On the Verifiability of the Activated Sludge System Dynamics

Ulf Jeppsson

Lund 1993

Department of
Industrial Electrical Engineering and
Automation (IEA)
Lund Institute of Technology
Box 118
S-221 00 LUND
Sweden

Published 1993
Printed in Sweden by KF Sigma

Official abstract page

Abstract

Wastewater treatment processes are inherently dynamic because of variations in the influent flow rate, concentration, and composition. The adaptive behaviour of the microorganisms further emphasizes this fact. Mathematical models and computer simulations are essential to describe, predict, and control the complicated interactions of the processes. Any attempt to model all details of the various reaction mechanisms are, however, destined to fail due to lacking knowledge and the extreme complexity required for such models.

A reduced order dynamic model for an activated sludge process performing carbonaceous removal, nitrification, and denitrification is presented herein. The identifiability of the model is investigated using both off-line and on-line methods and its dynamic behaviour is verified against simulations of a recognized model - the IAWPRC Activated Sludge Model No. 1.

The required data for the identification algorithms is based on directly measurable real time data. The simplified model may serve as a tool for predicting the dynamic behaviour of an activated sludge process since the parameters under varying operating conditions can be tracked on-line. The model is aimed for operation and control purposes as an integral part of a hierarchical control structure.

Abstract for the Novice

A model contains condensed knowledge of a physical process described by mathematical equations. The model may predict how the physical system will react under various conditions and is therefore an excellent tool for design, control, forecasting, education, etc. Though a model is seldom a perfect representation of a true system but more often a necessary simplification.

Depending on the purpose of the model, it may look quite different. For example, a combustion engine is an extremely

complex dynamic system. A *scientific* model which aims to get an insight into the intricate details of the process has to consider phenomena such as geometry of the cylinder, the mixing of air and fuel just as they meet the cylinder compartment, the chemical composition of the fuel, the propagation in time and space of the combustion, and the resulting movement of the piston. The timescale is from the millisecond range and upwards.

A model for the *control* of the air/fuel ratio will reflect a much more macroscopic view of the motor. Here the flow ratio of air to fuel has to be controlled close to the stochiometric relation. The spatial distribution of the combustion is not considered, only the mass flows of air and fuel. The timescale is no longer in the millisecond range, but rather 10-100 times longer. Finally, the *driver* requires another type of model for the motor. How the car acceleration responds to the throttle pedal becomes more important and the details of the combustion phenomena or of the air/fuel mixing process may be neglected.

The same reasoning holds for modelling of wastewater treatment plants. A number of scientific models have been proposed, but due to the high complexity they are not suited for control purposes. This work is aimed towards simplified models which only take the most important events into consideration. The situation is further complicated by factors like uncontrollable inputs to the process, many important quantities are not possible to measure accurately, and the various reactions have very different time constants. Moreover, the process (the activated sludge system) is a biological system and therefore adaptive to changing conditions. This means that the parameters of the models are not constants but have to be updated (identified and estimated) on a regular basis.

In order to check the results of the simplified models, they are tested (verified) against a scientific model to investigate whether they incorporate the same dynamic behaviour for the key variables or not. It is an aim to apply the reduced models for automatic control actions and thereby allowing the human operator at a wastewater treatment plant to focus on supervision and overall control without having to be concerned with the details of the processes and the low level reactions (like a driver of a car).

Contents

Preface

This thesis for the degree of Teknisk Licentiat[1] summarizes my current work on the subject of *Modelling and Control of the Activated Sludge Process* at the Department of Industrial Electrical Engineering and Automation (IEA) at Lund Institute of Technology. Parts of the work have already been accepted for publishing and presentation at the 6th IAWQ Workshop on Instrumentation, Control and Automation of Water and Wastewater Treatment and Transport Systems in Banff-Hamilton, Ontario, Canada, June 17-25, 1993 [Jeppsson and Olsson, 1993].

The following text contains numerous references to the *International Association on Water Pollution Research and Control* (IAWPRC). This organization was recently renamed and is today known as the *International Association on Water Quality* (IAWQ). The former name has been used throughout the thesis mainly for nostalgic reasons. It is *not* to be considered as a contribution to the ongoing debate on the appropriateness of the newly chosen name.

[1] A Swedish degree between M.Sc. and Ph.D., achieved after a minimum of two years of graduate studies.

Acknowledgements

I was introduced to the problems, possibilities and challenges of modern wastewater treatment by Professor Gustaf Olsson in 1989. During these four years he has shown an unlimited enthusiasm and support for my work and always an eager willingness to discuss new aspects and ideas of my research, also when his agenda was bursting - my sincere thanks. I would also like to thank all my fellow graduate students and the staff at the Department of Industrial Electrical Engineering and Automation (IEA) as well as my colleagues within the Swedish STAMP (Swedish abbreviation for *Control of Wastewater Treatment Plants*) research group, for their assistance.

For discussions, comments, and exchange of ideas and research material, I am also grateful to several persons abroad - especially Prof. Peter Dold, McMaster University, Hamilton, Ontario, Canada, Prof. (emer.) John Andrews, formerly at Rice University, Houston, Texas, USA, and Dr. Eduardo Ayesa, CEIT, San Sebastian, Spain.

This work was financially supported by the Swedish National Board for Industrial and Technical Development (NUTEK) under the STAMP program, contract no. 90-4882.

Ulf Jeppsson
Lund 1993

1

Introduction

Ever since Isaac Newton published his grand work *Mathematical Principles of Natural Philosophy* in 1687 where the fundamental laws of force and motion were formulated, the conclusion within the scientific community has been: *Nature has laws, and we can find them*. The importance of this statement can not be overestimated. It implies that every system, whether mechanical, electrical, biological or whatever, can be accurately modelled. Although proved wrong by the recently developed theory of chaos, the influence on the way scientists think has been enormous.

At the present time, models are usually the basis for understanding, controlling and predicting events and mechanisms in the world around us. Mathematical models describe everything from how electrons move and interact with other particles to how clusters of galaxies are formed in the universe. However, for a model to be really useful, a number of criteria exist which should be fulfilled. These criteria are related to the following areas:

- model validation;
- state and parameter identification;
- model verification.

First of all, a model should be properly *validated*, i.e. ensure that it provides results which are highly correlated with the true system and not only a fit of the specific data used for the modelling work. The conditions under which the discrepancy is such that a model can not be regarded as valid should also be clearly stated.

A set of methods for *identifying* and *estimating* the parameters of the model for calibration purposes should be given. The methods

may be be either mathematical or experimental, or a combination of both. If the model is mechanistic though, it is extremely important that the calibration process gives a unique solution for the different parameters, since a parameter often has a certain physical interpretation.

In order to *verify* the model, the state variables of it must in some way be comparable to measurable conditions (directly or indirectly) of the true process. This implies that the complexity of a model should be related to the amount of reliable measurements available from the physical process.

Unfortunately, for the activated sludge process a number of difficulties exists, which increase the problems when developing new mathematical models for it. First of all, results from experiments on full scale plants are seldom reproducable due to the high degree of site specific conditions. The input (flow rate and concentrations) is only to a small amount controllable and, moreover, highly variable. Historic data and weather forecasts may provide some possibilities of predicting flow rate variations and the sewer system or special equalization basins may be used as a buffer for smoothing these variations out, though only to a certain degree. Moreover, the entire system is time variant which implies that the dynamic behaviour is changing over time due to the fact that the microorganisms adapt to new environmental conditions. Parameters of a model must therefore also be updated but since available measurements from the activated sludge process are generally poor (especially the lack of on-line instruments), the situation is troublesome.

The complexity of wastewater treatment processes has increased over the years. In order to control and predict the behaviour of the plants, the appertunant mathematical models have also increased in complexity. The measurement problems have thus led to that some of the more complex models are not possible to verify. There are so many degrees of freedom that by adjusting the parameters of the model in one way, a good fit to experimental data can be achieved. A different set of parameters, however, may fit the experimental data equally well which implies that the models are too complex when compared to what can actually be measured. This is a serious problem since the models are highly mechanistic and the parameters have specific

physical interpretations which are used for comparing plants and to reach conclusions about their internal conditions and performance.

The purpose of this thesis is to point out the importance of concepts such as identifiability and verifiability when developing mathematical models, both generally and for the activated sludge process specifically. High model complexity does not serve a purpose of its own. It makes the intuitive understanding of the model more difficult and thereby hinders people from using it. The complexity should if possible coincide with what state variables and parameters are actually possible to measure in the real process. In this work a reduced order model for the activated sludge process is developed and results are compared to the IAWPRC 'Activated Sludge Model No. 1' [Henze et al., 1987a and 1987b] by means of computer simulations.

A simple example

Modelling of substrate uptake and cell growth in biological systems is highly complex as evidenced by the number of publications addressed to even the relatively simple biological system involving the uptake of glucose by yeast or red blood cells. A multiplicity of reactions and reaction mechanisms occur in even the simplest biological reaction. Adsorption, enzyme catalysis, and diffusional processes represent the major functional mechanisms which may control the uptake of a specific substrate. Furthermore, these mechanisms are dependent upon a number of physical, chemical, and biological variables within a given system.

The by far most common way of modelling how the growth rate of microorganisms (μ) depends on the substrate concentration (S) is by applying the Monod relationship[2] :

$$\mu(S) = \frac{\hat{\mu}\, S}{K_s + S} \qquad (1.1)$$

where: S = concentration of growth limiting substrate [mg/l];

[2] The notation used throughout this work coincide when possible to the notation proposed by the IAWPRC Task Group on Mathematical Modelling.

μ = maximum specific growth rate [day^{-1}];
K_S = substrate half-saturation coefficient [mg/l].

In a controlled laboratory environment in which no environmental conditions limit growth and a monoculture organism grows on a single substrate (only one specific carbon source), the Monod equation provides an adequate model. However, when applied to the activated sludge process, the situation is quite different.

