. These controllers are well tuned in the sense that we have made a significant effort to find PI-parameters that simultaneously give fast response and the smallest possible overshoot. Despite this effort, the response with the PI controller is slower than the response with the model based controller, and the

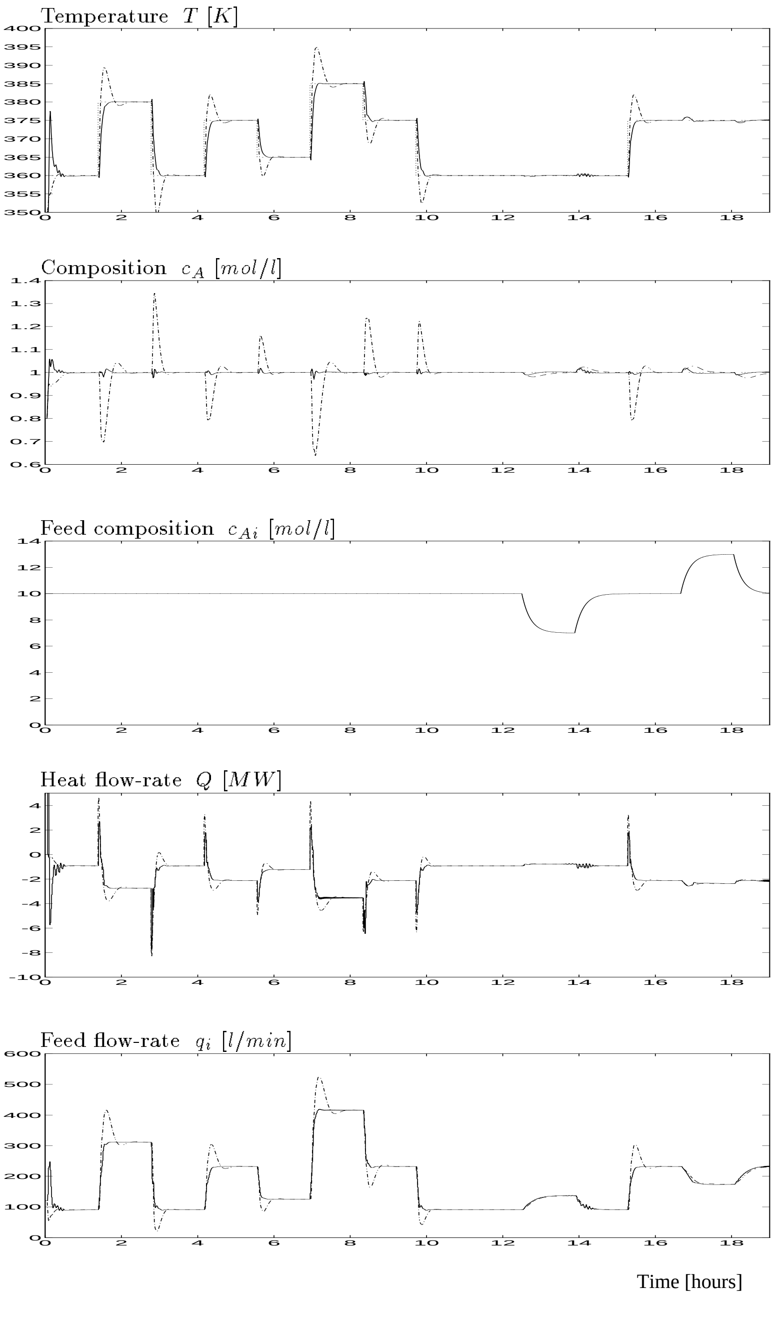


Figure 9.3: Simulation sequence. Dotted curves ($\cdots$) are set-points, solid curves ( — ) are simulation with adaptive controller, while dashed curves ($- \cdot - \cdot -$) are simulation with PI controller.

PI controller gives significant overshoot after set-point changes. Notice that the large overshoot is due to the combination of dead-time and system couplings, and cannot be avoided by re-tuning the PI controller. The model based controller is able to decouple the system, and therefore avoid overshoot. Both controllers are able to reject the disturbances in feed concentration, the adaptive controller being the somewhat better of the two. The adaptive model based controller is able to do this through the integral action in $G_j(q^{-1})$ and by adapting the parameter estimates. Since the estimator is slow compared to the control loop, the first effect is the most significant, but the second one is also important since the disturbance directly influences the dynamics of the system. The control authority of the two controllers are comparable, but the PI controller uses a little more control effort.

An alternative control structure using an operating regime based model applied to the same simulated system is reported in (Foss and Johansen 1992).

## 9.7 Discussion

In this chapter we have proposed an adaptive control structure based on local ARX models. Theoretical properties, like stability and robustness are investigated, and practical design trade-offs are briefly discussed. The theoretical analysis is based on reasonably weak assumptions on the system and model. The main assumptions are

- Invertibility of the model, and global exponential stability of the nominal inverse model. This excludes strong non-minimum-phase effects, and was made necessary by the particular control structure.
- The unstructured uncertainty is relatively small. This is a fundamental assumption.
- The nominal system is time-invariant. This assumption is made for convenience only, and can be relaxed to include slowly time-varying non-linear nominal systems, see Appendix A.

Unfortunately, the results are of a qualitative nature. This is due to both the generality of the setup, and the application of input/output stability theory. However, we believe that more quantitative, but somewhat weaker results can be found using Lyapunov theory, similar to Chen and Khalil (1991) and Polycarpou and Ioannou (1991).

The simulation example illustrates two points. First, the prior knowledge required for developing the model structure in this example is quite elementary and of a more qualitative character than what would be needed for the design of a mechanistic model structure. This is consistent with the results in Chapters 2 and 3. Second, the accuracy of the model is sufficient to give good performance when used in the model based controller.

The application of local models in the context of adaptive control is attractive because only *local* persistence of excitation is needed to avoid drift and bursting phenomena, rather than *global* persistence of excitation that may be needed with a non-local model representation.

# Chapter 10

# Conclusions

## 10.1 Operating Regime based Modeling

Some specific conclusions have been made in the discussion sections proceeding each chapter. The purpose of this section is to summarize and make some overall conclusions. This will lead to some advice regarding the applicability of the operating regime based modeling framework, and suggestions for future work.

### 10.1.1 Transparency

In this thesis we have analyzed a modeling framework based on the concept of *operating regimes*. This is appealing, since this is a concept that is well understood and applied by process engineers, and some operators and managers. Hence, the framework may support effective *interdisciplinary communication* about the system. The focus will be forced towards representation of knowledge related to the overall operation of the system in terms of operating regimes. Such knowledge is often qualitative and vague, and sometimes difficult to incorporate in other approaches. A major benefit of the operating regime concept is that it elevates the modeling problem to a more abstract level compared to purely equation based modeling. Of course, frameworks based on geometry/phenomena and some object oriented approaches, e.g. (Telnes 1992, Marquardt 1994), have this feature, too.

Transparency of the model is important because it simplifies both model development and application. In particular, validation, analysis, interpretation, and incorporation of prior knowledge benefits strongly from a transparent model. For example, we believe lack of transparency is one major reason why complex black-box model representations like neural nets have only had limited applicability to model-based control. The operating regime based modeling framework gives models that are reasonably transparent, and supports both the empirical and mechanistic modeling paradigms. The transparency is related to the possibility of interpreting the operating regimes in terms of either physical phenomena or different behaviors of the system. Moreover, it is related to our ability to interpret the local models independently, because the local models are typically simple enough to allow interpretation. This last point is not so important, as it is possible to linearize most models about interesting operating points to study their local behavior. The operating regime interpretation, however, is an inherent feature of the particular model representation.

### 10.1.2 Hybrid Modeling

The operating regime based modeling framework is flexible and supports hybrid modeling, in the sense that the different operating regimes or local models may have different

- levels of accuracy and detail,
- model representations,
- amounts and quality of relevant knowledge and data available,
- levels of transparency,
- computational algorithms associated with them, like estimators.

For large scale modeling problems, this flexibility may be useful. In particular, an essential part of hybrid modeling is the balanced combination of empirical data and prior knowledge in various forms.

The operating regime based modeling approach has found inspiration and applies methods from a number of different fields. At least superficially, the operating regime based modeling approach provides a unification of related modeling approaches that have arised in different fields, cf. Fig 10.1.

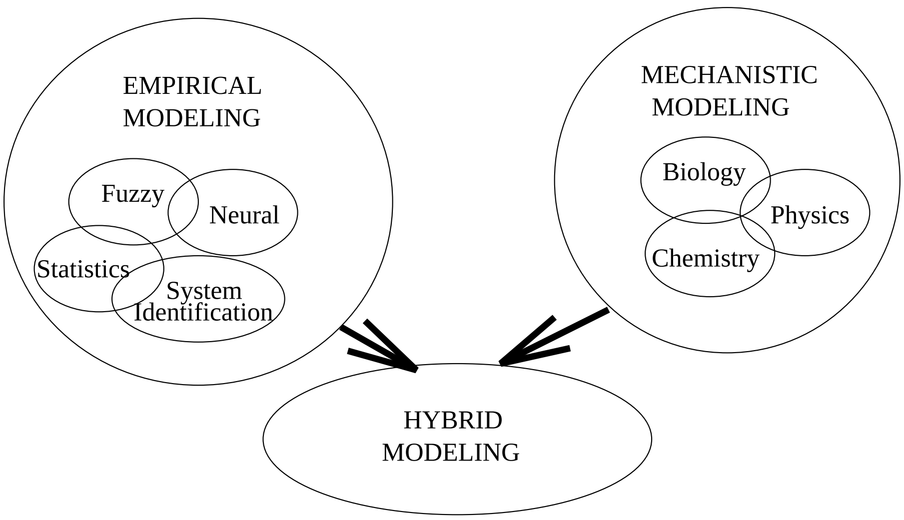


Figure 10.1: Elements of hybrid modeling.

### 10.1.3 Incremental and Iterative Modeling

Typically, process knowledge will improve with time, as experience increases, new phenomena are discovered, and more process data becomes available. Sometimes, such new knowledge is relevant only to particular operating conditions. If this is the case, the operating regime based modeling framework supports simple maintenance of the model, since one may only need to change one local model, or one may choose to decompose one operating regime into two or more sub-regimes. The remaining local models and operating regime can be left unchanged. In a sense, the operating regime based modeling approach can be thought of as modular, where the modules are the different operating regimes with local models.

Since some knowledge about the validity of each local model is explicit in the model representation, it is often possible to deduce operating conditions under which the system is not sufficiently well understood. Hence, experiment design support is directly available in the form of hints about which operating conditions experiments should be conducted under in order to improve the model. In addition, standard experiment design tools can be applied (Fedorov 1972, Box and Draper 1987).

The operating regime based modeling framework supports iterative model development. Consider for example the batch fermenter modeling example. One may start with a semi-empirical local linear model structure, cf. section 3.3.2, and gradually add more structure to both the local models and operating regimes, until one gets a semi-mechanistic model structure, cf. section 6.4.1. The identified local models provide new insight into the reaction kinetics, which may lead to a mechanistic kinetic model, and eventually a completely mechanistic process model.

### 10.1.4 Computer Aided Modeling

With examples and philosophical arguments we have explored some of the flexibility of the operating regime based modeling framework. The high flexibility is reflected by the fact that there are a large number of decisions to be made before a complete model is developed. The tasks are well defined and not too strongly interconnected. A typical modeling session would include iterations of the tasks

- Choice of model representation, e.g. state-space, input/output, distributed.
- Choice of which variables to apply to characterize the operating regimes.
- Design of experiments.
- Decomposition of the operating range into operating regimes.
- Development and identification of local model structures.
- Parameter identification.
- Model reduction.
- Model validation.