First of all, the biomass is a mix of many types of organisms, each with its own set of growth parameters. The influent material consists of a variety of carbon, nitrogen and phosphorus compounds mixed with artifically manufactured chemical compounds and toxic materials, which all organism types react differently to. Conditions like the dissolved oxygen concentration, temperature, and pH may vary within a treatment plant. All have a significant influence on the parameters of the model, although different organisms may react very different to a certain change. These transient behaviours of the process will undoubtedly result in frequent changes of the rate controlling mechanisms. Moreover, none of the model parameters above can be measured directly. By investigating the oxygen uptake rate, substrate concentration changes, and dry mass of the biomass and combining it with other models (for the oxygen transfer rate, for the decay material, what reactions take place within the flocs, etc.) the model may be adjusted if the yield factor (Y) and decay rate (b) of the organisms are also known (or estimated simultaneously).

All difficulties mentioned above introduce uncertainties and errors in the model and every required measurement also adds noise to the calibration process. One way of dealing with this is to increase the complexity of the Monod model. For example, growth on n multiple substrates in which the different components exhibit a competitive inhibition effect on the utilization of the other components, may look like:

$$\mu(S) = \sum_{i=1}^{n} \frac{\hat{\mu}_i S_i}{K_{Si} + \sum_{j=1}^{n} a_{ij} S_j} \qquad (1.2)$$

where $a_{ii} = 1$ and a_{ij} represents the inhibition effect of the j'th substrate on the utilization of the i'th substrate by the organism.

This more complex structure would however only take into account the multiple substrate problem but all the other difficulties would still remain unsolved. It would also be practically impossible to calibrate such a model mainly due to the problem of measuring several different substrates accurately and it would only improve model performance under very special conditions.

The relevance of the basic structure of the empiric Monod equation is also a matter of dispute. A large number of rival models which exhibit practically the same behaviour have been suggested and investigated [Boyle and Berthouex, 1974 and Dochain and Bastin, 1984]. Some of these variants are,

the Haldane model:

$$\mu(S) = \frac{\hat{\mu} S}{K_s + S + \frac{S^2}{K_i}} \qquad (1.3)$$

the Tiessier model:

$$\mu(S) = \theta_1 (1 - e^{\theta_2 S}) \qquad (1.4)$$

the second order analog of the Tiessier model:

$$\mu(S) = \theta_1 \left(1 - \frac{1}{\theta_1 \theta_2 S + 1}\right) \qquad (1.5)$$

the Powell model:

$$\mu(S) = \theta_1 \left(\frac{\theta_2 + \theta_3 + S}{2\,\theta_3} \right) \left(1 - \sqrt{1 - \frac{4\,\theta_3\,S}{(\theta_2 + \theta_3 + S)^2}} \right) \tag{1.6}$$

the Blackman model:

$$\mu(S) = \begin{cases} \dfrac{\hat{\mu}\,S}{K_b} & \text{if } S < K_b \\ \hat{\mu} & \text{if } S \geq K_b \end{cases} \tag{1.7}$$

and the Contois model:

$$\mu(S, X) = \frac{\hat{\mu}\,S}{K_c\,X + S} \tag{1.8}$$

where: X = concentration of microorganisms [mg/l].

A simple comparison based on computer simulations of some of the alternative models and their behaviour is shown in Figure 1.1. The similarities are obvious and an identification based on real measurements would most likely show that it is impossible to determine which one provides the best fit (and certainly which one best mimics the true reaction mechanism), based on the commonly used statistical criteria.

It is more fruitful to realize that the Monod model and the similar models are rough simplifications of the true reaction mechanisms. They can be accepted as reasonably good tools for determining the average growth rate of various organisms under different load conditions. The model parameters must however be adjusted and updated as the environmental and physical conditions of the process change. But this gives rise to a new problem - even if the low complexity equation (1.1) is used. Its inherent structure makes it difficult to determine the optimum values of the model parameters. The model is quite sensitive to measurement noise but it is generally insensitive to the number of measured points. In certain regions, however, it is essential to obtain several data points of high accuracy (to estimate μ it is

important to have measurements for large substrate concentrations (S_{max}), while data from substrate concentrations close to $\mu(S_{max})/2$ have to be provided to identify K_S). The growth rate can only be determined by indirect methods and as a part of a larger model. Estimation algorithms therefore often show poor convergence properties and in many cases the formulation lack practical identifiability.

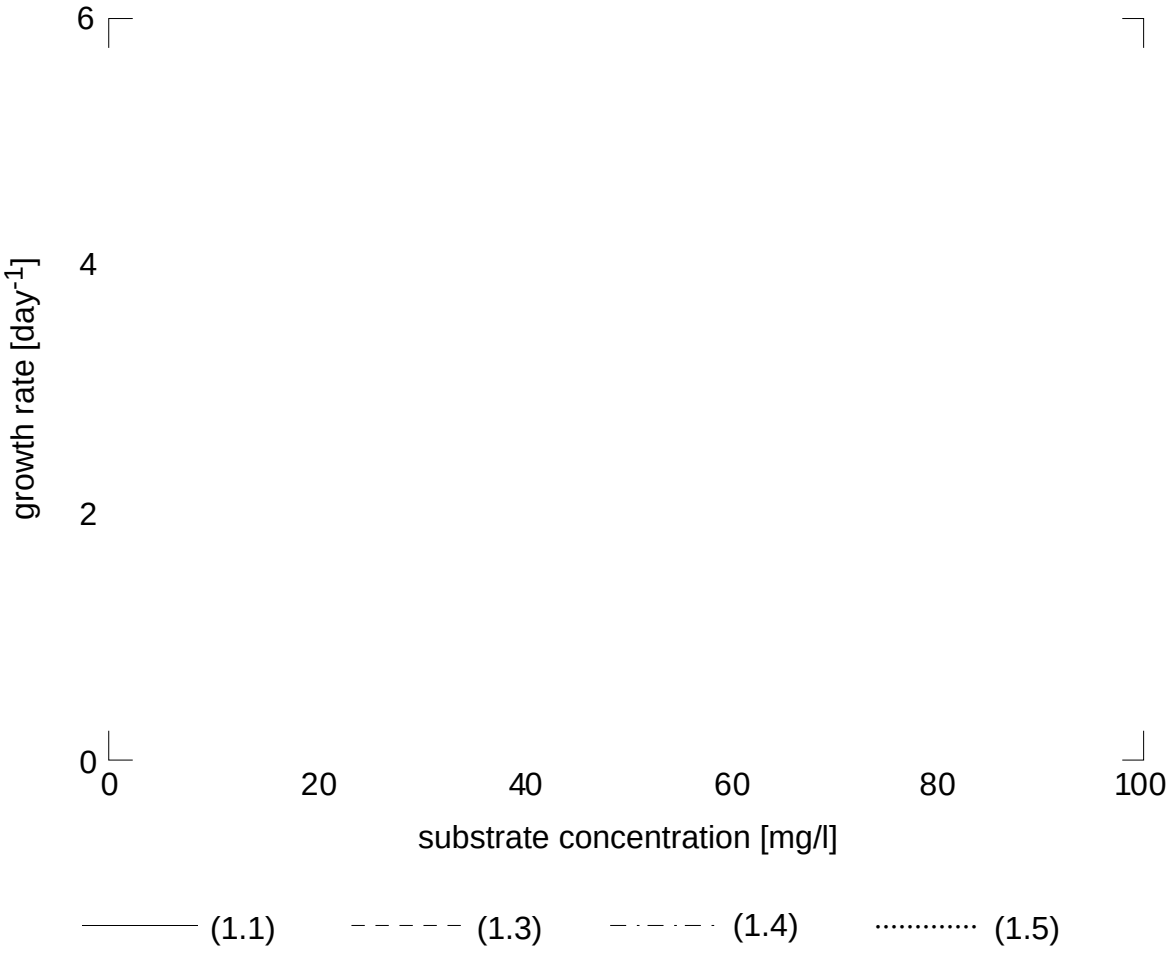


Figure 1.1 Simulations of the Monod (1.1), Haldane (1.3), Tiessier 1 (1.4), and Tiessier 2 (1.5) equations behaviour.

In Figure 1.2 it is illustrated how different sets of parameters for the Monod equation (1.1) produce similar results, in this case for quite low substrate concentrations (which is the normal operating condition for wastewater treatment plants receiving municipal wastewater). The same situation can be constructed for other regions of the graph. If a realistic amount of noise is

added to the system as well, an 'uncertainty deadband' - within which no guarantied better estimates can be found - occurs. A further investigation of this example is performed in Chapter 4.