It is clear that some of these tasks are usually better performed by computers than engineers (e.g. parameter identification), and vice versa (e.g. choice of model representation). Other tasks like structure identification (choice of regimes and local model structures), experiment design, model reduction, and model validation are more suitably solved by cooperation between engineer and computer, where the emphasize on man or computer may differ considerably among different modeling problems. We therefore suggest that the operating regime based modeling framework described here should be implemented as a *computer aided modeling* environment. In particular, the computer should support the user with the computation of statistical properties and visualization of the data, knowledge, model structure and parameters in a flexible, interactive modeling environment, and leave it to the user to pursue the more interesting alternatives and make the important decisions.

### 10.1.5 Applicability

The operating regime based modeling framework is intended to support the design of non-linear models. While linear modeling is sufficient for solving many of todays control or optimization problems, non-linear modeling may sometimes be of interest when the system is operating within a wide range of operating conditions, for example batch processes, continuous processes during startup and shutdown, product shifts, maintenance and other exceptional operating conditions.

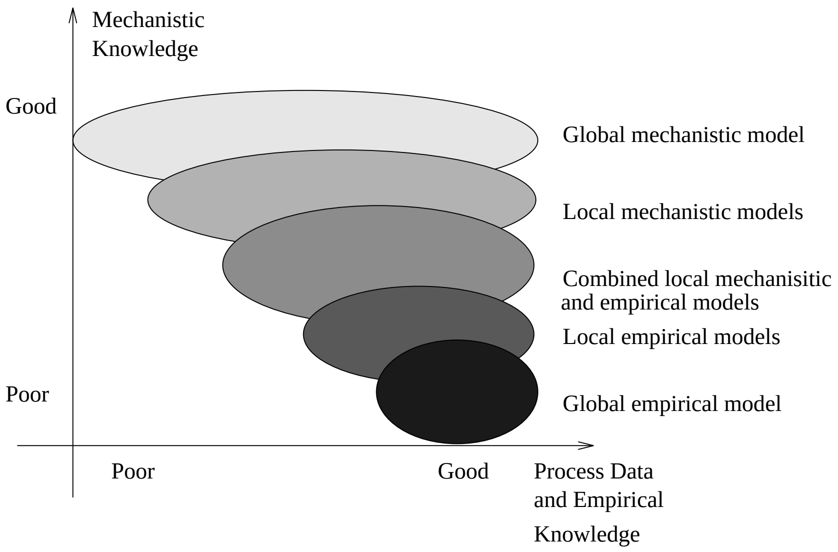


Figure 10.2: The figure illustrates which cases the operating regime based modeling framework is useful, as a function of the available mechanistic and empirical knowledge and data.

Fig. 10.2 illustrates our view on when the operating regime based modeling framework is useful, cf. Fig 1.1. The conditions are characterized by moderate amounts of both prior knowledge and empirical data, as opposed to the mechanistic modeling approach which requires very good prior knowledge, and the empirical modeling approach that requires large amounts of informative process data. The full range of knowledge levels in the gap between the empirical and mechanistic approaches can be addressed with the operating regime based modeling approach, as illustrated by the examples:

- **Local empirical models.** Fermenter in sections 3.3.2, 6.4.1, and 8, heat transfer process in section 2.4, hydraulic manipulator in section 6.4.3, and CSTR in section 9.6.
- **Mixed local empirical and mechanistic models.** pH neutralization example in section 6.4.2.
- **Local mechanistic models.** Fermenter in section 6.4.1.

Its flexibility combined with its transparency is the major advantage of this approach. For example, compared to a neural network based modeling approach, we have demonstrated that the operating regime based modeling approach has the same ability to fit empirical data, but due to its transparency we find it significantly more appealing. One the other hand, with a mechanistic local model structure where the unmodeled terms are modeled with simple empirical functions, the approach is very close to practical mechanistic modeling. It may be that the operating regime based modeling framework is not "the very best approach" for any given problem, but its flexibility makes it useful for a wide range of problems.

Recently, some industries have showed great interest in, and accepted certain model predictive control algorithms (Cutler and Ramaker 1979, Garcia et al. 1989). In general, these algorithms are based on linear empirical input/output models. In the future, the demand for improved plant operation may require controllers based on non-linear models to become standard, too. We believe that modeling and identification will be the major bottleneck in such control systems, and the success of empirical models for linear model predictive control makes it not unlikely that certain industries will insist on applying non-linear input/output models with a significant amount of empiricism. The operating regime based modeling approach discussed in this thesis is in our view an interesting approach for this purpose, cf. also Chapters 8 and 9.

### 10.1.6 Open questions

Although the results reported in this thesis are promising, several open questions remains:

- *How does the operating regime based modeling approach perform on complex and high-dimensional problems?* It may be that a significant amount of transparency is lost. Furthermore, prior knowledge may be available in forms that are not possible to utilize easily with this approach. These aspects are not covered by the simple examples in this thesis, nor is it convincingly demonstrated by the more complex examples in the literature, e.g. (Sugeno and Kang 1986), which are too incompletely documented to allow us to draw any conclusions.

- *Is this framework really easy to apply?* We have made a large effort to develop and understand this framework. This makes us expert users. The framework may possibly be less transparent and more difficult to apply for inexperienced users. We have observed that inexperienced users (through student projects and diploma theses) intuitively understand the idea and find it appealing. However, the interplay between the choice of local model structures and operating regimes seems to require some experience to understand completely.
- *Are the procedures and algorithms robust with respect to deficient and incomplete prior knowledge and data set?* The data sets we have applied here are either simulated or experimental data under rather ideal experimental conditions.
- *How should the structure of a computer aided modeling environment be?* We have only implemented some core algorithms. A software prototype has never been a partial goal for this thesis, and there remains a considerable effort before the structure of a computer aided modeling environment can be seen, in particularly related to man-machine interaction. It is important that the software environment does not reduce the transparency of the approach. Moreover, the environment must be flexible, yet so focused that it is easy to use and understand.

A complete validation of the framework, and an answer to the above questions, can only be accomplished through one or more tests on real modeling problems of industrial complexity. This is not within the scope of this thesis.

## 10.2 Empirical Modeling, System Identification, and Regularization

Even though the material in Chapter 7 is somewhat preliminary and not treated in as much depth as one might desire, we feel that it is an important contribution, and deserves some conclusions.

We have shown that the non-linear semi-empirical modeling problem can be addressed at a higher and more transparent level than the direct specification of a parameterized model structure. Prior knowledge in different forms can be applied to constrain the model. The method can be viewed in the context of regularization.

Regularization is a quite new tool in the empirical modeling tool-box, but has found wide applicability in solving inverse problems in a number of other fields in science and engineering (Tikhonov and Arsenin 1977, Nashed 1976, Sabatler 1987). With regularization, the number of parameters and the parameterization of the model, are of minor importance, as it is the model's effective number of parameters and the constraints on the parameter space introduced by the prior knowledge that are important. Hence, the underlying model structure is less important than the prior knowledge used in the regularization to constrain the parameter space. Of course, regularization can also be applied when a good parameterized model structure is known, in order to robustify the parameter identification problem. We have also seen that the modeling and identification problem can be approached in an infinite-dimensional function space rather than in a finite-dimensional parameter space. Although of limited practical applicability, this provides some fundamental insight into the interplay between prior knowledge and model structure.

It is interesting to observe that regularization can be implicit, like the singular value decomposition approach to matrix inversion, stopping an iterative optimization (Sjöberg and Ljung 1992, Ljung et al. 1993), and in local learning and smoothing algorithms (Hastie and Tibshirani 1990, Murray-Smith 1994*a*). Not understanding the effect of implicit regularization have led some researchers to criticize researchers in the neural network field for applying over-parameterized models. At least sometimes, this criticism has been unfair.

The practical limitations of regularization are mainly related to the additional computational complexity introduced. Various heuristics and short-cuts are necessary to approach high-dimensional problems with regularization. Another potential problem is the reduced transparency, as concepts like "effective number of parameters" and the effect of the constraints on the parameter space introduced by the prior knowledge may be difficult to grasp. However, we believe regularization has the potential to become a key tool in the non-linear system identification and semi-empirical modeling fields.

Structure identification has traditionally been given less attention than parameter identification. However, with more complex parametric model sets and computer power around, the structure identification problem has become more important, but certainly not simpler. Compared to todays practise, we feel that one should attempt to incorporate some more prior knowledge in both the structure and parameter identification criteria, in order to reduce the demand for empirical data and to improve the robustness with respect to data set deficiencies. The application of regularization to robustify the parameter identification problem is one possibility, and the bootstrap structure identification criteria introduced in Chapter 4 is a complementary method. However, several practical and theoretical issues remain unsolved.

## 10.3 Adaptive Control

This thesis contains some results on the stability and robustness of an adaptive control algorithm which applied to quite a large class of non-linear systems.

In Chapter 9, we apply a set of elementary analysis tools from input/output stability theory (some norm inequalities and the small gain theorem) to prove stability of an operating regime based adaptive controller. In Appendix A, a more general class of models is studied, using weighted $l_2$-norms and the Bellman-Gronwall lemma. The set of assumptions made is perhaps one of the least restrictive in the adaptive control literature. A continuous-time variant based on a non-linear state-space model representation is presented in (Johansen and Ioannou 1993). It is interesting to observe that the robust stability proofs are essentially the same as in the case of linear models. This is a feature of the input/output stability theory. Unfortunately, the results are even more qualitative, because the various bounds involved are potentially very conservative in the non-linear case. This is another feature of the input/output stability theory. Hence, there are no major practical significance of these results. They mainly serve as an illustration of the fundamental limitations and general design trade-offs.

# Appendix A

# Robust Adaptive Control of Slowly Time-varying Non-linear Systems

The application of input/output stability theory (Desoer and Vidyasagar 1975), and in particular weighted $L_2$-norms in the continuous-time case (Ioannou and Datta 1991, Tsakalis 1992), and weighted $l_2$-norms in the discrete time case (Datta 1993), has turned out to be attractive for the analysis of stability and robustness of adaptive control loops based on linear models.

The purpose of this appendix is to illustrate the application of weighted $l_2$-norms for analysis of adaptive controllers based on non-linear models. The control structure we have chosen is a simple feedback linearizing controller (Monaco and Normand-Cyrot 1987, Sastry and Isidori 1989, Nijmeijer and van der Schaft 1990) based on a non-linear discrete-time input/output model with slowly time-varying parameters. The parameter estimation algorithm applies normalization (Praly 1984) and a relative dead-zone (Kreisselmeier and Anderson 1986) to ensure robustness.

This appendix is organized as follows: First, we review the essentials of weighted $l_2$-norms. Then we present the model representation in section A.2, and the adaptive control structure in section A.3. The parameter estimation algorithm and the closed loop stability are analyzed in sections A.4 and A.5, respectively. The appendix ends with some discussion of the assumptions made, and some concluding remarks.

The material in this appendix is taken from (Johansen 1994$d$).