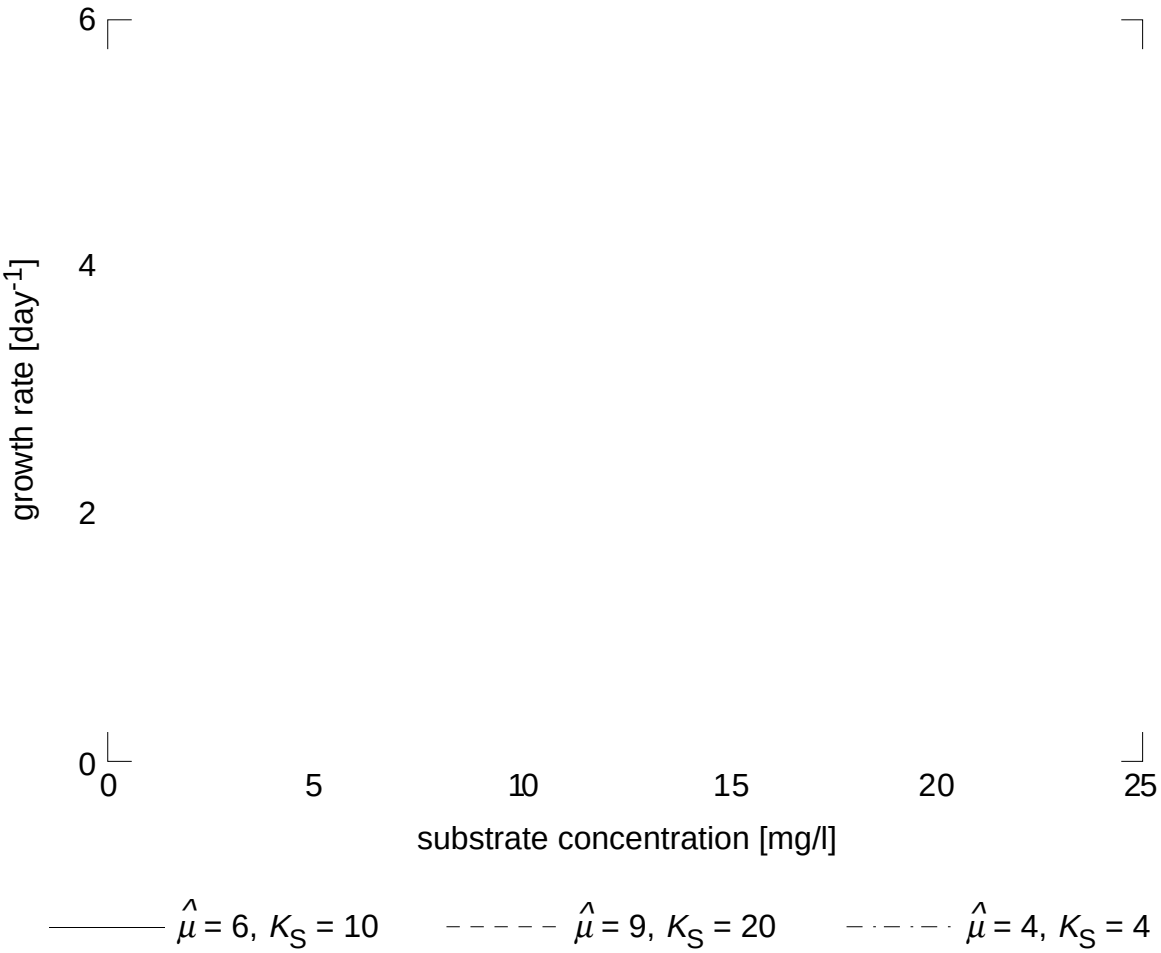


Figure 1.2 Simulations of the Monod equation (1.1) for different sets of model parameters.

This introduction is presented in order to point out some of the problems addressed in this work and motivate the need for simple, well structured models. The purpose of this discussion is not to reject the Monod equation, which has been used with success on many occasions; the purpose is to point out some general problems. These difficulties are naturally present - only much more so - in the traditional mathematical models for biological wastewater treatment. The Monod formulation is here merely a small part but due to the overall increased complexity the identifiability and verifiability difficulties are more severe.

Goals and contributions

The main goal for the work presented in this thesis is to combine knowledge of the activated sludge process dynamics with mathematical methods for estimation and identification as well as model reduction methods in order to obtain the simplest possible model capable of reasonably describing the carbonaceous and nitrogeneous activities at a wastewater treatment plant. The reduced model should preferably be globally identifiable from on-line measurements and thereby provide an appropriate basis for the future development of automatic control strategies applying feed forward and adaptive control algorithms. The reactions should be modelled in a physical reasonable manner whenever possible. This goal has to a large part been accomplished.

The last decade has seen a dramatic upheaval in the field of environmental protection. The immense environmental challenges facing the world in the beginning of the next millenium will require scientists in these areas to focus on better and more cost-effective solutions. The field of wastewater treatment is no exception. However, in the western world the majority of the wastewater treatment plants needed are now either operating or under construction. The dominant problems of this field have thus shifted from those of design and construction to those of plant operation. A better understanding of the dynamic behaviour of these plants and the use of control systems to convert unsatisfactory to satisfactory dynamic behaviour have significant potential for solving operational problems as well as reducing operational costs. Thus, the work of this thesis is motivated.

The results of this work are summarized in Chapter 6. The major contributions are believed to be:

- reduced order models aimed for operation of activated sludge systems together with suggestions for further refinement;
- a thorough analysis of the identifiability properties of the reduced models provided for a large number of cases using different methods;
- an approach for model reduction including the combination of

sophisticated mathematical algorithms with process knowledge of the activated sludge system dynamics;

- an extensive bibliography which summarizes much of the innovative work in the area of model reduction, identification, and verification of the activated sludge system dynamics;
- a base for operational models suitable for on-line parameter tracking and process diagnosis.

Outline of the thesis

An introduction to mathematical modelling of the activated sludge process is given in Chapter 2, by reviewing a number of biological reactor models developed over the last twenty years. Depending on what quantities can be measured and controlled, the influence on different state variables and parameters is discussed. In Chapter 3 the concepts of validity, identifiability, and verifiability of mathematical models are introduced. A short literature review of some of the earlier work in this field is also given here. Principles for reducing complex models are discussed in Chapter 4 and a simplified model of the activated sludge system is developed and examined. Ways of estimating state variables and parameters - which are not possible to measure directly - by mathematical means are also described and tested on the reduced order model. The simplified model is compared to a 'state-of-the-art' model for different operating conditions by means of computer simulations in Chapter 5 involving both off-line and on-line methods. A discussion of the advantages and drawbacks of the different approaches is given. Chapter 6 contains some final conclusions.

A guide for the reader

Due to the fact that modelling of wastewater treatment processes requires knowledge from several different scientific areas - not possible to describe in detail in this thesis - some references are given where the interested reader may improve his knowledge in fields not usually encountered.
A comprehensive guide to the principles of biological wastewater treatment is given in [Grady and Lim, 1980]. Hundreds of

references and suggestions to further reading are also presented there. Another excellent book describing the same area is [Henze et al., 1992]. Both books incorporate numerous solved examples where the reader may verify his (or her) new knowledge.

A number of good books describe the fundamentals of mathematical modelling, process dynamics, and automatic control in general, for example [Seborg et al., 1989, Åström and Wittenmark, 1990, and Kuo, 1991]. All these aspects are of great importance to understand the intentions of this thesis.

Unfortunately, a great deal of mathematical knowledge is required of the reader to fully understand the three books above. For the novice seeking a basic understanding of system dynamics and control methods a better choice is [Olsson and Piani, 1992]. Several examples from wastewater treatment plants are used to illustrate these areas.

Mathematical methods for estimating and identifying state variables and parameters which are not directly measurable, as well as determining the optimal set of model parameters for a specific set of data is thoroughly described in [Ljung, 1987, Söderström and Stoica, 1989 and Fletcher, 1987]. These methods are, however, complicated and need not be understood in detail.

2

Biological Reactor Models

This chapter reviews a number of different models and modelling approaches for activated sludge wastewater treatment plants. The investigated is restricted to biological reactor models performing carbon oxidation, nitrification, denitrification and to some extent biological phosphorus removal. Firstly, a number of fundamental aspects concerning mathematical modelling is discussed. A commented literature review is given in section 2.2 and a further examination of the most commonly used state variables and parameters which influence the models, is performed in sections 2.3 and 2.4. Finally, the possibilities of measuring and manipulating different state variables and process parameters are commented in the last two sections.

2.1 Modelling - Purpose and Problems

During the last twenty years the complexity of biological waste water treatment processes has increased dramatically from simple plug flow reactors performing carbon oxidation, to plants with advanced flow schemes combined with nitrification-denitrification processes and recently even biological phosphorus removal. These new process developments have created an increasing demand for reliable, yet simple and understandable mathematical models, able to describe not only the stationary but also the dynamic behaviour of the activated sludge system.

Mathematical models are powerful tools by which a great deal of insight and knowledge about the processes can be gained as well as being an important instrument for practical operations at a modern plant. However, before starting the development of a new

model it is extremely important to clearly state the *purpose* of the model since this is most likely to influence its structure. Some of the more general purposes are listed below.

- Design - models allow the exploration of the impact of changing system parameters and development of plants designed to meet the desired process objectives at minimal cost.
- Research - models serve as a tool to build and test hypotheses and thereby gaining new knowledge about the processes.
- Process control - models allow the development of new control strategies through the possibilty of investigating the system response to a wide range of inputs without endangering the actual plant.
- Forecasting - models are used to predict future plant performance when exposed to foreseen input changes and provide a framework for testing appropriate counteractions.
- Performance analysis - models allow analysis of total plant performance over time when compared to laws and regulations and what the impact of new effluent requirements on plant design and operational costs will be.
- Education - models provide students with a tool to actively explore new ideas and improve the learning process as well as allowing plant operators constant training and thereby increasing their ability to handle unforseen situations.

Another decision which often coincides with the purpose of the model, is to determine who should be the end user. Operators, design engineers, process engineers, plant managers, researchers - all have different goals and reasons for using a model, leading to an abundance of more or less suitable model structures.

Mathematical models may serve many purposes but they can also be of different *types*. At one extreme, the models are highly mechanistic, linking the inputs and the outputs together through rate equations which seek to mimic the true reaction mechanisms. The equations are then glued together through mass balance equations. The mechanistic model serves as an excellent tool for researchers trying to understand the events

occuring in the real system as well as for educational purposes. At the other extreme, the models are highly empirical (time-series models). Some experiments are performed on the system and a model is then fitted to the recorded data by assigning suitable numerical values to its parameters. Numerous methods for this type of system identification are given in [Söderström and Stoica, 1989 and Ljung, 1987]. The main disadvantages of these 'black-box' models are their limited validity, they provide little physical insight of the process, and some of the important variables may be difficult or impossible to measure.

Since the dynamics of the activated sludge process change with time, the use of pure empirical models are limited. On the other hand, the processes are so complicated that it may be practically impossible to obtain reasonable models using only physical insight as the tool for modelling. Between the two extreme model types described above, exists a wide spectrum of alternatives leading to more or less 'grey-box' models, which combine the main physical understanding of the processes with the possibility to identify and estimate unknown states and parameters (with no direct physical interpretation) in the model by mathematical methods.

A problem which is often more or less neglected when a model is developed, is the fact that many of the parameters in the model are not possible to measure. Although numerous variables are measurable in the activated sludge process, some of the most essential ones are not - at least not without lengthy and expensive laboratory investigations. Furthermore, uncertainties connected with sampling and measurements introduce severe noise into the models and the lack of reliable on-line measurements is a permanent source of problems. This implies that any model which is to be used successfully, must take this into account.

The development of mathematical models often imply a balance between high complexity and the possibility of intuitive understanding of the model. Potential users are reluctant to use models that are overly complex and even more important - the more complex a model is, the more difficult it is to verify and to determine and identify the parameters involved. This problem is associated with the measurement problem mentioned earlier.

Therefore, unless a *significant* improvement in model performance (according to some preset criteria) can be statistically proved, new complexities should not be added - except for research purposes where possible hypotheses must be tested in order to gain new knowledge.

A final difficulty when modelling the activated sludge process is the different time constants involved. The timescale for the process varies from seconds (for the airflow) to seasons (for the temperature and certain flow patterns), [Olsson, 1987 and Olsson, 1989]. This leads to a very stiff system of differential equations which is not only hard to solve numerically (this will not be further discussed in this thesis) but also give rise to the question if different model structures should be used for different time horizons. For long term goals, the model might emphasize other parts of the process than if the fast dynamical behaviour is being examined.

2.2 Literature Review

A large number of different models for the activated sludge system have been presented over the last twenty years. To review all of them is outside the scope of this thesis. Though in order to give an indication of how the work has proceeded, some examples are discussed below.

Mathematical models for some parts of the wastewater treatment process have quite a long history. For example, Thomann proposed a dynamic model for river water quality in 1963 [Thomann, 1963], Eckenfelder presented a biokinetic model for the activated sludge process in 1966 [Eckenfelder, 1966], and Andrews developed a dynamic model for anaerobic digestion, which dates back to 1969 [Andrews, 1969]. During the 1970s the improved computing power together with the falling price for computers liberated mathematical modelling from most constraints. Large systems of partial and ordinary differential equations could now be numerically solved and a number of new models were presented. However, the complexity of these models

also seemed to increase as the computers became more sophisticated and the essential questions - whether theory reasonably approximated reality and how the results should be interpreted in order to verify new knowledge about the processes - were often not sufficiently investigated.

Some of the most important and impressive work concerning modelling of the activated sludge process has been performed at the University of Cape Town (UCT), South Africa. In 1976, a steady state aerobic model was presented [Marais and Ekama, 1976] for carbonaceous and nitrogenous conversion and removal. This model was in turn based on earlier proposals [McKinney, 1962 and McKinney and Ooten, 1969] which had suggested a number of new approaches. Some of these are listed below.

- The mixed liquor can be divided into three volatile solids fractions: active, endogenous-inert and inert (from the influent).
- A relationship between the mass of substrate utilized and the active mass of organisms synthesized was stated.
- An accumulation of endogenous-inert solids takes place because of endogenous respiration.
- A relationship between the oxygen demand and the organisms synthesized and the active mass loss due to endogenous respiration was stated.
- An accumulation of inert solids takes place due to the presence of this material in the influent wastewater.

Marais and Ekama accepted these proposals but added the idea from [Lawrence and McCarty, 1970] linking the specific organism growth rate to the substrate concentration via the Monod relationship (the Monod equation will be further discussed later on). They further suggested that influent carbonaceous material should be divided into three fractions:

- biodegradable;
- unbiodegradable particulate;
- unbiodegradable soluble;

and the influent nitrogen should be divided into four fractions:

- unbiodegradable soluble;
- unbiodegradable particulate;
- biodegradable organic;
- free/saline ammonia.

Again the Monod approach was used to describe the conversion of ammonia to nitrate. The biochemical oxygen demand (BOD) was rejected as a suitable parameter for defining the carbonaceous material and instead they accepted the electron donating capacity in its equivalent form, the chemical oxygen demand (COD). The oxygen utilization rate was also recognized as the most sensitive parameter against which to test the behaviour of proposed models to the activated sludge process.

The Marais/Ekama model was furher refined to include the denitrification process and the dynamic behaviour of the activated sludge system. In accordance with practical experiments with cyclic loadings, it was proposed that the biodegradable COD in the influent actually consisted of two fractions: readily biodegradable and slowly biodegradable COD [Ekama and Marais, 1979]. This was the *bi-substrate hypothesis*. The readily biodegradable COD was assumed to consist of simple molecules able to pass through the cell wall and immediately be used for synthesis by the organisms. The slowly biodegradable COD, which consisted of larger complex molecules, were enmeshed by the sludge mass, adsorbed and then required extracellular enzymatic breakdown (often referred to as hydrolysis) before beeing transferred through the cell wall and used for metabolism.

Based on the above concept, an aerobic activated sludge kinetic model including nitrification was presented, which produced very good predictions of system behaviours under cyclic load and flow conditions [Dold et al., 1980]. Another new approach was also introduced, namely the *death-regeneration hypothesis*. This was an attempt to single out the different reactions which take place when organisms die. The traditional endogenous respiration concept described how a fraction of the organism mass disappeared to provide energy for maintenance. However,

practical experiments with anaerobic tanks in connection with aerobic (or anoxic) ones showed that the endogenous respiration model was not satisfactory. In the death-regeneration model, the cell material was released through lysis. One fraction was unbiodegradable and remained as an unbiodegradable residue while the remaining fraction was considered to be slowly biodegradable. It could thus return to the process and be used by the remaining organisms as substrate through hydrolysis.

Finally, denitrification was incorporated in the model to produce a general nitrification/denitrification activated sludge kinetic model [Van Haandel et al., 1981]. This was done by using the same formulations regarding readily and slowly biodegradable COD under aerobic conditions for anoxic conditions instead. The main difference being that the hydrolysis rate under anoxic conditions had to be reduced.

In 1983, the International Association on Water Pollution Research and Control (IAWPRC) formed a task group which were to review existing models for the activated sludge process. The main goal was to develop the simplest model having the capability of predicting the performance of single sludge systems carrying out carbon oxidation, nitrification, and denitrification. The result was presented in 1987 [Henze et al., 1987a and 1987b] as the IAWPRC Activated Sludge Model No. 1 (Appendix B).

Many basic concepts were adapted from the UCT model discussed earlier, such as the bi-substrate hypothesis and the death-regeneration hypothesis. Again the standard Monod relationship was used to determine the growth rate of both Heterotrophs and Autotrophs. COD was selected as the suitable parameter for defining the carbonaceous material.

Some substantial changes were also proposed by the IAWPRC task group in regard to the UCT model. Firstly, the enmeshed slowly biodegradable substrate was not considered to be adsorbed on the organism mass but directly hydrolysed and released to the bulk liquid as readily biodegradable substrate. Secondly, the fate of organic nitrogen and source of organic nitrogen for synthesis were treated somewhat differently.

As a comparison, the fourteen process equations of the UCT model were reduced to eight equations in the IAWPRC model. An evaluation of the two models [Dold and Marais, 1986 and Dold et al., 1991] showed more or less identical predictions under most conditions when properly calibrated. The task group also introduced the concept of switching functions to turn process rate equations on and off as the environmental conditions were changed (mainly between aerobic and anoxic conditions). The switching functions were Monod-like expressions which were mathematically continuous and thereby reduced the problems of numerical instability during simulations. Furthermore, the work of the group promoted the structural presentation of biokinetic models via a matrix format, which was easy to read and understand. The IAWPRC model also consolidated much of the existing knowledge on the activated sludge process and is today considered as the principle 'state-of-the-art' model. It will be further examined and discussed later on.

Another approach to develop a structured kinetic model for the activated sludge system is given in [Padukone and Andrews, 1989]. The proposed model is stated to be the simplest one capable of giving a realistic description of the contact stabilization process for carbonaceous removal although no validation of the model using experimental data is presented in the paper. Based on a traditional storage/metabolism hypothesis for the substrate, the rate equations are chosen in a way which reduce them to the Monod equation during 'balanced growth' (when the external conditions to which the cell is exposed change so slowly that its composition remains perfectly acclimated to them, for example in the completely mixed activated sludge process). Because the rate equations are linear, the cell growth and substrate uptake in a stirred tank can be defined exactly in terms of the average composition of the biomass. The composition of the flocs is described by the ratio of stored substrate to active biomass. The number and type of parameters and state variables make this model difficult to verify and would require lengthy experiments in order to update the parameters for changing environmental conditions.

A simplified model is presented in [Fujie et al., 1988]. It predicts the concentration of organic material in the aeration basins and in the effluent from a wastewater treatment plant performing

only carbonaceous removal. Only soluble organic substance is modelled since the particulate material is considered to be immediately adsorbed by the activated sludge and thereby remain within the system. The model is easily verified since practically all parameters and state variables are directly available through simple measurements.

In the paper the predictions are validated against experimental data and they show a large degree of agreement. It has to be emphasized however, that the effluent concentration of organic substrate is not the most suitable variable for modelling a modern treatment plant receiving municipal wastewater. This concentration is usually so low that the uncertainty of any measurement is considerable. Since most modern plants also perform nitrification/denitrification, the sludge age is usually so long that the effluent concentration of organic soluble material is more or less negliable.

A number of mechanistically simplified models for the organic substrate and the active biomass are given and tested against each other in [Sheffer et al., 1984]. Ways of automatically selecting the best possible model for a certain purpose are also discussed as well as the need for on-line updating of model parameters. A similar comparison between different levels of mechanistic simplification of the IAWPRC model to experimental data is given in [Gujer and Henze, 1991]. Complete models for entire wastewater treatment processes, including primary settling, aeration, secondary settling, gravity thickening, anaerobic digestion, waste disposal etc., have also been proposed, for example [Tang et al.,1988]. Such large models are usually only valid under steady state conditions and are mainly used to analyse the most cost-effective approach for an entire plant.

A somewhat different modelling approach is suggested in [Benefield and Molz, 1984]. It is based on a modified Monod relationship and the transfer of nutrients into the flocs is modelled as spherical molecular diffusion depending on the floc radius. Biological phosphorus removal is also included in the model though in a very rudimentaryform. The rate of the removal is simply stated to be directly proportional to the rate of microbial growth. The model is further investigated and validated in [Benefield and Reed, 1985].

Several new models including biological phosphorus removal has been proposed in recent years, e.g. [Wentzel et al., 1986, Dupont and Henze, 1989, and Dold, 1992]. The phosphorus part is often added on as an extension to an earlier accepted model, for example the IAWPRC model. Due to the complexity of the processes controlling the phosphorus removal mechanism, any attempt to model it will significantly enlarge the model. As an example, the mathematical model suggested by Dupont increases the number of state variables from thirteen to twentyone and the number of rate equations from eight to twentyone when compared to the standard IAWPRC model, on which it is based. This implies a huge problem when trying to calibrate the model to any realistic set of data and an impossible task to verify each individual parameter. The need for model simplifications is again emphasized.

As a complement to the mechanistic models discussed above, models based on stochastic analysis and other methods of system identification are gaining interest. Although often incorporating some physical knowledge of the process, the possibilities of adapting such models to the amount of information available is good. The identifiability and verifiability become easier. Maybe more important is the potential of estimating and identifying parameters and state variables of traditional mechanistic models, which are not accessible through direct measurements. The difficulties of model calibration and lacking model reliability can this way be minimized. These possibilities will be further discussed later on.