## A.1 Preliminaries

The Euclidean norm of a vector $x \in R^n$ is $||x|| = \sqrt{x^T x}$. The difference sequence $\Delta s$ of a sequence $s = (s(t))$ is defined by $\Delta s(t) = s(t) - s(t-1)$ for all $t \geq 0$. The truncation of a sequence $s$ at time $t$ is denoted $s_t$. The exponentially weighted $l_2$-norm of a sequence is defined by

$$||s_t||_2^\delta = \sqrt{\sum_{\tau=0}^{t} \delta^{t-\tau} ||s(\tau)||^2}$$

for any $\delta \in [0,1]$ and $t \geq 0$. The norm of a discrete-time transfer function $H(q)$, where $q$ is the one-step prediction operator and $H(q)$ has all poles strictly inside the unit disc, is defined by

$$||H(q)||_\infty = \sup_{\omega \in [0,2\pi]} |H\left(e^{j\omega}\right)|$$

If $H(q)$ is proper and has all poles on or inside the disc $|c| \leq \sqrt{\delta}$, then the exponentially weighted norm is defined by $||H(q)||_\infty^\delta = ||H(\sqrt{\delta}q)||_\infty$. The mixed notation $s_2(t) = H(q^{-1})s_1(t)$, where $H(q^{-1})$ is an exponentially stable transfer function, means the convolution $s_2(t) = h(t) \star s_1(t)$, where $h(t)$ is the inverse $z$-transform of the transfer function $H(q^{-1})$. The exponentially decaying term, which may appear as a consequence of non-zero initial conditions, is consistently neglected in this work, as it does not affect the stability analysis. The following lemmas will be used frequently:

**Lemma 1** *Let $s_2(t) = H(q^{-1})s_1(t)$, where $H(q^{-1})$ is a proper causal transfer function that has all poles in the disc $|c| \leq \sqrt{\delta}$, where $\delta \in (0,1]$, and suppose $\sum_{\tau=0}^{t} s_1^2(\tau) < \infty$. Then $||(s_2)_t||_2^\delta \leq ||H(q^{-1})||_\infty^\delta ||(s_1)_t||_2^\delta$ for all $t \geq 0$*

**Proof:** Given in (Datta 1993).

□

**Lemma 2** *Let $s_1$ be a positive sequence, i.e. $s_1(t) \geq 0$ for all $t$. Let the transfer function $H(q^{-1})$ be defined by $H(q^{-1}) = K(1 - \sigma q^{-1})^{-1}$ for some $K \geq 0$ and $\sigma \in [0,1)$. Suppose the sequence $s_2$ satisfy*

$$s_2(t) \leq H(q^{-1})\left(\gamma_0 s_1(t) + \gamma_1 s_1(t-1) + ... + \gamma_n s_1(t-n)\right)$$

*for some non-negative integer $n$ and constants $\gamma_0, ..., \gamma_n \geq 0$. Then,*

$$s_2(t) \leq H(q^{-1})\left(\gamma_0 + \gamma_1 \sigma^{-1} + ... + \gamma_n \sigma^{-n}\right) s_1(t)$$

**Proof:** Simple.

□

## A.2 Model Representation

We consider linearly parameterized non-linear discrete-time input/output models, written in the predictor form

$$y(t+d) = \varphi^T(y(t), ..., y(t-m), u(t), ..., u(t-r))\theta^\star(t) + \nu(t+d) \qquad \text{(A.1)}$$

with scalar input and output sequences $u$ and $y$, and $m$ and $r$ are non-negative integers. The integer $d \geq 1$ is the system's time-delay. Alternatively, it can be interpreted as the system's relative degree (Monaco and Normand-Cyrot 1987). The sequence $\nu$ is unstructured uncertainty, which contains unmodeled dynamics (modeling error and reduced order model), disturbances, noise, effects due to sampling of continuous-time signals etc. The function $\varphi$ is possibly non-linear, and the parameter sequence $\theta^\star$ may be slowly time-varying.

**AP.** There exists a constant $\underline{\theta} \geq 0$, such that $||\Delta\theta^\star(t)|| \leq \underline{\theta}$ for all $t \geq 0$. In addition, we know a convex and compact set $\Theta$ such that $\theta^\star(t) \in \Theta$ for all $t \geq 0$. $\square$

This means that the parametric variation from time $t-1$ to $t$ is bounded in norm by the small constant $\underline{\theta}$. Contrary to the time-invariant case, we need to assume that $\Theta$ is compact, in order to show boundedness of the parameter estimate. As a consequence, there exists a $\overline{\theta} \geq 0$ such that $\theta_1, \theta_2 \in \Theta$ implies $||\theta_1 - \theta_2|| \leq \overline{\theta}$.

**AB.** The sign of the high-frequency gain is uniformly constant and known. Without loss of generality, we assume here that it is positive:

$$\frac{\partial \varphi^T(y(t), ..., y(t-m), u(t), ..., u(t-r))\theta}{\partial u(t)} \geq \beta_0 \ > \ 0 \tag{A.2}$$

holds uniformly for all $\theta \in \Theta$, input sequences $u$, and output sequences $y$. It is assumed that $\varphi$ is a continuously differentiable function of $u(t)$. Moreover, $\varphi$ is bounded by an affine function, i.e. there exists a constant $C \geq 0$ such that

$$||\varphi(\cdot)|| \leq C(1 + |y(t)| + ... + |y(t-m)| + |u(t)| + ... + |u(t-r)|)$$

$\square$

AB contains a strong global controllability assumption. In the usual global controllability definition, it is required that any state can be reached from any initial state within an unspecified, but finite time (Nijmeijer and van der Schaft 1990). Our definition requires this time to be $d$, which is the shortest possible. On the other hand, we are only concerned about the output, not the full state.

**AR.** Let $y^\star$ be a bounded reference sequence that is known $d$ steps ahead in time. The bound is denoted $K^\star = \sup_{t \geq 0} |y^\star(t)|$. $\square$

The problem we address here is the one of asymptotically tracking this reference. In other words, if the tracking error is defined as $\tilde{y} = y^\star - y$, then the control objective is to make $|\tilde{y}(t)|$ as small as possible as $t \to \infty$.

## A.3 Adaptive Control Structure

With the above model representation, the adaptive certainty equivalence feedback linearizing controller is defined by the implicit equation

$$v(t) \ = \ \varphi^T(y(t), ..., y(t-m), u(t), ..., u(t-r))\hat{\theta}(t) \tag{A.3}$$

where $\hat{\theta}(t)$ is an estimate of $\theta^\star(t)$. From (A.2) it follows that there always exists a control input $u(t)$ that satisfies (A.3), and it is unique. The closed loop behavior is now described by

$$y(t+d) = v(t) + \varphi^T(t)(\theta^\star(t) - \hat{\theta}(t)) + \nu(t+d)$$

The following linear control algorithm is applied to control this nominally linear system

$$v(t) = y^\star(t+d) + G(q^{-1})(y^\star(t) - y(t)) \tag{A.4}$$

where $G(q^{-1}) = P(q^{-1})/Q(q^{-1})$. Now the closed loop satisfy

$$\tilde{y}(t) = M(q^{-1})\left(\varphi^T(t-d)(\hat{\theta}(t-d) - \theta^\star(t-d)) - \nu(t)\right) \tag{A.5}$$

where

$$M(q^{-1}) = \frac{1}{1 + G(q^{-d})q^{-d}} = \frac{Q(q^{-1})}{Q(q^{-1}) + q^{-d}P(q^{-1})}$$

The effect of the parametric and unstructured uncertainty in (A.5) will be analyzed in the remaining of this paper.

## A.4 Parameter Estimation

The estimate $\hat{\theta}(t)$ of $\theta^\star(t)$ is based on the predictor $\hat{y}(t|t-1) = \varphi^T(t-d)\hat{\theta}(t-1)$ which gives the prediction error

$$e(t) = \hat{y}(t|t-1) - y(t) = \varphi^T(t-d)\hat{\theta}(t-1) - \varphi^T(t-d)\theta^\star(t-d) - \nu(t) \tag{A.6}$$

and normalized prediction error $\varepsilon(t) = e(t)/n(t)$, where the scalar normalizing sequence $n$ is defined by

$$n^2(t) = \delta n^2(t-1) + 1 + ||\varphi(t-d)||^2$$

with $n(0) > 0$ and $\delta \in (0,1)$ given. The normalization is introduced to ensure that the normalized unstructured uncertainty is bounded (Praly 1984):

**AU.** The unstructured uncertainty satisfies $|\nu(t)| \leq \mathcal{V}n(t)$, where $\mathcal{V} \geq 0$. □

We apply a parameter estimation algorithm with normalization and relative dead-zone (Kreisselmeier and Anderson 1986):

$$\hat{\theta}'(t) = \hat{\theta}(t-1) - \lambda\frac{\varphi(t-d)}{1 + \varphi^T(t-d)\varphi(t-d)}n(t)\mathcal{D}(\varepsilon(t)) \tag{A.7}$$

$$\hat{\theta}(t) = \mathcal{P}_\Theta\left(\hat{\theta}'(t)\right) \tag{A.8}$$

$$\mathcal{D}(\varepsilon(t)) = \begin{cases} \varepsilon(t) + d_0 & \text{if } \varepsilon(t) < -d_0 \\ 0 & \text{if } |\varepsilon(t)| \leq d_0 \\ \varepsilon(t) - d_0 & \text{if } \varepsilon(t) > d_0 \end{cases} \tag{A.9}$$

where $\mathcal{P}_\Theta$ is a continuous parameter projection that projects its argument to the closest point (using Euclidean norm) in $\Theta$. The continuous function $\mathcal{D}$ is referred to as a dead-zone function, the constant $d_0 \geq 0$ is the dead-zone, and $\lambda > 0$ is the estimator gain. For convenience, the parameter error sequence is defined by $\tilde{\theta}(t) = \hat{\theta}(t) - \theta^\star(t)$. Next, we examine the properties of this algorithm:

**Lemma 3** *Suppose the outputs are generated by*

$$y(t+d) = \varphi^T(t)\theta^\star(t) + \nu(t+d)$$

*and consider the parameter estimation algorithm (A.7)-(A.9) with initial estimate* $\hat{\theta}(0) \in \Theta$. *If assumptions AP and AU hold, the gain satisfies* $\lambda \in (0,2)$, *and the dead-zone is chosen such that* $d_0 = \mathcal{V}$, *then the algorithm has the properties*

$$\begin{aligned} \sum_{\tau=0}^{t} \mathcal{D}^2(\varepsilon(\tau)) &\leq \left(\frac{2}{\lambda(2-\lambda)}\left((d\underline{\theta})^2 + 2(1+\lambda)d\underline{\theta}\overline{\theta}\right) + 2(d-1)^2\underline{\theta}^2\right)t \\ &\qquad + \frac{2}{\lambda(2-\lambda)}\overline{\theta}^2 \qquad \text{(A.10)} \\ |\varepsilon(t)| &\leq \mathcal{V} + |\mathcal{D}(\varepsilon(t))| \qquad \text{(A.11)} \\ \sum_{\tau=0}^{t} ||\Delta\hat{\theta}(\tau)||^2 &\leq \left(\frac{\lambda}{2-\lambda}\left((d\underline{\theta})^2 + 2(1+\lambda)d\underline{\theta}\overline{\theta}\right) + (d-1)^2\underline{\theta}^2 + 2\lambda(d-1)\underline{\theta}\overline{\theta}\right)t \\ &\qquad + \frac{\lambda}{2-\lambda}\overline{\theta}^2 \qquad \text{(A.12)} \end{aligned}$$

*for all* $t \geq 0$.

**Proof:** Rewriting the equation for the prediction error (A.6) gives

$$\varepsilon(t) = \frac{\varphi^T(t-d)\tilde{\theta}(t-1)}{n(t)} - \frac{\varphi^T(t-d)}{n(t)}\left(\theta^\star(t-d) - \theta^\star(t-1)\right) - \frac{\nu(t)}{n(t)}$$

Since $d_0 = \mathcal{V}$ there exists a sequence $\beta$ with the properties $0 \leq \beta(t) \leq 1$ and

$$\mathcal{D}(\varepsilon(t)) = \beta(t)\frac{\varphi^T(t-d)\tilde{\theta}(t-1)}{n(t)} - \beta(t)\frac{\varphi^T(t-d)}{n(t)}\left(\theta^\star(t-d) - \theta^\star(t-1)\right)$$

From (A.7) we find

$$\tilde{\theta}'(t) = \left(I - \lambda\beta(t)\frac{\varphi(t-d)\varphi^T(t-d)}{1+\varphi^T(t-d)\varphi(t-d)}\right)\tilde{\theta}(t-1) + \gamma(t) \qquad \text{(A.13)}$$

where

$$\gamma(t) = -\Delta\theta^\star(t) + \lambda\beta(t)\frac{\varphi(t-d)\varphi^T(t-d)}{1+\varphi^T(t-d)\varphi(t-d)}\left(\theta^\star(t-d) - \theta^\star(t-1)\right)$$

Defining the function $V(t) = ||\tilde{\theta}(t)||^2$, we find $V(t) \leq ||\tilde{\theta}'(t)||^2$ from the convexity of $\Theta$. Hence,

$$\begin{aligned} V(t) - V(t-1) \leq &\left(\gamma^T(t) + \tilde{\theta}^T(t-1)\left(I - \lambda\beta(t)\frac{\varphi(t-d)\varphi^T(t-d)}{1+\varphi^T(t-d)\varphi(t-d)}\right)\right) \\ &\cdot\left(\left(I - \lambda\beta(t)\frac{\varphi(t-d)\varphi^T(t-d)}{1+\varphi^T(t-d)\varphi(t-d)}\right)\tilde{\theta}(t-1) + \gamma(t)\right) - \tilde{\theta}^T(t-1)\tilde{\theta}(t-1) \end{aligned}$$

Assume $\beta(t) \neq 0$, then

$$
\begin{aligned}
V(t) - V(t-1) \leq\ & \gamma^T(t)\gamma(t) + 2\gamma^T(t)\left(I - \lambda\beta(t)\frac{\varphi(t-d)\varphi^T(t-d)}{1+\varphi^T(t-d)\varphi(t-d)}\right)\tilde{\theta}(t-1) \\
& - \left(\frac{2\lambda}{\beta(t)} - \lambda^2\right)\left(\frac{n^2(t)}{1+\varphi^T(t-d)\varphi(t-d)}\right) \\
& \cdot \left(\mathcal{D}(\varepsilon(t)) + \beta(t)\frac{\varphi^T(t-d)}{n(t)}\left(\theta^\star(t-d) - \theta^\star(t-1)\right)\right)^2
\end{aligned}
$$

or

$$
\begin{aligned}
& \left(\mathcal{D}(\varepsilon(t)) + \beta(t)\frac{\varphi^T(t-d)}{n(t)}\left(\theta^\star(t-d) - \theta^\star(t-1)\right)\right)^2 \leq \\
& \qquad \frac{\beta(t)}{\lambda(2-\beta(t)\lambda)}\left(||\gamma(t)||^2 + 2(1+\lambda)||\gamma(t)|| \cdot ||\tilde{\theta}(t-1)|| - \Delta V(t)\right) \qquad \text{(A.14)}
\end{aligned}
$$

If $\beta(t) = 0$, then $V(t) = V(t-1)$ and $\mathcal{D}(\varepsilon(t)) = 0$. Hence, from the properties of $\beta$ it follows that

$$
\begin{aligned}
\mathcal{D}^2(\varepsilon(t)) \leq\ & \frac{2}{\lambda(2-\lambda)}\left(||\gamma(t)||^2 + 2(1+\lambda)||\gamma(t)|| \cdot ||\tilde{\theta}(t-1)|| - \Delta V(t)\right) \\
& + 2||\theta^\star(t-d) - \theta^\star(t-1)||^2
\end{aligned}
$$

Using the fact that $||\tilde{\theta}(t)|| \leq \overline{\theta}$, which implies $V(t) \leq \overline{\theta}^2$ for all $t \geq 0$, together with the fact that AP implies $||\gamma(t)|| \leq d\underline{\theta}$, we get (A.10). By the definition of the dead-zone function, (A.11) follows. Finally, we get from (A.13)

$$
\begin{aligned}
\sum_{\tau=0}^{t} ||\Delta\hat{\theta}(\tau)||^2 \leq\ & \sum_{\tau=0}^{t} ||\hat{\theta}'(\tau) - \hat{\theta}(\tau-1)||^2 \\
\leq\ & \sum_{\tau=0}^{t} \Big( (\gamma(\tau) + \Delta\theta^\star(\tau))^T(\gamma(\tau) + \Delta\theta^\star(\tau)) - \\
& 2(\gamma(\tau) + \Delta\theta^\star(\tau))^T \lambda\beta(\tau)\frac{\varphi(\tau-d)\varphi^T(\tau-d)}{1+\varphi^T(\tau-d)\varphi(\tau-d)}\tilde{\theta}(\tau-1) + \\
& \lambda^2\beta^2(\tau)\tilde{\theta}^T(\tau-1)\frac{\varphi(\tau-d)\varphi^T(\tau-d)\varphi(\tau-d)\varphi^T(\tau-d)}{(1+\varphi^T(\tau-d)\varphi(\tau-d))^2}\tilde{\theta}(\tau-1)\Big) \\
\leq\ & \sum_{\tau=0}^{t} \Big( ||\gamma(\tau) + \Delta\theta^\star(\tau)||^2 + 2\lambda||\gamma(\tau) + \Delta\theta^\star(\tau)|| \cdot ||\tilde{\theta}(\tau-1)|| + \\
& \quad \lambda^2 \frac{n^2(\tau)}{1+\varphi^T(\tau-d)\varphi(\tau-d)} \\
& \quad \cdot \left(\mathcal{D}(\varepsilon(\tau)) + \beta(\tau)\frac{\varphi^T(\tau-d)}{n(\tau)}\left(\theta^\star(\tau-d) - \theta^\star(\tau-1)\right)\right)^2 \Big)
\end{aligned}
$$

Now, (A.12) follows from (A.14).

□

The relative dead-zone is a modification that turns the parameter estimator off when the normalized prediction error becomes small, and avoids therefore drift phenomena that otherwise might be excitated by the unstructured uncertainty. There are, however, several other modifications available, that leads to parameter estimation algorithms with similar properties and a robust adaptive control system. These include different variations of $\sigma$-modification, $\epsilon$-modification, and parameter projection (Narendra and Annaswamy 1989, Ioannou and Datta 1991, Datta 1993, Ydstie 1992).

## A.5 Closed Loop Stability

With the feedback linearizing control structure, it is well known that certain states may be unobservable (Byrnes and Isidori 1984, Monaco and Normand-Cyrot 1987). In order to ensure boundedness of these states, we must study the behavior of the inverse system. By the Implicit Function Theorem, e.g. (Nijmeijer and van der Schaft 1990), and assumption AB, it is evident that the inverse system is globally defined by a function $g$:

$$u(t) = g(u(t-1), ..., u(t-r), y(t+d), y(t), ..., y(t-m), \nu(t+d), \theta^{\star}(t)) \quad \text{(A.15)}$$

where $y$ and $\nu$ are viewed as inputs, and $u$ as the output.

**AI.** The inverse system (A.15) is globally uniformly exponentially stable, in the sense that its impulse response coefficients are bounded by an exponentially decaying sequence. □

Under this assumption, we can prove the following lemma, which essentially shows that the input is bounded by the output and the unstructured uncertainty:

**Lemma 4** *Suppose AB and AI hold, then there exist constants $\sigma \in [0,1)$ and $K_\sigma \geq 0$ such that for any $\kappa \in (\sigma, 1)$*

$$||(q^{-d}u)_t||_2^\kappa \leq \frac{K_\sigma}{1-\sqrt{\sigma/\kappa}}(||y_t||_2^\kappa + ||\nu_t||_2^\kappa)$$

**Proof.** It follows directly from AI that for some $K_\sigma \geq 0$

$$|u(t)| \leq K_\sigma \sum_{\tau=0}^{t} \sqrt{\sigma}^{t-\tau} \left(|y(\tau+d)| + |\nu(\tau+d)|\right)$$

the result follows from Lemma 1 together with $||(1-\sqrt{\sigma}q^{-1})^{-1}||_\infty^\kappa = (1-\sqrt{\sigma/\kappa})^{-1}$.

□

We are now in position to prove the main result:

**Theorem 1** *Suppose the system (A.1) is controlled by (A.3) and (A.4), and the algorithm (A.7)-(A.9) is applied to estimate the parameter sequence $\theta^\star$. Assume AB, AR, AP, AU and AI hold, and let $\rho \in [0,1)$ be a constant such that the impulse response coefficients of $M(q^{-1})$ decays faster than $\sqrt{\rho}^t$. Suppose the dead-zone is chosen as $d_0 = \mathcal{V}$. Let $\kappa$ be arbitrary in the open interval $\max(\delta, \sigma, \rho) < \kappa < 1$. If $\underline{\theta}$ and $\mathcal{V}$ are both sufficiently small, then for arbitrary initial conditions and $\hat{\theta}(0) \in \Theta$, all variables in the closed loop are bounded, and the average squared tracking error satisfies*

$$\frac{1}{t}\sum_{\tau=0}^{t} \tilde{y}^2(\tau) \le c_1 + \frac{1}{t}c_2 \tag{A.16}$$

*for all $t > 0$, where $c_2$ is proportional to $\overline{\theta}^2$, and $c_1 \to 0$ when $\underline{\theta} \to 0$ and $\mathcal{V} \to 0$.*

**Proof:** Let us define

$$\tilde{Y}(t) = \varphi^T(t-d)\tilde{\theta}(t-d) - \nu(t) \tag{A.17}$$

Combining (A.6) and (A.17), we get

$$\frac{\tilde{Y}(t)}{n(t)} = \varepsilon(t) + \frac{\varphi^T(t-d)}{n(t)}\left(\hat{\theta}(t-d) - \hat{\theta}(t-1)\right) \tag{A.18}$$

From the definition of the normalizing signal, assumption AB and Lemma 2, it is evident that there exists a constant $C$ such that

$$n^2(t) \le \frac{1}{1-\delta} + C^2 \sum_{\tau=0}^{t} \kappa^{t-\tau}\left(1 + |y(\tau-d)| + |u(\tau-d)|\right)^2$$

Using Lemma 4 and the triangle inequality, we find

$$n^2(t) \le \frac{1+3C^2}{1-\kappa} + 3C^2\left(\left|\left|\left(q^{-d}y\right)_t\right|\right|_2^\kappa\right)^2 + \frac{3C^2K_\sigma^2}{(1-\sqrt{\sigma/\kappa})^2}\left(||(|y|+|\nu|)_t||_2^\kappa\right)^2$$

because $\kappa > \sigma$. Using Lemma 2, we get

$$n^2(t) \le \frac{1+3C^2}{1-\kappa} + \left(K_y\,||y_t||_2^\kappa\right)^2 + \left(K_\nu\,||\nu_t||_2^\kappa\right)^2$$

where $K_y = \sqrt{3}C(\kappa^{-d/2} + \sqrt{2}K_\sigma(1-\sqrt{\sigma/\kappa})^{-1})$ and $K_\nu = \sqrt{6}CK_\sigma(1-\sqrt{\sigma/\kappa})^{-1}$. By the triangle inequality, the bound on $y^\star$, and (A.5), we get

$$n^2(t) \le K_0 + \left(K_y(1-\sqrt{\rho/\kappa})^{-1}||\tilde{Y}_t||_2^\kappa\right)^2 + \left(K_\nu||\nu_t||_2^\kappa\right)^2$$

where $K_0 = (1 + 3C^2 + 3C^2K_y^2(K^\star)^2)/(1-\kappa)$. This can be written

$$n^2(t) \le K_0 + K_y^2\frac{\kappa}{(\sqrt{\kappa}-\sqrt{\rho})^2}\sum_{\tau=0}^{t}\kappa^{t-\tau}\tilde{Y}^2(\tau) + K_\nu^2\sum_{\tau=0}^{t}\kappa^{t-\tau}\nu^2(\tau)$$

After some straightforward manipulations and using (A.18)

$$n^2(t) \le K_0 + K_y^2 \frac{\kappa}{(\sqrt{\kappa} - \sqrt{\rho})^2} \sum_{\tau=0}^{t} \kappa^{t-\tau} s^2(\tau) n^2(\tau) + K_\nu^2 \sum_{\tau=0}^{t} \kappa^{t-\tau} \nu^2(\tau)$$

where we have defined

$$s^2(t) = \left( \varepsilon(t) + \frac{\varphi^T(t-d)}{n(t)} \left( \hat{\theta}(t-d) - \hat{\theta}(t-1) \right) \right)^2 \tag{A.19}$$