An excellent introduction to the potential of process identification in wastewater treatment is given in [Beck, 1989] and a comprehensive survey of different examples of identification for wastewater treatment applications is presented in [Beck, 1986]. Often the identification methods are applied to certain specific but important parts of the treatment process. The processes are usually identified in specially designed experiments created to excite the unknown variables as much as possible. For example, the respiration rate and oxygen transfer rate can be identified on-line by allowing the dissolved oxygen setpoint to oscillate in a certain way, [Holmberg, 1991]. In a similar manner, special experiments may be used for identification of the nitrification

process [Ossenbruggen et al., 1991]. A nonlinear stochastic model for adsorption in batch reactors is suggested in [Argyelan et al., 1991].

Finally, one of the most fundamental relationships for modelling wastewater treatment systems is discussed. The Michaelis-Menten equation was originally proposed to describe the enzymatic reaction rate as a function of the substrate concentration. Monod adapted this relationship [Monod, 1942] to model the microorganism growth rate μ as a function of the limiting substrate concentration S, defined by equation (1.1). It was originally presented as an empirical expression for monocultures but is today used in practically every dynamical model for wastewater treatment systems in the original or slightly modified form.

The growth rate is one of the most sensitive parameters in most models. It has been shown [Holmberg and Ranta, 1982] that both μ and K_S are theoretically identifiable from perfect measurements but the situation is more difficult when using true measurements including noise. This means that unique sets of parameters can rarely be obtained. Parameters estimated from data obtained during apparently similar conditions show considerable variations and parameter estimation methods show poor convergence properties, [Holmberg, 1982].

In [Vialas et al., 1985] ways of improving the practical identification of the Monod equation by using different sample times depending of the current state of the process, is suggested. Both linear and nonlinear regression techniques are applied for estimating the growth model parameters as well as the yield coefficient Y and the decay rate b from true plant data in [Vaccari and Christodoulatos, 1990].

The estimates for K_S and b were, however, not significant at the 95% confidence level and therefore the use of a simple first order rate equation instead of the Monod expression is proposed. A comparison of the nonlinear Monod equation and a linear simplification is also performed in [Derco et al., 1990a and Derco et al., 1990b]. The investigations show that a linear rate model is not as good for predicting actual transient responses in biomass

and substrate concentration as the traditional formulation. The standard Monod equation does, however, not provide perfect results either, when compared to true data.

2.3 State Variables

The state variables characterize a dynamic system; each describing an internal process variable at a certain time. These states may have a physical significance, be artificial, or even symbolize statistic probabilities. By integrating the state equations forward in time, the future state of a system can be predicted.

In the specific case of a biological wastewater treatment process performing carbon oxidation and nitrification/denitrification, the state variables of most models are related to concentrations of different types of microorganisms, carbonaceous fractions, nitrogen fractions and dissolved oxygen. They have a direct physical interpretation though some of them have been more or less adapted because of their empirical importance and without regard to the possibility of verification. This will be discussed in section 2.5.

The number of states in most activated sludge models presented in the literature, vary between ten and fifteen for a completely mixed system. If biological phosphorus removal also is to be included in a model, between five and ten more state variables are usually added. Though for every new state, a differential equation describing the reaction rates has to be included, which indirectly implies adding new parameters to the model. Since many of the state variables are non-measurable, the complex models are extremely difficult to verify.

Altogether, increased complexity may actually make a model worse due to calibrational problems, lack of unique solutions, and lost observability and controllability. For similar reasons, the possibilities to develop improved automatic control systems for the processes and better parameter estimators based on the

current models is reduced, due to the higher complexity. Finally, the financial investments in personal and equipment for measurement and analysis - in order to keep such models 'on-track' when used in real applications at wastewater treatment plants - are increased.

As an example, the different state variables introduced by the IAWPRC and the UCT models are presented in Table 2.1.

State Variables	IAWPRC	UCT
Soluble inert organic matter	S_I	S_{us}
Readily biodegradable substrate	S_S	S_{bs}
Particulate inert organic matter	X_I	-
Inert mass	-	Z_I
Endogenous mass	-	Z_E
Slowly biodegradable substrate	X_S	-
Adsorbed slowly biodegradable substrate	-	S_{ads}
Enmeshed slowly biodegradable substrate	-	S_{enm}
Active heterotrophic biomass	$X_{B,H}$	Z_{BH}
Active autotrophic biomass	$X_{B,A}$	Z_{BA}
Particulate products arising from biomass decay	X_P	-
Oxygen	S_O	O
Nitrate and nitrite nitrogen	S_{NO}	N_{o3}
Ammonia nitrogen	S_{NH}	N_a
Soluble biodegradable organic nitrogen	S_{ND}	N_{obs}
Particulate biodegradable organic nitrogen	X_{ND}	N_{obp}
Alkalinity	S_{ALK}	Alk

Table 2.1 State variables of the IAWPRC and UCT models.

2.4 Model Parameters

The selection of values for the kinetic and stochiometric coefficients of a mathematical model is known as model calibration. In the case of activated sludge models, the calibration has traditionally been carried out through specific and well controlled experiments at pilot and bench scale plants assuming constant operating conditions. The values obtained in such a way may not be reliable mainly for two different reasons. The first reason being the difficulty of configuring and operating a small scale plant in exactly the same way as a full scale plant and therefore introducing a risc of changing the behaviour of the microorganism population and also the conditions which influence the values of the parameters which should be determined. The second reason is that the experiments and calculations are often based on the fact that the coefficients are constants. Since the experiments may take several weeks to perform, they are not carried out very often. Many of the parameters are time variant and some of them may show considerable change over a limited period of time. Factors such as plant configuration, operating conditions, microorganism population dynamics, degree of inhibition by toxic compounds, composition of incoming wastewater, temperature, pH, etc., all affect the values of the process parameters.

By examining the sensitivity, variability, and uncertainty of the model parameters, an indication is given as to which coeffients are most important to determine accurately. Such an investigation is performed in [Henze, 1988]. It is stated that for plants including nitrification/denitrification, models show very little sensitivity to the COD. The parameters which are most important are the heterotrophic decay rate, denitrifying growth rate, denitrifying hydrolysis rate, hydrolysis rate and saturation, and maximum growth rate of nitrifyers (for the IAWPRC Activated Sludge Model No. 1). It is also shown how different sets of parameter values may lead to approximately the same model behaviour.

Another problem is the fact that many coefficients are correlated. This implies that parameters can often not be adjusted one by one, but rather the whole set must be tuned to each other. Some examples of these interrelations are given here.

- Growth rate and decay rate - increased growth and decay rate may give identical net growth rate but will increase the oxygen demand and speed up the substrate cycling.
- Yield and growth rate - increased yield and growth rate may outbalance each other with respect to substrate conversion rate but will increase the oxygen consumption.
- Yield and heterotrophs in inflow - high yield and low concentration of heterotrophs in feed equal low yield and high concentration of heterotrophs in feed.

As an example, the model parameters used in the IAWPRC model are presented together with their default values (as suggested by the IAWPRC task group for 20°C) in Table 2.2. As a comparison, values commonly found in the literature are provided for some of the coefficients.

The same type of problems are even more emphasized for characterizing the incoming wastewater. Some mechanistic models distinguish between ten different fractions of nitrogen and organic material in the influent although only three or four of these fractions are possible to measure on-line. But while the parameters discussed above may change their values considerable over a few days, the characteristics of the incoming wastewater may change significantly within a few hours. The fact that the influence of the influent wastewater composition on model behaviour is usually large, further amplifies these difficulties.

The situation outlined above is an indication that methods for identifying and estimating the non-measurable state variables and model parameters have to be employed. This should be done in order to extract all possible information from the actual on-line measurements.

IAWPRC model parameters	Default	Literature
Heterotrophic max. specific growth rate [day^{-1}]	6.0	1.5-13.2
Heterotrophic decay rate [day^{-1}]	0.62	0.05-1.6
Half-saturation coefficient for Heterotrophs [g COD m^{-3}]	20.0	5-180
Oxygen half-saturation coefficient for Heterotrophs [g O_2 m^{-3}]	0.20	0.01-0.20
Nitrate half-saturation coefficient for denitrifying Heterotrophs [g NO_3-N m^{-3}]	0.50	0.1-0.5
Autotrophic max. specific growth rate [day^{-1}]	0.80	0.2-1.0
Autotrophic decay rate [day^{-1}]	0.20	0.05-0.20
Oxygen half-saturation coefficient for Autotrophs [g O_2 m^{-3}]	0.4	0.5-2.0
Ammonia half-saturation coefficient for Autotrophs [g NH_3-N m^{-3}]	1.0	-
Correction factor for anoxic growth of Heterotrophs [dimensionless]	0.8	0.6-1.0
Ammonification rate [$m^3 \cdot$(g COD$\cdot$ day)$^{-1}$]	0.08	-
Max. specific hydrolysis rate [g slowly biodeg. COD (g cell COD$\cdot$day)$^{-1}$]	3.0	-
Half-saturation coefficient for hydrolysis of slowly biodegradable substrate [g slowly biodeg. COD (g cell COD)$^{-1}$]	0.03	-
Correction factor for anoxic hydrolysis [dimesionless]	0.4	-
Heterotrophic yield [g cell COD formed (g COD oxidized)$^{-1}$]	0.67	0.46-0.69
Autotrophic yield [g cell COD formed (g N oxidized)$^{-1}$]	0.24	0.07-0.28
Fraction of biomass yielding particulate products [dimensionless]	0.08	-
Mass N/Mass COD in biomass [g N(g COD)$^{-1}$ in biomass]	0.086	-
Mass N/Mass COD in products from biomass [g N(g COD)$^{-1}$ in endogenous mass]	0.06	-

Table 2.2 Parameters of the IAWPRC model (20 °C).

2.5 Measurable Quantities

The earliest models for the activated sludge system used state variables which were readily measurable. Calibration techniques were based on results obtained by operating continuous plants at steady state for different sludge retention times. Since then, mechanistic models have evolved considerably. In order to more precisely explain the phenomena taking place, many state variables and parameters which are not directly measurable, have been introduced.