It is straightforward to show that there exists a finite constant $B \ge 1$ such that

$$\frac{n^2(t+1)}{n^2(t)} \le B$$

for all $t \ge 0$. Hence

$$n^2(t) \le K_0 + \sum_{\tau=0}^{t-1} \kappa^{t-\tau} w^2(\tau) n^2(\tau) \tag{A.20}$$

where $w^2(t) = B\kappa^{-1} K_\nu^2 \mathcal{V}^2 + BK_y^2 (\sqrt{\kappa} - \sqrt{\rho})^{-2} s^2(t)$. By the Gronwall lemma, e.g. (Desoer and Vidyasagar 1975, Datta 1993), and the relation between arithmetic and geometric means, e.g. (Rudin 1986), we get from (A.20)

$$n^2(t) \le K_0 + K_0 \sum_{\tau=0}^{t-1} \kappa^{t-\tau} w^2(\tau) \left( 1 + \frac{\kappa^{-1}}{t-\tau-1} \sum_{i=\tau+1}^{t-1} w^2(i) \right)^{t-\tau-1}$$

From Lemma 3, it follows after some manipulation that

$$\sum_{\tau=0}^{t} w^2(\tau) \le \mu_0 t + \mu_1 \tag{A.21}$$

where

$$\begin{aligned} \mu_0 &= \left( B\kappa^{-1} K_\nu^2 + \frac{2BK_y^2}{(\sqrt{\kappa} - \sqrt{\rho})^2} \right) \mathcal{V}^2 + \frac{4Bd^2 K_y^2}{(\sqrt{\kappa} - \sqrt{\rho})^2} \left( \frac{3+\lambda}{\lambda(2-\lambda)} \right) \underline{\theta}^2 + \\ &\qquad \frac{2BdK_y^2}{(\sqrt{\kappa} - \sqrt{\rho})^2} \left( \frac{\lambda^3 + 4\lambda^2 + 4\lambda + 4}{\lambda(2-\lambda)} \right) \overline{\theta}\underline{\theta} \\ \mu_1 &= \frac{2BK_y^2}{(\sqrt{\kappa} - \sqrt{\rho})^2} \left( \frac{\lambda^2 + 2}{\lambda(2-\lambda)} \right) \overline{\theta}^2 \end{aligned} \tag{A.22}$$

Now,

$$n^2(t) \le K_0 + K_0 \sum_{\tau=0}^{t-1} \kappa^{t-\tau} w^2(\tau) (1 + \kappa^{-1}\mu_0)^{t-\tau-1} \left( 1 + \frac{\mu_1}{(\kappa + \mu_0)(t-\tau-1)} \right)^{t-\tau-1}$$

Since $(1+x^{-1})^x \leq e$ for all $x > 0$, it follows that

$$n^2(t) \leq K_0 + K_0(1+\kappa^{-1}\mu_0)^{-1} \exp\left(\frac{\mu_1}{\kappa+\mu_0}\right) \sum_{\tau=0}^{t-1} (\kappa+\mu_0)^{t-\tau} w^2(\tau)$$

Taking $\underline{\theta}$ and $\mathcal{V}$ sufficiently small, it follows from (A.22) that $\mu_0 + \kappa < 1$. Hence, (cf. Lemma 2.4 in (Datta 1993))

$$n^2(t) \leq K_0 + \frac{K_0(\mu_0+\mu_1)\exp\left(\frac{\mu_1}{\kappa+\mu_0}\right)}{(1-\kappa-\mu_0)(1+\kappa^{-1}\mu_0)} = \overline{n}^2$$

and it is proved that $n$ is a bounded sequence. From (A.5), Lemma 1, and (A.19) it follows that

$$\begin{aligned} \sum_{\tau=0}^{t} \tilde{y}^2(\tau) &\leq \frac{1}{(1-\sqrt{\rho})^2} \sum_{\tau=0}^{t} \tilde{Y}^2(\tau) \leq \frac{1}{(1-\sqrt{\rho})^2} \sum_{\tau=0}^{t} s^2(\tau) n^2(\tau) \\ &\leq \frac{2\overline{n}^2}{(1-\sqrt{\rho})^2} \sum_{\tau=0}^{t} \left(\varepsilon^2(\tau) + (d-1)^2 \left|\left|\Delta\hat{\theta}(\tau)\right|\right|^2\right) \end{aligned}$$

From Lemma 3, it is evident that (A.16) holds, and $c_1$ and $c_2$ have the stated properties.

□

## A.6 Discussion and Concluding Remarks

The contribution of this paper is a demonstration of the use of weighted $l_2$-norms for stability and robustness analysis of adaptive controllers based on non-linear models. Sufficient conditions for robust stability of the adaptive control loop are provided by Theorem 1. Stability requires the unstructured uncertainty $\mathcal{V}$ and parametric variations $\underline{\theta}$ both to be sufficiently small. Because the various bounds involved are potentially very conservative, the performance result (A.16) is conservative and should be interpreted in a qualitative way. The average squared tracking error will be bounded by a sum of two terms. The first term is the asymptotic performance bound that scales with the uncertainty and parameter variations. The second term is a bound on the transient performance that scales with the bound $\overline{\theta}^2$ on the parameter error, and vanishes asymptotically at the rate $1/t$. In addition, we have left out a third exponentially decaying term that is due to non-zero initial conditions.

It is clear that the model representation (A.1) contains two major limitations

1. Assumption AB requires $y(t+d)$ to depend strictly monotoneously on $u(t)$. Moreover, the model cannot contain actuator saturation. Typically, if this is violated, a solution to (A.3) will only exist for $v(t)$ (or $y^*(t)$) restricted to some subset of $R$. Then it may be necessary to modify the controller for

$v(t)$ to ensure that $v(t)$ remains within its prescribed subset, and analyze the effects of these modifications on the closed loop. Since these problems are of a more general nature and not related to the use of weighted $l_2$-norms, we have chosen to apply the restrictive assumption AB in order to simplify the presentation.

2. The predictor is linearly parameterized. This assumption can be relaxed to include non-linearly parameterized models

$$y(t+d) = f(y(t), ..., y(t-m), u(t), ..., u(t-r), \theta^\star(t)) + \nu(t+d) \quad (A.23)$$

where $f$ is a non-linear function that is twice continuously differentiable with respect to $\theta$. With the certainty equivalence controller

$$v(t) = f(y(t), ..., y(t-m), u(t), ..., u(t-r), \hat{\theta}(t))$$

we get using Taylor's theorem

$$\begin{aligned}\tilde{y}(t) = M(q^{-1}) \Bigg( & \frac{\partial f}{\partial \theta}(\cdot, \hat{\theta}(t-d))\tilde{\theta}(t-d) + \\ & \frac{1}{2}\tilde{\theta}(t-d)\frac{\partial^2 f}{\partial \theta^2}(\cdot, \overline{\theta}(t-d))\tilde{\theta}(t-d) + \nu(t)\Bigg)\end{aligned}$$

where $\overline{\theta}(t) = \alpha(t)\theta^\star(t) + (1-\alpha(t))\hat{\theta}(t)$ for some $\alpha(t) \in [0, 1]$. Defining

$$\begin{aligned}\varphi(t) &= \frac{\partial f}{\partial \theta}(\cdot, \hat{\theta}(t)) \\ \eta(t) &= \nu(t) + \frac{1}{2}\tilde{\theta}(t-d)\frac{\partial^2 f}{\partial \theta^2}(\cdot, \overline{\theta}(t-d))\tilde{\theta}(t-d)\end{aligned}$$

we get

$$\tilde{y}(t) = M(q^{-1})(\varphi^T(t-d)\tilde{\theta}(t-d) + \eta(t))$$

Under similar conditions as before, it is clear that there exists a constant $K \geq 0$ such that $\eta(t) \leq \mathcal{V}' n(t)$ where

$$\mathcal{V}' = \mathcal{V} + K\overline{\theta}^2$$

If $\underline{\theta}, \overline{\theta}$ and $\mathcal{V}$ are all sufficiently small, then Theorem 1 still holds. Hence, the price to pay for the relaxed assumption is that the parametric uncertainty must be sufficiently small.

Assumptions AP, AU, and AR are standard in the certainty equivalence adaptive control literature. The invertability condition AI is imposed by the particular control algorithm, while the bound on $\varphi$, cf. AB, are directly related to the use of input/output global stability theory arguments, see also (Chen and Khalil 1991) for a related local Lyapunov approach. While the affine boundedness appears to strongly restrict the class of mathematical systems (Sastry and Isidori 1989, Schwartz and Mareels 1992), it is clear from a more practical point of view that systems where the states do not diverge at a rate faster than exponential, will satisfy this assumption. We cannot think of any man-made physical system that does not satisfy this assumptions.

We have seen that the application of weighted $l_2$-norms gives a quite simple proof under rather weak and transparent assumptions. However, while the results are qualitatively appealing and transparent, their quantitative significance is doubtful. This is related to the general conservativeness of global input/output stability theory applied to non-linear systems.

# Bibliography


Adams, R. A. (1975), *Sobolev spaces*, Adademic Press, New York.

Akaike, H. (1969), 'Fitting autoregressive models for prediction', *Ann. Inst. Stat. Math.* **21**, 243–247.

Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE Trans. Automatic Control* **19**, 716–723.

Aksenova, T. I. (1989), 'Sufficient covergence conditions for external criteria for model selection', *Soviet J. Automation and Information Sciences* **22**, 49–53.

Albus, J. (1972), Theoretical and experimental aspects of a cerebellar model, PhD thesis, University of Maryland.

Aoyama, A. and Venkatasubramanian, V. (1993), Integrating neural networks with first-principles knowledge for bioreactor modeling and control, Paper 147i, 1993 Annual AIChE Meeting, November, St. Louis.

Bailey, J. E. and Ollis, D. F. (1986), *Biochemical Engineering Fundamentals*, McGraw-Hill, Singapore.

Balchen, J. G., Ljungquist, D. and Strand, S. (1992), 'State-space predictive control', *Chemical Engineering Science* **47**, 787–807.

Barron, A. R. (1993), 'Universal approximation bounds for superpositions of a sigmoidal function', *IEEE Trans. Information Theory* **39**, 930–945.

Bastin, G. and Dochain, D. (1988), Non linear adaptive control algorithms for fermentation processes, *in* 'Proc. American Control Conference, Atlanta', pp. 1124–1128.

Bellman, R. and Dreyfus, S. (1962), *Applied Dynamic Programming*, Princeton University Press.

Bellman, R. E. (1961*a*), *Adaptive Control Processes*, Princeton Univ. Press.

Bellman, R. E. (1961*b*), 'On the approximation of curves by line segment using dynamic programming', *Comm. Assoc. for Comp.*

Benveniste, A., Juditsky, A., Delyon, B., Zhang, Q. and Glorennec, P.-Y. (1994), Wavelets in identification, *in* 'Preprints 10th IFAC Symp. System Identification, Copenhagen', Vol. 2, pp. 27–48.

Bertero, M., De Mol, C. and Pike, E. R. (1985), 'Linear inverse problems with discrete data. I: General formulation and singular system analysis', *Inverse Problems* **1**, 301–330.

Bezdek, J. C., Coray, C., Gunderson, R. and Watson, J. (1981*a*), 'Detection and characterization of cluster substructure. II. fuzzy c-varities and complex combinations thereof', *SIAM J. Applied Mathematics* **40**, 352–372.

Bezdek, J. C., Coray, C., Gunderson, R. and Watson, J. (1981*b*), 'Detection and characterization of cluster substructure. I. Linear structure: Fuzzy c-lines', *SIAM J. Applied Mathematics* **40**, 339–357.

Billings, S. A. and Chen, S. (1989), 'Extended model set, global data and threshold model identification of severly non-linear systems', *Int. J. Control* **50**, 1897–1923.

Billings, S. A. and Voon, W. S. G. (1987), 'Piecewise linear identification of non-linear systems', *Int. J. Control* **46**, 215–235.

Bishop, C. (1991), 'Improving the generalization properties of radial basis function neural networks', *Neural Computation* **3**, 579–588.

Bohlin, T. (1989), The fundamentals of modelling and identification, Technical Report TRITA-REG-89/00002, Dept. Automatic Control, Royal Institute of Technology, Stockholm.

Bohlin, T. and Graebe, S. F. (1994), Issue in nonlinear stochastic grey-box identification, *in* 'Preprints IFAC Symposium on System Identification, Copenhagen', Vol. 3, pp. 213–218.

Box, G. E. P. and Draper, N. R. (1987), *Empirical Model-Building and Response Surfaces*, John Wiley & Sons, New York.

Box, G. E. P. and Hunter, W. G. (1965), 'The experimental study of physical mechanismns', *Technometrics* **7**, 23–42.

Box, G. E. P. and Jenkins, G. M. (1970), *Time series analysis: Forecasting and control*, Holden-Day, San Francisco, Ca.

Box, G. E. P. and Youle, P. V. (1955), 'The exploration and exploitation and response surfaces: An example of the link between the fitted surface and the basic mechansims of the system', *Biometrics* **11**, 287–323.

Breiman, L. (1993), 'Hinging hyperplanes for regression, classification, and function approximation', *IEEE Trans. Information Theory* **39**, 999–1013.

Breiman, L. and Meisel, W. S. (1976), 'General estimates of the intrinsic variability of data in nonlinear regression models', *J. American Statistical Association* **71**, 301–307.

Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984), *Classification and Regression Trees*, Wadsworths & Brooks, Monterey, Ca.

Broomhead, D. S. and Lowe, D. (1988), 'Multivariable functional interpolation and adaptive networks', *Complex Systems* **2**, 321–355.

Brown, R. H., Ruchti, T. L. and Feng, X. (1993), Artificial neural network identification of partially known dynamic nonlinear systems, *in* 'Proc. 32nd IEEE Conf. Decision and Control, San Antonio, TX', pp. 3694–3699.

Byrnes, C. I. and Isidori, A. (1984), A frequency domain philosophy for nonlinear systems, *in* 'Proceedings of the 1984 IEEE Conf. Decision and Control, Las Vegas, Nevada', pp. 1569–1573.

Carlin, M., Kavli, T. and Lillekjendlie, B. (1994), 'A comparison of four methods for non-linear data modeling', *Chemometrics and Intelligent Laboratory Systems* **23**, 163–178.

Carlstein, E. (1992), Resampling techniques for stationary time-series: Some recent developments, *in* D. Brillinger et al., ed., 'New Direction in Time Series Analysis. Part I', Springer-Verlag, New York, NY, pp. 75–85.

Chalmers, A. F. (1982), *What is this thing called science? 2nd Ed.*, Open University Press, Buckingham, UK.

Chen, F.-C. and Khalil, H. K. (1991), Adaptive control of nonlinear systems using neural networks - a dead-zone approach, *in* 'Proc. American Control Conference, Boston', pp. 667–672.

Chen, S. and Billings, S. A. (1989), 'Representation of non-linear systems: The NARMAX model', *Int. J. Control* **49**, 1013–1032.

Chen, S., Billings, S. A. and Grant, P. M. (1990*a*), 'Non-linear system identification using neural networks', *Int. J. Control* **51**, 1191–1214.

Chen, S., Billings, S. A., Cowan, C. F. N. and Grant, P. (1990*b*), 'Practical identification of NARMAX models using radial basis functions', *Int. J. Control* **52**, 1327–1350.

Craven, P. and Wahba, G. (1979), 'Smoothing noisy data with spline functions. Estimating the correct degree of smoothing by the method of generalized cross-validation', *Numerical Math.* **31**, 317–403.

Cutler, C. R. and Ramaker, B. L. (1979), Dynamic matrix control - a computer control algorithm, *in* 'AIChE 86th National Meeting, Houston'.

Cybenko, G. (1989), 'Approximations by superpositions of a sigmoidal function', *Mathematics of Control, Signals and Systems* **2**, 303–314.

Cybenko, G., Saarinen, S., Gray, R., Wu, Y. and Khrabrov, A. (1994), On the unreasonable effectiveness of memory-based methods, Preprint.

Cyrot-Normand, D. and Mien, H. D. V. (1980), Non-linear state-affine identification methods: Application to electrical power plants, *in* 'Proc. IFAC Symposium on Automatic Control in Power Generation, Distribution and Protection', pp. 449–462.

Datta, A. (1993), 'Robustness of discrete-time adaptive controllers: An input-output approach', *IEEE Trans. Automatic Control* **38**, 1852–1857.

De Veaux, R. D., Psichogios, D. C. and Ungar, L. H. (1993), A tale of two non-parametric estimation schemes: MARS and neural networks, *in* 'Proc. 4th Int. Conf. Artificial Intelligence and Statistics'.

Demuth, H. and Beale, M. (1993), *Neural Network Toolbox User's Guide (MATLAB)*, The MathWorks, Inc.

Desoer, C. A. and Vidyasagar, M. (1975), *Feedback Systems: Input-Output Properies*, Academic Press, New York.

Dorofeyuk, A. A., Kasavin, A. D. and Torgovitsky, I. S. (1970), Application of automatic classification methods to process identification in industry, *in* 'Preprints 2nd IFAC Symposium on Identification and Process Parameter Identification, Prague, Czechoslovakia', p. 4.1.

Doyle, J. (1978), 'Guaranteed margins for LQG regulators', *IEEE Trans. Automatic Control*.

Dyn, N. (1989), Interpolation and approximation by radial and related functions, *in* C. K. Chui, L. L. Schumaker and J. D. Ward, eds, 'Approximation Theory IV', Vol. 1, Academic Press, Inc., pp. 211–234.

Efron, B. and Tibshirani, R. (1986), 'Bootstrap methods for standard error, confidence intervals, and other measures of statistical accuracy', *Statistical Science* **1**, 54–77.

Farmer, J. D. and Sidorowich, J. J. (1987), 'Predicting chaotic time series', *Physical Review Letters* **59**(8), 845–848.

Fedorov, V. V. (1972), *Theory of Optimal Experiments*, Academic Press, New York.

Foss, B. A. and Johansen, T. A. (1992), Parallel nonlinear decoupling for process control – A NARMAX approach, *in* 'Preprints IFAC Symposium on AI in Real-Time Control, Delft, Holland', pp. 209–214.

Foss, B. A. and Johansen, T. A. (1993), On local and fuzzy modeling, *in* 'Proceedings of the 3rd Int. Conf. Industrial Fuzzy Control and Intelligent Systems, Houston, TX', pp. 80–87.

Foss, B. A., Johansen, T. A. and Sørensen, A. V. (1994), Nonlinear predictive control using local models - applied to a batch process, *in* 'Preprints IFAC Symp. Advanced Control of Chemical Processes (ADCHEM), Kyoto, Japan (also accepted for Control Engineering Practise)', pp. 225–230.

Freedman, D. (1984), 'On bootstrapping two-stage least-squares estimates in time series analysis', *The Annals of Statistics* **12**, 827–842.

Friedman, J. H. (1991), 'Multivariable adaptive regression splines (with discussion)', *The Annals of Statistics* **19**, 1–141.

Friedman, J. H. and Stuetzle, W. (1981), 'Projection pursuit regression', *J. American Statistical Association* **76**, 817–823.

Garcia, C. (1984), Quadratic/dynamic matrix control of nonlinear processes, *in* 'AIChE Annual Meeting, San Francisco'.

Garcia, C. E., Prett, D. M. and Morari, M. (1989), 'Model predictive control: Theory and practice - A survey', *Automatica* **25**, 335–348.

Gelb, A., ed. (1974), *Applied Optimal Estimation*, MIT Press, Cambridge.

Ghose, T. K. and Ghosh, P. (1976), 'Kinetic analysis of gluconic acid production by pseudomonas ovalis', *J. Applied Chemical Biotechnology* **26**, 768–777.

Gill, P., Murray, W. and Wright, M. (1981), *Practical optimization*, Academic Press, Inc.

Girosi, F., Jones, M. and Poggio, T. (1994), Priors, stabilizers and basis functions: From regularization to radial, tensor and additive splines, Technical Report AI Memo 1430, MIT, Cambridge.

Goldberg, D. E. (1989), *Genetic algorithms in search, optimization and machine learning*, Addison-Wesley.

Goodwin, G. C., Ramadge, P. J. and Caines, P. E. (1980), 'Discrete-time multivariable adaptive control', *IEEE Trans. Automatic Control* **25**, 449–456.

Grace, A. (1990), *Optimization Toolbox User's Guide (MATLAB)*, The MathWorks, Inc.

Granger, C. W. J. and Andersen, A. P. (1978), 'On the invertibility of time series models', *Stochastic Processes and their Application* **8**, 87–92.

Gustafsson, T. K. and Waller, K. V. (1983), 'Dynamic modeling and reaction invariant control of pH', *Chemical Engineering Science* **38**, 389–398.

Haber, R., Vajk, I. and Keviczky, L. (1982), Nonlinear system identification by "linear" systems having signal-dependent parameters, *in* 'Preprints 6th IFAC Symp. on Identification and System Parameter Identification, Washington D.C.', pp. 421–426.

Hall, R. C. and Seborg, D. E. (1989), Modelling and self-tuning control of a multivariable pH neutralization process. Part I: Modelling and multiloop control, *in* 'Proc. American Control Conference, Pittsburgh', Vol. 2, pp. 1822–1827.

Hastie, T. J. and Tibshirani, R. J. (1990), *Generalized Additive Models*, Chapman & Hall, London.

Hathaway, R. J. and Bezdek, J. C. (1993), 'Switching regression models and fuzzy clustering', *IEEE Trans. Fuzzy Systems* **1**, 195–204.

Hengjie, Z., Jianzhong, L., Shuqing, W. and Jicheng, W. (1989), Nonlinear feedback control of a fed-batch spriamycin fermentation process, *in* 'Proc. DYCORD+'89, Maastricht', pp. 155–160.

Hilaly, A., Karim, M. and Linden, J. (1994), 'Use of an extended Kalman filter and development of an automated system for xylose fermentation by recombinantescherichia coli', *J. Industrial Microbiology* **13**, 83–89.

Hilhorst, R. A. (1992), Supervisory Control of Mode-Switch Processes, PhD thesis, University of Twente - Electrical Engineering Department.

Hjalmarsson, H. (1993), Aspects on Incomplete Modeling in System Identification, PhD thesis, University of Linköping.

Huber, P. J. (1985), 'Projection pursuit (with discussion)', *The Annals of Statistics* **13**, 435–525.

Impe, J. V., Nicola, B., Vanrolleghem, P., Spriet, J., Moor, B. D. and Vandewalle, J. (1994), 'Optimal control of the Pencillin G fed-batch fermentation', *Optimal control appl. and meth.* **15**, 13–34.

Ioannou, P. and Datta, A. (1991), 'Robust adaptive control: A unified approach', *Proceedings of the IEEE* **79**, 1736–1768. Also in *Foundations of Adaptive Control*, P. V. Kokotović, Ed., Springer-Verlag, Berlin, 1991.

Isidori, A. (1989), *Nonlinear Control Systems, 2nd Ed.*, Springer Verlag, Berlin.

Ivakhnenko, A. G. and Yurachkovsky, Y. P. (1988), System structure identification by sets of observation data on the base of unbiasness principles, *in* 'Preprints IFAC Symposium on Identification and System Parameter Estimation, Beijing', pp. 953–965.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J. and Hinton, G. E. (1991), 'Adaptive mixtures of local experts', *Neural Computation* **3**, 79–87.

Janssen, P., Stoica, P., Söderström, T. and Eykhoff, P. (1988), Cross-validation ideas in model structure selection for multivariate systems, *in* 'Preprints 8th IFAC/IFORS symposium on Identification and System Parameter Estimation, Beijing, Aug. 27-31, 1988', pp. 961–965.

Jian, C. (1993), 'A predicting system based on combining an adaptive predictor and a knowledge base as applied to a blast furnace', *Journal of Forecasting* **12**, 93–102.

Joerding, W. H. and Meador, J. L. (1991), 'Encoding a prior information in feedforward networks', *Neural Networks* **4**, 847.