It is obvious that the measurement technique has not evolved as fast as the complexity of the models. Due to this fact there are two principle ways to proceed:

- to accept large models;
- to use simplified, reduced order models.

The first implies accepting the complex models in their current form and improve matters by measuring as much as possible, designing special identification experiments, and developing new and better measuring techniques and instruments. Identifiability may be improved by means of exciting the system, perturbing the input and control signals in an optimal manner, properly choose the sampling instants, use various methods of signal processing, etc. The best design of identification experiments is a very troublesome task but of the outmost importance in order to produce reliable results and is further discussed in [Söderström and Stoica, 1989 and Ljung, 1987]. The major drawbacks are the large amount of resources required (both equipment and personnel), the high degree of uncertainty (two equally skilled persons may reach quite different results when performing identical experiments due to the need for subjective interpretation of many results), the long time to perform certain experiments (some results may be obsolete by the time they are reached), and the lack of standarized methods (different methods for determining the same quantity may show considerable variations). Furthermore, many parameters used by complex models have to be considered as constants because of the practical

difficulties of performing identification experiments as often as would be needed in order to keep track of their variation.

The second possibility is to use simplified, reduced order models and fully take advantage of the quantities which are measurable on-line. In combination with mathematical identification-estimation algorithms the model should be automatically calibrated on-line and always be tuned to the current situation of the plant (adaptive models). The drawbacks in this case are the lack of reliable on-line sensors (only a few quantities can be measured). The lack of data may leed to an overly simplified model for producing any realistic predictions (the number of parameters which can be accurately estimated is directly related to the amount and quality of the data available). Moreover, the cost and need of maintenance for advanced instrumentation are quite high.

A combination of the two described approaches is naturally an alternative. Depending on the purpose of the model and for what time scale the model is to be used, the best procedure can be selected. A model used for design of new plants simulates plant behaviour over long periods of time. The variations which are of real relevance are those with time constants of days and weeks. On the other hand, an oxygen regulator for an aerobic reactor reacts within minutes and parameter changes with time constants of seconds and minutes have to be detected. This can only be accomplished by on-line measurements.

The main measurement problem for wastewater treatment plants is usually not lack of data. On the contrary, a large amount of information is being logged at a modern plant. The number of inputs from sensors usually vary between a few hundred and several thousand. The problem is often how to obtain the most important quantities for modelling purposes with adequate accuracy and sampling frequency.

Traditionally, on-line measurements have been restricted to physical/chemical variables such as flow rates (of both water, sludge and air), power to pumps, levels in reactors, temperature, pH, redox potential, water conductivity, etc. The introduction of on-line meters for measuring the oxygen concentration of the wastewater led to intensive research and development of

regulators, estimators and models for controlling the dissolved oxygen concentration as a key variable. This work has continued over the last two decades. The meters are today quite reliable and the oxygen consumption is today the most commonly used parameter for verifying mathematical models as well as being used for control purposes.

On-line meters for measuring the COD (chemical oxygen demand), the ammonia concentration, and the nitrate concentration in the wastewater have also become available. Due to this fact, the possibilities of developing adaptive and selftuning models have increased considerably. New technologies for advanced sensors also provide interesting possibilities for the future. Ultrasonic sensors for on-line determining the sludge level as well as the concentration profile in the settler and optical UV sensors for on-line measurements of the bacteriological activity are now becoming available. On-line respirometers of various construction and philosophy [Spanjers and Olsson, 1992 and Vanrolleghem and Verstraete, 1993] are also being prototyped. Furthermore, the substrate concentration can be estimated from respirometry data [Spanjers et al., 1993]. In the future, models will have to be modified to include information from these types of sensors.

A large number of parameters and quantities can today be determined off-line using special experimental setups, either by direct or indirect methods. Such experiments are generally batch or dynamic tests in continuous plants operated in bench or pilot scale under very special conditions. In this manner only the coefficients to be determined are supposed to influence the measurable variables throughout the experiment. Whether the acquired results are directly applicable to full scale plants or not are a matter of debate. Since this thesis deals mainly with on-line measurements it is beyond its scope to describe the methods any further but a number of possible procedures and experiments are described in [Cech et al., 1985, Ekama et al., 1986, Grady et al., 1989, and Kappeler and Gujer, 1991].

As an example, the IAWPRC model contains five stochiometric coefficients, fourteen kinetic parameters, and five non-measurable state variables [Larrea et al., 1991]. A general description of an extensive experimental procedure to determine

the unknown values of the model is given in [Henze et al., 1987a]. This is a combination of practical experiments and curve-fitting procedures and it is clearly stated that an error introduced when determining certain coefficients will be compensated when determining another parameter. This might seem satisfactory but is actually an indication of lacking identifiability (non-unique solutions) of the model since different sets of model parameters will produce identical results. Consequently, the model will fit practically any set of data if the model coefficients are allowed to vary sufficiently.

The possibility of employing mathematical estimation algorithms for model calibration on-line during normal plant operation will be discussed in Chapters 4 and 5.

2.6 Manipulative Variables

Traditionally, biological wastewater treatment processes have been regarded as more or less self-controlled and quite inflexible in their operation. The plants normally function under pseudo steady state conditions for long periods of time which are suddenly interrupted by abrupt failures. Some of these instabilities can be attributed to high frequency, external disturbances of high amplitude but most are probably due to the propagation of slowly variable, internal perturbations in the largely inaccessible microbiological state of the system. The available control usually depends on the expertise of the plant operator in combination with a few automatic, single-loop controllers (mainly for the concentration of dissolved oxygen and other local variables).

Apart from the previously discussed problems of reliable and relevant measurements, the difficulties of controlling the activated sludge process are further emphasized by lack of suitable control variables. Flow rates can usually be controlled - both of wastewater and sludge. For example, a desired sludge age can be determined by the sludge recycle flow and the sludge waste flow. Different modes of operation may also be imposed,

such as step feed, contact-stabilization, pre-fermentation, etc.

Apart from changing the internal flow schemes and controlling the concentration of dissolved oxygen, the ways to influence a plant are quite limited. At some modern plants tests are being performed in order to investigate the possibility of using the sewer system or large equalization basins to smooth out dramatic flow variations of the influent as well as adding different sources of substrate at specific points of the plant to improve certain reactions, for example the denitrification. Different types of chemical inhibitors may also be added to slow down other biological reactions, the pH may be changed, etc. Finally, the dissolved oxygen setpoint or spatial distribution can be used to favour growth of certain organism species.

One of the major problem when attempting to control an activated sludge plant is the lack of control of the different concentrations in the influent wastewater. The control strategy is therefore often based on empirical knowledge and statistic information for each specific plant. Factors such as temperature (which is basically not controllable) also affect plant performance dramatically.

The preferred situation would naturally be that every parameter of the system could be individually manipulated but the situation is quite the opposite. When the internal flow scheme is changed and various substances added, practically all process parameters are indirectely affected in a very complex way. Over longer periods of time, the system behaviour may change significantly. This problem is often not considered in most mathematical models. Instead the model coefficients are considered to be more or less constant. More adequate and adaptive models are necessary in order to predict both the long and short time behaviour of a plant and to determine suitable control strategies.

3

Validity of Mechanistic Models

In this chapter the concept of model validity is discussed. A short introduction is given in section 3.1 and a commented literature review is performed in the following section, concerning issues such as verifiability, identifiability, and validity of biological wastewater treatment models. The problem of model complexity versus model verifiability is discussed in section 3.3. Finally, a more detailed investigation of model identifiability and verifiability implications is performed in the last section.

3.1 What is a valid Model?

Writing a mathematical model is generally easier than verifying it. Just as it is extremely important to determine the purpose of a model prior to its development, it is equally important to test the validity of a model once it is implemented, e.g. investigate how well and under what specific conditions a model realistically mimics the true system behaviour and verify that the purpose of the model is fulfilled. A model may for example be valid for:

- steady state behaviour (no transients);
- various types of dynamic behaviour and time horizons;
- certain operating conditions;
- certain input (amplitude, variability, frequency) conditions;
- specific noise distributions;
- qualitative comparisons.

This situation is especially true for highly complex processes like

those of biological wastewater treatment, where no such thing as *the true model* exist. Different models have different advantages and purposes.

Model discrimination and model verification are closely related because in both cases the model must be subjected to suitable practical tests to discover its weaknesses. It is important that the performed tests put the model in jeopardy and a godness of fit is not a sufficient condition for model acceptance. Several models may in fact fit the same data (as shown in the small example in Chapter 1). A related problem for model verification is that different postulated reaction mechanisms may lead to the same mathematical function, thus making it impossible to verify certain mechanisms by model fit.

Since most models for biological processes are mechanistic, it is actually not sufficient that only the output of the model is validated against the true process. Most model parameters have a direct physical interpretation which implies that the validation has to include an evaluation of those parameters when compared to the actual process parameters. However, mechanistic models nearly always have empirical qualities and it is unlikely that any biological or biochemical system has ever been described exactly by a theoretical model. Thus, a 'true' mechanistic model is one which describes the mechanisms well enough to assist understanding and to allow useful - but not exact - extrapolation. Clearly, the classification of biological models as empirical or mechanistic depends on what is expected of the ultimate model.

Models are often considered to be validated when they have been identified from a couple of different data sets. Strictly speaking such models are simply a convenient mean of describing collected data or of summarizing current knowledge. Situations may arise in which two or more models based on partly contradictory hypotheses may explain experimental results equally well [Holmberg, 1981]. It should, however, be noted that no methods exist which provide a completely validated model. A certain degree of uncertainty and untested conditions always remain. The ultimate validation of a model can only be performed using it in practice and checking the results over a longer period of time [Söderström and Stoica, 1989].

3.2 Literature Review

As the complexity of wastewater treatment models have increased, so have the needs for verification and identification. During the last five years the number of publications dealing with adaptive models and control, fuzzy logic, on-line parameter estimation, and other quite sophisticated methods applied to wastewater treatment have risen dramatically. This may be considered as a new trend in an otherwise rather conservative research area.

A very good introduction to the possibilities and difficulties of identification, estimation and control of biological wastewater treatment processes is given in [Beck, 1986]. A list of more than one hundred relevant references covering much of the work produced in this area up to 1986 is also provided. In [Beck, 1991] the concept of model calibration versus model uncertainty is further emphasized.