Johansen, T. A. (1994*a*), Adaptive control of MIMO nonlinear systems using local ARX models and interpolation, *in* 'Preprints IFAC Symp. Advanced Control of Chemical Processes (ADCHEM), Kyoto, Japan', pp. 159–166.

Johansen, T. A. (1994*b*), 'Fuzzy model based control: Stability, robustness, and performance issues', *IEEE Trans. Fuzzy Systems* **2**, 221–234.

Johansen, T. A. (1994*c*), Identification of non-linear systems using empirical data and prior knowledge - an optimization approach, Submitted to Automatica.

Johansen, T. A. (1994*d*), Weighted $l_2$-norms for analysis of an adaptive control loop based on a non-linear model, Submitted to IEEE Trans. Automatic Control.

Johansen, T. A. (1995), On the optimality of the Takagi-Sugeno-Kang fuzzy inference mechanism, Accepted for the 4th IEEE Conf. on Fuzzy Systems, Yokohama, Japan.

Johansen, T. A. and Foss, B. A. (1992*a*), 'A NARMAX model representation for adaptive control based on local models', *Modeling, Identification, and Control* **13**, 25–39.

Johansen, T. A. and Foss, B. A. (1992*b*), Nonlinear local model representation for adaptive systems, *in* 'Proc. IEEE Int. Conf. on Intelligent Control and Instrumentation, Singapore', Vol. 2, pp. 677–682.

Johansen, T. A. and Foss, B. A. (1992*c*), Representing and learning unmodeled dynamics with neural network memories, *in* 'Proceedings of the American Control Conference, Chicago, Il.', pp. 3037–3043.

Johansen, T. A. and Foss, B. A. (1993*a*), 'Constructing NARMAX models using ARMAX models', *Int. J. Control* **58**, 1125–1153.

Johansen, T. A. and Foss, B. A. (1993*b*), State-space modeling using operating regime decomposition and local models, *in* 'Preprints 12th IFAC World Congress, Sydney, Australia, 19-23 July. Extended paper in Technical Report 93-40-W, Department of Engineering Cybernetics, Norwegain Institute of Technology, Trondheim', Vol. 1, pp. 431–434.

Johansen, T. A. and Foss, B. A. (1994*a*), A dynamic modeling framework based on local models and interpolation - combining empirical and mechanistic knowledge and data, Submitted to Computers and Chemical Engineering.

Johansen, T. A. and Foss, B. A. (1994*b*), Identification of non-linear system structure and parameters using regime decomposition, *in* 'Preprints IFAC Symposium on System Identification, Copenhagen (also accepted for Automatica).', Vol. 1, pp. 131–136.

Johansen, T. A. and Foss, B. A. (1995), Empirical modeling of a heat transfer process using local models and interpolation, Submitted to the 1995 American Control Conference, Seattle, Wa.

Johansen, T. A. and Ioannou, P. A. (1993), Robust adaptive control of minimum phase nonlinear systems, Submitted to Int. J. Adaptive Control and Signal Processing.

Johansen, T. A. and Weyer, E. (1995), Model structure identification using separate validation data - asymptotic properties, Submitted to the European Control Conference, Rome.

Johnson, A. (1987), 'The control of fed-batch fermentation processes - A survey', *Automatica* **23**, 691–705.

Jones, R. D. and co-workers (1991), Nonlinear adaptive networks: A little theory, a few applications, Technical Report 91-273, Los Alamos National Lab., NM.

Jones, R. D., Lee, Y. C., Barnes, C. W., Flake, G. W., Lee, K., Lewis, P. S. and Qian, S. (1989), Function approximation and time series prediction with neural networks, Technical Report 90-21, Los Alamos National Lab., NM.

Jordan, M. I. and Jacobs, R. A. (1993), Hierarchical mixtures of experts and the EM algorithm, Technical Report 9301, MIT Computational Cognitive Science.

Jörgensen, S. B. and Jensen, N. (1989), Dynamics and control of chemical reactors - selectively surveyed, *in* 'Preprints DYCORD+ 89, Maastricht', pp. 359–371.

Kasavin, A. D. (1972), 'Adaptive piecewise approximation algorithms in the identification problem', *Automation and Remote Control* **33**, 2001–2006.

Kashyap, R. L. (1977), 'A bayesian comparison of different classes of dynamic models using empirical data', *IEEE Trans. Automatic Control* **5**, 715–727.

Kavli, T. (1990), Nonuniformly partitioned piecewise linear representation of continuous learned mappings, *in* 'Proceedings of IEEE Int. Workshop on Intelligent Motion Control, Istanbul', pp. 115–122.

Kavli, T. (1993), 'ASMOD -An algorithm for adaptive spline modelling of observation data', *Int. J. Control* **58**, 947–967.

Kavli, T. and Weyer, E. (1995), ASMOD – Some theoretical and experimental results, *in* K. Kunt, G. Irwin and K. Warwick, eds, 'Advances in Neural Networks for Control Systems', Springer-Verlag, Berlin.

Keulers, M. (1993), *Identification and Control of a Fed-Batch Process*, Dr. thesis, Technical University, Eindhoven.

Kirkpatrick, S., Gelatt Jr., C. D. and Vecchi, M. P. (1983), 'Optimization by simulated annealing', *Science* **220**, 671–680.

Kolmogoroff, A. (1936), 'Sulla theoria di Volterra della lotta per l'esistenza', *G. Istit. Ital. Degli Attuari* **7**, 74–80.

Konstantinov, K. and Yoshida, T. (1989), 'Physiological state control of fermentation processes', *Biotechnology and Bioengineering* **33**, 1145–1156.

Kortmann, M. and Unbehauen, H. (1988), Two algorithms for model structure determination of nonlinear dynamic systems with applications to industrial processes, *in* 'Preprints 8th IFAC/IFORS symposium on Identification and System Parameter Estimation, Beijing, Aug. 27-31, 1988', pp. 939–946.

Kramer, M. A., Thompson, M. L. and Phagat, P. M. (1992), Embedding theoretical models in neural networks, *in* 'Proc. American Control Conference, Chicago, Il.', pp. 475–479.

Kreisselmeier, G. and Anderson, B. D. O. (1986), 'Robust model reference adaptive control', *IEEE Trans. Automatic Control* **31**, 127–133.

Kreyszig, E. (1978), *Introductory Functional Analysis with Applications*, Krieger, Malabar, FL.

Lane, S. H., Handelman, D. A. and Gelfand, J. J. (1992), 'Theory and development of higher-order CMAC neural networks', *IEEE Control System Magazine* **12**(2), 23–30.

Larsen, J. (1992), A generalization error estimate for nonlinear systems, *in* 'Proc. IEEE Workshop on Neural Networks for Signal Processing, Piscataway, NJ', pp. 29–38.

Larsen, J. (1994), Generalization performance of regularized neural network models, *in* 'Proc. IEEE Workshop on Neural Networks for Signal Processing, Ermioni, Greece'.

Leontaritis, I. J. and Billings, S. A. (1985), 'Input-output parametric models for non-linear systems', *Int. J. Control* **41**, 303–344.

Lim, H. and Lee, K. (1991), Control of bioreactor systems, *in* H.-J. Rehm and G. Reed, eds, 'Biotechnology: Measuring, Modeling, and Control', Vol. 4, VCH, Weinheim.

Lindskog, P. and Ljung, L. (1994), Tools for semi-physical modeling, *in* 'Preprints IFAC Symposium on System Identification, Copenhagen', Vol. 3, pp. 237–242.

Ljung, L. (1978), 'Convergence ananlysis of parametric identification methods', *IEEE Trans. Automatic Control* **23**, 770–783.

Ljung, L. (1987), *System Identification: Theory for the User*, Prentice-Hall, Inc., Englewood Cliffs, NJ.

Ljung, L. (1994), 'System identification in a Modeling, Identification, and Control perspective', *Modeling, Identification, and Control* **15**, 153–160.

Ljung, L., Sjöberg, J. and McKelvey, T. (1993), On the use of regularization in system identification, *in* 'Preprints 12th IFAC World Congress, Sydney'.

Luenberger, D. G. (1969), *Optimization by Vector Space Methods*, John Wiley.

Luenberger, D. G. (1984), *Introduction to Linear and Nonlinear Programming, 2nd Ed.*, Addison-Wesley, Inc., Reading, MA.

Madych, W. R. and Nelson, S. A. (1990), 'Multivariate interpolation and conditionally positive definite functions. II', *Mathematics of Computation* **54**, 211–230.

Mallows, C. L. (1973), 'Some comments on $c_p$', *Technometrics* **15**, 661–667.

Marquardt, W. (1994), Trends in computer-aided process modeling, *in* 'Proc. Process Systems Engineering, Seoul, South Korea', pp. 1–24.

Mavrovouniotis, M. L. and Chang, S. (1992), 'Hierarchical neural networks', *Comp. Chem. Engr.* **16**, 347–369.

Næs, T. (1991), 'Multivariate calibration when data are split into subsets', *J. Chemometrics* **5**, 487–501.

Næs, T. and Isaksson, T. (1991), 'Splitting of calibration data by clustering analysis', *J. Chemometrics* **5**, 49–65.

Monaco, S. and Normand-Cyrot, D. (1987), Minimum-phase nonlinear discrete-time systems and feedback stabilization, *in* 'Proc. IEEE Conf. Decision and Control, Los Angeles, Ca.', pp. 979–986.

Moody, J. and Darken, C. J. (1989), 'Fast learning in networks of locally-tuned processing units', *Neural Computation* **1**, 281–294.

Mosca, E. (1970), System identification by reproducing kernel hilbert space methods, *in* 'Preprints 2nd IFAC Symposium on Identification and Process Parameter Estimation, Prague', p. Paper 1.1.

Murray-Smith, R. (1992), A fractal basis function neural net for modeling, *in* 'Proceedings Int. Conference on Automation, Robotics, and Computer Vision, Singapore', pp. NW–2.6.1 – NW–2.6.5.

Murray-Smith, R. (1994*a*), A Local Model Network Approach to Nonlinear Modelling, PhD thesis, The University of Strathclyde, Glasgow.

Murray-Smith, R. (1994*b*), Local model networks and local learning, *in* 'Fuzzy Duisburg', pp. 404–409.

Murray-Smith, R. and Gollee, H. (1994), A constructive learning algorithm for local model networks, *in* 'Proceedings of the IEEE Workshop on Compuer-Intensive Methods in Control and Signal Processing, Prague, Czech Republic', pp. 21–29.

Nakamori, Y. and Ryoke, M. (1994), 'Identification of fuzzy prediction models through hyperellipsoidal clustering', *IEEE Trans. Systems, Man, and Cybernetics* **24**, 1153–1173.

Nakamori, Y., Suzuki, K. and Yamanaka, T. (1992), A new design of a fuzzy model predictive control system for nonlinear processes, *in* T. Terano, M. Sugeno, M. Mukaidono and K. Shigemasu, eds, 'Fuzzy Engineering Toward Human Friendly Systems', IOS Press, Amsterdam, pp. 788–799.

Narendra, K. S. and Annaswamy, A. M. (1989), *Stable Adaptive Systems*, Prentice-Hall, Englewood Cliffs, NJ.

Nashed, M. Z., ed. (1976), *Generalized Inverses and Applications*, Academic Press, New York.

Nguyen, D. H. and Widrow, B. (1990), 'Neural networks for self-learning control systems', *IEEE Control Systems Magazine* **10**, 18–23.

Nijmeijer, H. and van der Schaft, A. J. (1990), *Nonlinear Dynamical Control Systems*, Springer-Verlag, New York.

Omohundro, S. M. (1987), 'Efficient algorithms with neural network behavior', *J. Complex Systems* **1**, 273–347.

Opoitsev, V. I. (1970), 'Identification of static plants by means of piecewise linear functions', *Automation and Remote Control* **31**, 809–815.

O'Sullivan, F. (1986), 'A statistical perspective on ill-posed inverse problems', *Statistical Science* **1**, 502–527.

Park, J. and Sandberg, I. W. (1993), 'Approximation and radial-basis-function networks', *Neural Computation* **5**, 305–316.

Parzen, E. (1962), 'On estimation of a probability density function and mode', *Ann. Meth. Stat.* **33**, 1065–1076.

Pearl, J. (1984), *Heuristics: Intelligent Search Strategies for Computer Problem Solving*, Addison-Wesley.

Peterson, T., Hernandez, E., Arkun, Y. and Schork, F. (1989), Nonlinear predictive control of a semi batch polymerization reactor by extended DMC, *in* 'Proc. American Control Conference, Pittsburg', pp. 1534–1539.

Pollard, D. (1984), *Convergence of Stochastic Processes*, Springer-Verlag, New York.

Polycarpou, M. M. and Ioannou, P. A. (1991), Identification and control of nonlinear systems using neural network models: Design and stability analysis, Technical Report 91-09-01, Dept. Electrical Enginering – Systems, University of Southern California, Los Angeles.

Polycarpou, M. M., Ioannou, P. A. and Ahmed-Zaid, F. (1992), Neural networks and on-line approximators for discrete-time nonlinear system identification, Preprint.

Pomerleau, Y., Perrier, M. and Dochain, D. (1989), Adaptive nonlinear control of the bakers' yeast fed-batch fermentation, *in* 'Proc. American Control Conference, Pittsburgh', pp. 2424–2429.

Pottmann, M., Unbehauen, H. and Seborg, D. E. (1993), 'Application of a general multi-model approach for identification of highly non-linear processes - A case study', *Int. J. Control* **57**, 97–120.

Powell, M. J. D. (1987), Radial basis function approximations to polynomials, *in* '12th Biennal Numerical Analysis Conference, Dundee', pp. 223–241.

Praly, L. (1984), Robust model reference adaptive controllers, Part I: Stability analysis, *in* 'Proc. IEEE Conf. Decision and Control, Las Vegas, Nevada', pp. 1009–1014.

Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. (1988), *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press.

Priestley, M. B. (1988), *Non-linear and Non-stationary Time Series Analysis*, Academic Press, London.

Proll, T. and Karim, M. (1994), 'Real-time design of an adaptive nonlinear predictive controller', *Int. J. Control* **59**, 863–889.

Psichogios, D. C. and Ungar, L. H. (1992), 'A hybrid neural network - first principles approach to process modeling', *AIChE J.* **38**, 1499–1511.

Rai, V. R. and Constantindes, A. (1973), 'Mathematical modeling and optimization of the gluconic acid fermentation', *AIChE Symposium Series* **69**, 114–122.

Rajbman, N. S., Dorofeyuk, A. A. and Kasavin, A. D. (1981), Identification of nonlinear processes by piecewise approximation, *in* P. Eykhoff, ed., 'Trends and Progress in System Ientification', Pergamon Press, Oxford, pp. 185–238.

Rawlings, J. B., Meadows, E. S. and Muske, K. R. (1994), Nonlinear model predictive control: A tutorial and survey, *in* 'Preprints IFAC Symposium ADCHEM, Kyoto, Japan', pp. 203–214.

Rescigno, A. and Richardson, I. W. (1973), The deterministic theory of population dynamics, *in* R. Rosen, ed., 'Foundations of Mathematical Biology', Vol. 3, Academic Press, New York, NY, pp. 283–360.

Rippin, D. W. T. (1989), Control of batch processes, *in* 'Proceedings DYCORD+ 89, August, Maastrict, The Netherlands', pp. 115–125.

Rissanen, J. (1978), 'Modeling by shortest data description', *Automatica* **14**, 465–471.

Rissanen, J. (1980), Consistent order-estimates of autoregressive processes by shortest decscription of data, *in* O. L. R. Jacobs, M. H. A. Davis, M. A. H. Dempster, C. J. Harris and P. C. Parks, eds, 'Analysis and Optimization of Stochastic Systems', Academic Press, London, pp. 451–461.

Rudin, W. (1986), *Real and Complex Analysis*, McGraw-Hill, New York.

Rugh, W. J. (1991), 'Analytical framework for gain scheduling', *IEEE Control Systems Magazine* **11**, 79–84.

Sabatler, P. C., ed. (1987), *Inverse Problems: An Interdisciplinary Study*, Academic Press, London.

Sanger, T. D. (1991), 'A tree-structured adaptive network for function approximation in high-dimensional spaces', *IEEE Transactions on Neural Networks* **2**, 285–293.

Sargantanis, J., Karim, M., Murphy, V. and Ryoo, D. (1993), 'Effect of operating conditions on solid substrate fermentation', *Biotech. and Bioengr.* **42**, 149–158.

Sastry, S. S. and Isidori, A. (1989), 'Adaptive control of linearizable systems', *IEEE Trans. Automatic Control* **34**, 1123–1131.

Sbarbaro, D. (1994), Context sensitive networks for modelling nonlinear dynamic systems, Preprint.

Schwartz, C. A. and Mareels, I. M. Y. (1992), 'Comments on "Adaptive control of linearizable systems"', *IEEE Trans. Automatic Control* **37**, 698–701.

Scott, G. M. and Ray, W. H. (1993), 'Creating efficient nonlinear neural network process models that allow model interpretation', *J. Process Control* **3**, 163–178.

Shamma, J. S. and Athans, M. (1990), 'Analysis of gain scheduled control for nonlinear plants', *IEEE Trans. Automatic Control* **35**, 898–907.

Shibata, R. (1976), 'Selection of the order of an autoregressive model by Akaike's information criterion', *Biometrica* **63**, 117–126.

Shorten, R. and Murray-Smith, R. (1994), On normalising radial basis function networks, *in* 'Proc. Irish Neural Network Conference, Dublin'.

Silverman, B. W. (1986), *Density Estimation for statistics and data analysis*, Chapman & Hall, London.

Sjöberg, J. and Ljung, L. (1992), Overtraining, regularization, and searching for minimum in neural networks, *in* 'Preprints IFAC Symposium on Adaptive Systems in Control and Signal Processing, Grenoble, France', pp. 669–674.

Sjöberg, J., Hjalmarsson, H. and Ljung, L. (1994), Neural networks in system identification, *in* 'Preprints 10th IFAC Symp. System Identification, Copenhagen', Vol. 2, pp. 49–72.

Skeppstedt, A., Ljung, L. and Millnert, M. (1992), 'Construction of composite models from observed data', *Int. J. Control* **55**, 141–152.

Sørheim, E. (1990), 'A combined network architecture using ART2 and back propagation for adaptive estimation of dynamical processes', *Modeling, Identification and Control* **11**, 191–199.

Söderman, U., Top, J. and Strömberg, J.-E. (1993), The conceptual side of mode switching, *in* 'Proc. IEEE Conf. Systems, Man, and Cybernetics, Le Touquet, France', pp. 245–250.

Söderström, T. and Stoica, P. (1988), *System Identification*, Prentice Hall, Englewood Cliffs, NJ.

Stepashko, V. S. (1988), 'Asymptotic properties of external criteria for model selection', *Soviet J. Automation and Information Sciences* **21**, 84–92.

Stoica, P., Eykhoff, P., Janssen, P. and Söderström, T. (1986), 'Model-structure selection by cross-validation', *Int. J. Control* **43**, 1841–1878.

Stokbro, K. and Umberger, D. K. (1990), Forecasting with weighted maps, *in* 'Proc. 1990 Workshop on Nonlinear Modeling and Forecasting, Santa Fe Institute'.

Stokbro, K., Hertz, J. A. and Umberger, D. K. (1990), 'Exploiting neurons with localized receptive fields to learn chaos', *J. Complex Systems* **4**, 603.

Stone, M. (1974), 'Cross-validatory choice and assessment of statistical predictions', *J. Royal Statistical Soc. B* **36**, 111–133.

Strömberg, J.-E., Gustafsson, F. and Ljung, L. (1991), Trees as black-box model structures for dynamical systems, *in* 'Proc. European Control Conference, Grenoble', pp. 1175–1180.

Su, H.-T., Bhat, N. and McAvoy, T. J. (1992), Integrated neural networks with first principles models for dynamic modeling, In Preprints IFAC DYCORD+ '92, College Park, Maryland.

Sugeno, M. and Kang, G. T. (1986), 'Fuzzy modelling and control of multilayer incinerator', *Fuzzy Sets and Systems* **18**, 329–346.

Sugeno, M. and Kang, G. T. (1988), 'Structure identification of fuzzy model', *Fuzzy Sets and Systems* **26**, 15–33.

Takagi, T. and Sugeno, M. (1985), 'Fuzzy identification of systems and its application to modeling and control', *IEEE Trans. Systems, Man, and Cybernetics* **15**, 116–132.

Tanaka, Y., Fukushima, M. and Ibaraki, T. (1988), 'A comparative study of several semi-infinite nonlinear programming algorithms', *European Journal of Operational Research* **36**, 92–100.

Telnes, K. (1992), Computer Aided Modeling of Dynamic Processes based on Elementary Physics, PhD thesis, The Norwegian Institute of Technology.

Thompson, M. L. and Kramer, M. A. (1994), 'Modeling chemical processes using prior knowledge and neural networks', *AIChE J.* **40**, 1328–1340.

Tikhonov, A. N. and Arsenin, V. Y. (1977), *Solutions of Ill-posed problems*, Winston, Washington DC.

Tong, H. and Lim, K. S. (1980), 'Threshold autoregression, limit cycles and cyclical data', *J. Royal Stat. Soc. B* **42**, 245–292.

Tsakalis, K. S. (1992), 'Robustness of model reference adaptive controllers: Input-Output properties', *IEEE Trans. Automatic Control* **37**, 556–565.

Tulleken, H. J. A. F. (1993), 'Grey-box modelling and identification using physical knowledge and bayesian techniques', *Automatica* **29**, 285–308.

Vapnik, V. (1982), *Estimation of Dependences based on Empirical Data*, Springer-Verlag, New York.

Vogler, H. S. (1986), *Elements of Chemical Reaction Engineering*, Prentice-Hall, Englewood Cliffs, NJ.

Volterra, V. (1931), *Lecons sur la théoris mathématique de la lutte pour la vie*, Gauthier-Villars, Paris.

Wahba, G. (1990), *Spline Models for Observational Data*, SIAM, Philadelphia.

Wang, L.-X. and Mendel, J. M. (1992), 'Fuzzy basis functions, universal approximation, and orthogonal least-squares learning', *IEEE Trans. Neural Networks* **3**, 807–814.

Weigend, A. S., Huberman, B. A. and Rumelhart, D. E. (1990), 'Predicting the future: A connectionist approach', *International Journal of Neural Systems* **1**, 193–209.

Weyer, E., Williamson, R. C. and Mareels, I. M. Y. (1994), System identification in the behavioral framework. Part II: Analysis, Submitted for publication.

Yager, R. R. and Filev, D. P. (1993), 'Unified structure and parameter identification of fuzzy models', *IEEE Trans. Systems, Man, and Cybernetics* **23**, 1198–1205.

Ydstie, B. E. (1992), 'Transient performance and robustness of direct adaptive control', *IEEE Trans. Automatic Control* **37**, 1091–1105.

Yoshinari, Y., Pedrycz, W. and Hirota, K. (1993), 'Construction of fuzzy models through clustering techniques', *Fuzzy Sets and Systems* **54**, 157–165.

Zadeh, L. A. (1965), 'Fuzzy sets', *Information and Control* **8**, 338–353.

Zames, G. (1966), 'On the input-output stability of time-varying nonlinear feedback systems. Part I: Conditions derived using concepts of loop gain, conicity, and positivity', *IEEE Trans. Automatic Control* **11**, 228–238.

Zhang, X.-C., Visala, A., Halme, A. and Linko, P. (1994), 'Functional state modelling approach for bioprocesses: Local models for aerobic yeast growth processes', *J. Process Control* **4**, 127–134.