In a paper by [Thomann, 1982] the need for verification of water quality models is discussed. The conclusions are, however, valid for all kinds of complex models and quantitative statistical measures such as regression analysis, relative error, root mean square error, and comparision of means are suggested to be useful. Acollection of more general methods for identifiability analysis is presented in [Godfrey and DiStefano, 1985]. Five different approaches are thoroughly described, though only one of them is suitable for nonlinear models, i.e. the Taylor series expansion of observations. The difficult question of identifiability in the presence of real, noisy data is also considered.

One of the more extensive investigations concerning identifiability, verifiability, and state estimation in wastewater engingeering is presented in [Holmberg, 1981]. The Michaelis-Menten equation is analyzed and it is shown that the parameters are not generally identifiable in the presence of noisy measurements. The usefullness of sensitivity analysis as a mean for model calibration, experiment planning, and model reduction is also described.

The term 'flexible modelling' is introduced in [Sheffer et al., 1984]. Depending on the situation and quality of the available data a suitable model - from a set of mechanistic models for carbon removal at different levels of simplification - is automatically selected according to a predefined criteria. The model comparison program is further connected to a program for on-line parameter fitting although the identifiability problem is not investigated.

The use of state and model coefficients estimation for the activated sludge process is investigated for model calibration purposes in [Ayesa et al., 1991]. A modified Kalman filter algorithm is proposed and tested on a simplified form of the IAWPRC model, excluding denitrification. Unique solutions are reached also when noise is present in the computer simulated data. The work is continued in [Larrea et al., 1991] where suitable identification experiments are suggested, to be used in combination with the identification algorithms of Ayesa, though verified only in theory.

In order to determine (by direct measurements) the kinetic parameters of heterotrophic biomass and the COD wastewater fractions of the IAWPRC model, a simple method based on three different types of batch tests is proposed in [Kappeler and Gujer, 1991]. The results are achieved simply by measuring the oxygen respiration and are used for model calibration. It is also useful as a mean for verifying the results based on different mathematical estimation algorithms.

The identifiability of the respiration rate and oxygen transfer rate is investigated in [Holmberg, 1991]. The two parameters are found to be impossible to identify from a single experiment and any identification algorithm will produce a biased result.

It is interesting to study the evaluation of the IAWPRC model performed in [Dold and Marais, 1986]. Although based on an excellent understanding of the process kinetics involved and a very thorough model validation based on actual plant data under various conditions, the identifiability problem is not discussed at all. It is essential to realize that the more complex a model is, the greater is the necessity for such an analysis in order to ensure that model results are not misused or misunderstood.

3.3 Mechanistic Model Complexity

The choice of a model structure is in practice greatly influenced by the intended use of the model. A stabilizing regulator can often be based on a crude low-order model, whereas more complex and detailed models are necessary if a model is aimed at giving physical insight into a process.

The complexity of a system makes model simplification necessary and the final model is probably a compromise between a practical need for easily handled models and a desire for biologically explanatory models. The most common statistical criteria to discriminate between models are normally based on [Boyle and Berthouex, 1974]:

- minimum sum of squares of residual errors;
- fewest parameters;
- simplest form.

From these criteria it is apparent that a low model complexity is as important as the actual fit of the model. Notice that these two demands are often contradictory and a compromise has to be reached.

When modelling the activated sludge process, the difficulty in relating measurable variables to model variables constitute a serious problem. A further difficulty is the fact that the process is time variant. Even in a complex model the parameters still have to be updated in some way. Strong nonlinearities and varying time constants of the process do not improve matters.

By combining the most essential processes of the mechanistic models with some empirical input/output relationships, a low complexity 'grey-box' model may be developed which can be made adaptive by the use of mathematical estimation algorithms and thereby useful for automatic control purposes. Such a model should, however, not be expected to explain all the phenomena involved in the physical process. Some statistical means of determining whether a model is overly complex or not are

discussed in [Söderström and Stoica, 1989].

It should also be realised that not only the complexity of a model determines whether there will exist an identifiability problem or not. Various algebraical structures and nonlinearities within an otherwise simple differential equation may give rise to similar difficulties. On the other hand, a linearization of a nonlinear system may very well be non-identifiable whereas the original system shows no such weaknesses [Godfrey and DiStefano, 1985] - if it is at all possible to perform an analytical investigation.

3.4 Identifiability and Verifiability

The concept of identifiability can be introduced in a number of ways. In a mathematical sense it may be defined as follows. An identification method I applied to a parametric model structure M will yield the resulting estimate denoted by $\theta(N; S, M, I, E)$. The estimate will depend not only on I and M but also on the number of data points N, the true system S, and the experimental condition E. The system S is *system identifiable* under M, I, and E, abbreviated SI(M, I, E), if

$$\theta(N; S, M, I, E) \dashrightarrow D_T(S, M) \qquad \text{as } N \dashrightarrow \infty \tag{3.1}$$

with probability one and $D_T(S, M)$ is nonempty. The set $D_T(S, M)$ is a description of those parameter vectors for which M gives a perfect description of S and may theoretically be empty (model under-parametrization), consist of one point (ideal - the true parameter vector) or consist of several points (model over-parametrization).

The system S is further considered to be *parameter identifiable* under M, I and E, abbreviated PI(M, I, E), if it is SI(M, I, E) and $D_T(S, M)$ consists of exactly one point. The parameter estimate will both be unique for large values of N and consistent, i.e. converge to the true value. The system identifiability property

basically depends on the identification method. It is most desirable that this is true for as general experimental conditions as possible and it is then 'only' the model parametrization or model structure that determines whether the system is also parameter identifiable.

Most tests developed for studying the identifiability properties of a system are aimed at linear model structures. A general test which can also be applied to nonlinear systems is described in [Pohjanpalo, 1978] but practical limitations makes this method unsuitable for more complex models. A further difficulty is the fact that a model structure which is theoretically identifiable may lack practical parameter identifiability due to poor measurements, noise, etc. This may give rise to problems that unique sets of parameters can rarely be obtained, parameters estimated from data obtained during apparently similar conditions show considerable variations, and that the estimation methods show poor convergence properties. Estimation algorithms where the results vary depending on the selected initial values of the parameters, is also an indication to proceed with care. Altogether, it is often easy to obtain sets of parameters which give a good model fit. Since these parameters may be far from the correct ones, situations where they are given an exact biological interpretation should be avoided.

As far as verifiability is concerned, a truly verifiable mechanistic model would not have any internal state variables which are not possible to measure in the real system. As an example, most models for wastewater treatment include two such troublesome state variables - *active* heterotrophic biomass ($X_{B,H}$) and *active* autotrophic biomass ($X_{B,A}$). There is today no way of directly determining these quantities accurately and therefore it is not possible to verify that a model based on $X_{B,H}$ and $X_{B,A}$ really mimics the real process. Several models based on different reaction mechanisms may fit the measurable data well and be valid but provide different estimates for the active biomass concentration. The problem of verifying which model is 'most true' is an awkward task. Although this is not a very serious problem, it should be recognized. Otherwise, it may lead to a situation where new models are rejected not because they do not fit actual measurable data but because the estimates of the non-

measurable variables do not equal the ones provided by more traditional and accepted models - which are not necessarily true.

On the other hand, a model which shows a high correlation between its output signals and the corresponding signals from the real process is regarded to be valid (at least under certain conditions) whether or not the model provides a realistic description of the different reactions taking place. This situation is typical for 'black-box' (empiric) models where the internal states seldom have any physical relevance at all and implies that a model may actually be valid, although it is not fully verifiable.

Most models used today are only partly verifiable, i.e. only a few of the internal state variables are possible to measure in the real process. The awkward situation where the models are used to quantify the real process behaviour instead of the other way around - due to measurement difficulties - may arise, although the mechanistic models used today are still largely based on theoretical hypotheses. It must also be realised that a model is *always* a simplification of reality. This is made obvious by the fact that a model which appears to produce reliable results at one specific treatment plant may very well give misleading results if tested on another plant. This is because the internal reactions are not fully understood and many model quantities are neither measurable nor identifiable.

4

Model Reduction

In this chapter the mathematical modelling approach is taken one step further by introducing some means for reducing model complexity and applying them to the activated sludge process. A brief discussion of the importance of using simple models is given in section 4.1 and the small example from Chapter 1 is further investigated. In the following section a simplifed model for the activated sludge system is developed, mainly based on physical reasoning. The limitations and assumptions for the model are also discussed and its structure is analysed. The issue of model verification is penetrated in section 4.3 together with methods for off-line and on-line model calibration, state and parameter estimation of important non-measurable quantities, etc. The effectiveness of some of the estimation algorithms are also illustrated in the simple example under various operating and initial conditions.

4.1 Why are Reduced Models needed?

A model is nothing more than a mathematical abstraction of a real process. The equation or set of equations that comprise the model are at best an approximation of the true process. Hence, the model can not incorporate all of the features, both macroscopic and microscopic, of the real system. The engineer normally must seek a compromise involving the cost of obtaining the model, that is, the time and effort required to obtain and verify it and the expected benefits to be derived from its use. The ultimate application and purpose of the model finally determines how accurate it needs to be.

In general, modelling is still much of an art. The modeller must bring a significant level of creativity to the task, namely to make a set of simplifying assumptions that result in a realistic model. An 'optimal' model incorporates all of the important dynamic effects, is no more complicated in its structure than necessary, and keeps the number of equations and parameters at a reasonable level. The failure to choose an appropriate set of simplifying assumptions invariably leads to either a rigorous but overly complicated model or models that are overly simplistic. Both extremes should be avoided.

Activated sludge models are often derived from simpler unit operations and later combined to large plant models. The model parameter values consequently may not be the same. Moreover, several parameter combinations can often explain the same dynamical behaviour. This is further accentuated when the influent wastewater composition is taken into consideration; a change in its characteristics can quite often be explained by kinetic parameter changes.

Even if a major problem of the models has to do with the complex structure and the too large number of states and parameters to be uniquely identified, instrumentation problems amplify the difficulties. As earlier discussed in section 2.5, available on-line instrumentation and laboratory procedures are usually not adequate to verify the details of such a complex model. Furthermore, for a reliable identification result, the operation has to be perturbed (or purposefully disturbed) in such a way that all interesting dynamical modes of the process are excited. This creates a demand not only in amplitude but also in the time frame of the disturbances.

It is of course practically impossible to develop a model which is reliable on a microscopic level - this would require a system of several hundred equations. Though available models are quite complex they are still greatly simplifying the representation of many species of organisms. As the microbial population changes this would be reflected in changing kinetic parameters and even adding new state variables. For example, filamentous organisms ought to be represented during many operating conditions. On the other hand, a quite simple model can be used effectively if

key model parameters are properly fitted to operating data of the actual process, especially if the process is time variant.

A simplified model does not provide a fully explanatory model for every physical reaction. Several parts of the process are often lumped together in order to reduce the complexity. For a non-expert, the intuitive understanding of the process is however often enhanced by such models (if its basic structure is mechanistic). Furthermore, in many cases the model output needs only to be qualitatively significant, e.g. show trends and whether a variable is increasing or decreasing, without providing exact quantitative results. This may allow effective use of simplified models for highly complex processes.

The activated sludge process is suited for a hierarchical control structure based on several simple models. The process can in a natural way be devided into unit operations - aerobic reactor, anaerobic reactor, anoxic reactor, clarifier, thickener, presettler, sludge digester, etc. It can also be modularized based on different process time constants - oxygen (minutes), flows (hours), carbon oxidation/denitrification (a few days), nitrification (many days), temperature (months), etc. Each model controls and predicts the behaviour of its specific area in some optimal way but is also syncronized with a high level control system which optimizes the performance of the entire plant according to preset criteria which are often contradictory. The inclusion of a knowledge based system at the top level to allow for logical reasoning, diagnosis, and decision support would further enhance the possibilities. In Figure 4.1 a schematic view of such a hierarchical system is suggested.

Simplified models like the ones aimed at in this work should therefore not be evaluated and judged separately but put in a broader perspective and in the context of a full scale hierarchical control structure. This type of distributed automation has been successfully applied to many complex industrial applications, for example chemical, paper, and pulp processes. A similar approach could be applied to wastewater treatment (WWT) processes.

If the main purpose of a model is control, the need for simplicity is evident. Due to the internal structure of a closed loop system, a reasonably small error will automatically be compensated for.

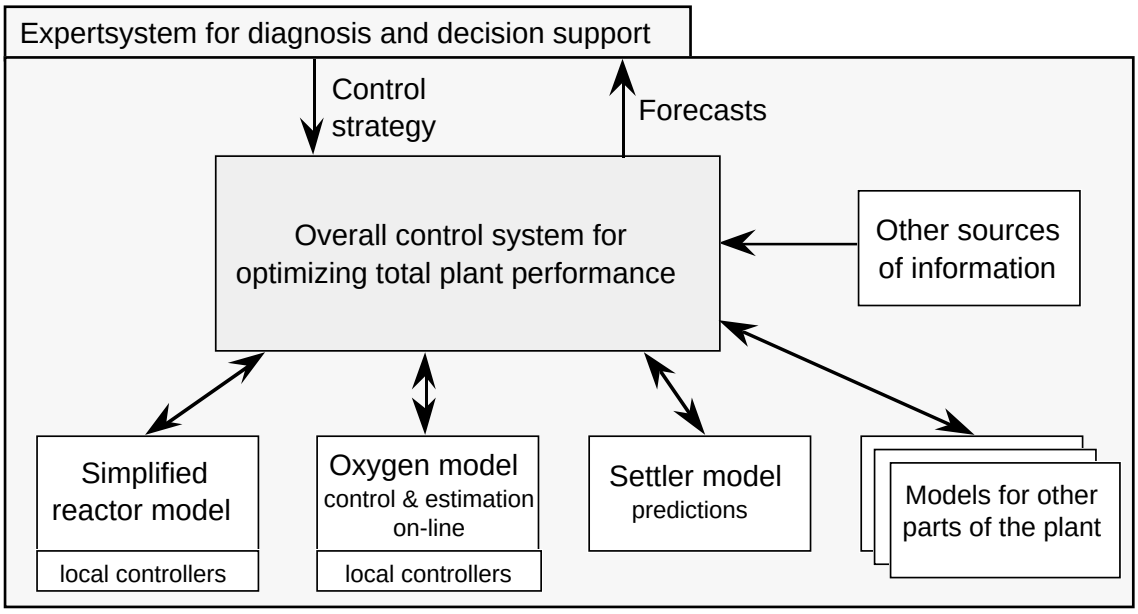


Figure 4.1 Hierarchical control structure of the WWT process.

As the practical control possibilities for the activated sludge process are quite limited, it is even more important to use the ones available to their full extent. Due to the large time constants and the difficulties to early detect problems, traditional control strategies based on feedback are probably not the best solution. Methods using feed forward, prediction, or adaptive algorithms appear to be better suited for this purpose. However, all these methods require a process model which is relatively simple in its structure, robust, uniquely identifiable, and possible to update on-line as the operational conditions change.

A model for operation and control has to be sufficiently complex to describe the major phenomena taking place but still so limited that its parameters can be updated while the plant is runing normally either by taking advantage of the natural disturbances of the process or by introducing small deliberate perturbations. The need for highly complex models is recognized for design purposes so the operational model has to be considered as a special case, either for certain operational levels or for particular time scales. In an operational model it is not always necessary to know the absolute values of certain parameters but rather their relative change.

A simple example (continued)

One of the simplest possible models to include the Monod growth equation describe a single-substrate/single-organism batch reactor with no other growth limitations. It can be formulated as:

$$\begin{cases} \dfrac{dX}{dt} = \mu(S)\,X - bX \\ \dfrac{dS}{dt} = -\dfrac{1}{Y}\mu(S)\,X \end{cases} \qquad (4.1)$$

where: X = concentration of microorganisms [mg/l];
S = concentration of growth limiting substrate [mg/l];
Y = yield factor [g cell COD formed (g COD oxidized)$^{-1}$];
b = decay rate [day^{-1}];
$\mu(S)$ = specific growth rate defined by equation (1.1).

It is assumed that both X and S are possible to measure directly and that only X and S are time variant. The measured data is further assumed to be *perfect*, e.g. no noise and continuously available.

Question: Can all model parameters (μ, K_S, Y, b) be uniquely determined from such measurements?

This may be investigated using various methods. In this case the simplest one is to analytically analyze the system (this is seldom practically feasible for more complex systems). For every time t perfect measurements of $X(t)$ and $S(t)$ are assumed to exist.

Define the following nomenclature:

$$\begin{cases} X_0 = X(0) \\ S_0 = S(0) \end{cases} \qquad (4.2)$$

$$\begin{cases} \dot{X}(0) = \dfrac{\hat{\mu}\,S_0}{K_S + S_0}\,X_0 - bX_0 = X_1 \\ \dot{S}(0) = -\dfrac{1}{Y}\dfrac{\hat{\mu}\,S_0}{K_S + S_0}\,X_0 = S_1 \end{cases} \qquad (4.3)$$

$$\mu = \frac{\hat{\mu}\, S_0}{K_S + S_0} \tag{4.4}$$

Equation (4.3) can now be formulated in a less complex way:

$$\begin{cases} X_1 = (\mu - b)\, X_0 \\ S_1 = -\dfrac{\mu}{Y} X_0 \end{cases} \tag{4.5}$$

If equation (4.3) is differentiated one more time, it gives after some simplifications:

$$\begin{cases} \ddot{X}(0) = (\mu - b)\, X_1 + \dfrac{\mu\, K_S\, X_0\, S_1}{(K_S + S_0)\, S_0} = X_2 \\ \ddot{S}(0) = -\dfrac{\mu}{Y}\left(X_1 + \dfrac{K_S\, X_0\, S_1}{(K_S + S_0)\, S_0}\right) = S_2 \end{cases} \tag{4.6}$$

From equations (4.5b) and (4.6b) an analytic expression for the substrate half-saturation coefficient K_S can be determined:

$$K_S = \frac{S_2\, S_0^2\, X_0 - S_1 S_0^2\, X_1}{S_1^2\, X_0 - S_2\, S_0\, X_0 + S_1\, S_0 X_1} \tag{4.7}$$

From equations (4.5a), (4.6a), and (4.7) the expression for μ can be formulated as:

$$\mu = \frac{X_2 X_0\, S_1 - X_1^2\, S_1}{S_2 X_0^2 - S_1 X_1\, X_0} \tag{4.8}$$

from which an expression for the maximum specific growth rate μ can be determined by applying equations (4.4) and (4.7):

$$\hat{\mu} = \frac{\left(X_2\, X_0\, S_1 - X_1^2\, S_1\right)(K_S + S_0)}{S_2\, S_0 X_0^2 - S_1 S_0\, X_1\, X_0} \tag{4.9}$$

The decay rate b is easily found from equations (4.5a) and (4.8):

$$b = \frac{S_1 X_2 - S_2 X_1}{S_2 X_0 - S_1 X_1} \tag{4.10}$$

and for the yield factor Y from equations (4.5b) and (4.8):

$$Y = \frac{X_1^2 - X_2 X_0}{S_2 X_0 - S_1 X_1} \tag{4.11}$$

The analysis shows that all parameters μ, K_S, b, and Y of the model (4.1) are *theoretically* globally identifiable (e.g. when perfect data is available) from measurements of X and S if neither of the values for $X(0)$ and $S(0)$ are equal to zero.

The method described above is actually a special case of the *Taylor series expansion of observations* method. The analysis is often quite difficult because for nonlinear systems there is no theoretical upper limit to the number of derivates which may provide new information and simpler methods are therefore needed.

An easier way to test for *local* identifiability is to examine the rank of the Jacobian for the model [Godfrey and DiStefano, 1985]. Another approach is to linearize the model about a suitable operating point (if such a point exists) and apply one of the many methods of analysis for linear systems. However, fewer identifiable parameter combinations than for the full nonlinear model may result and parameters of the nonlinear model may not even appear in the linearized one. Therefore unidentifiability of a linearized system does not necessary indicate that the original nonlinear model is unidentifiable.

Finally, a sensitivity analysis may be performed which gives an indication of possible identifiability problems. Such an analysis was performed in [Holmberg, 1982] on this exact model (4.1) and reveals a difficulty of distinguishing between the effects of μ and K_S from measurements of X and S during a batch experiment.

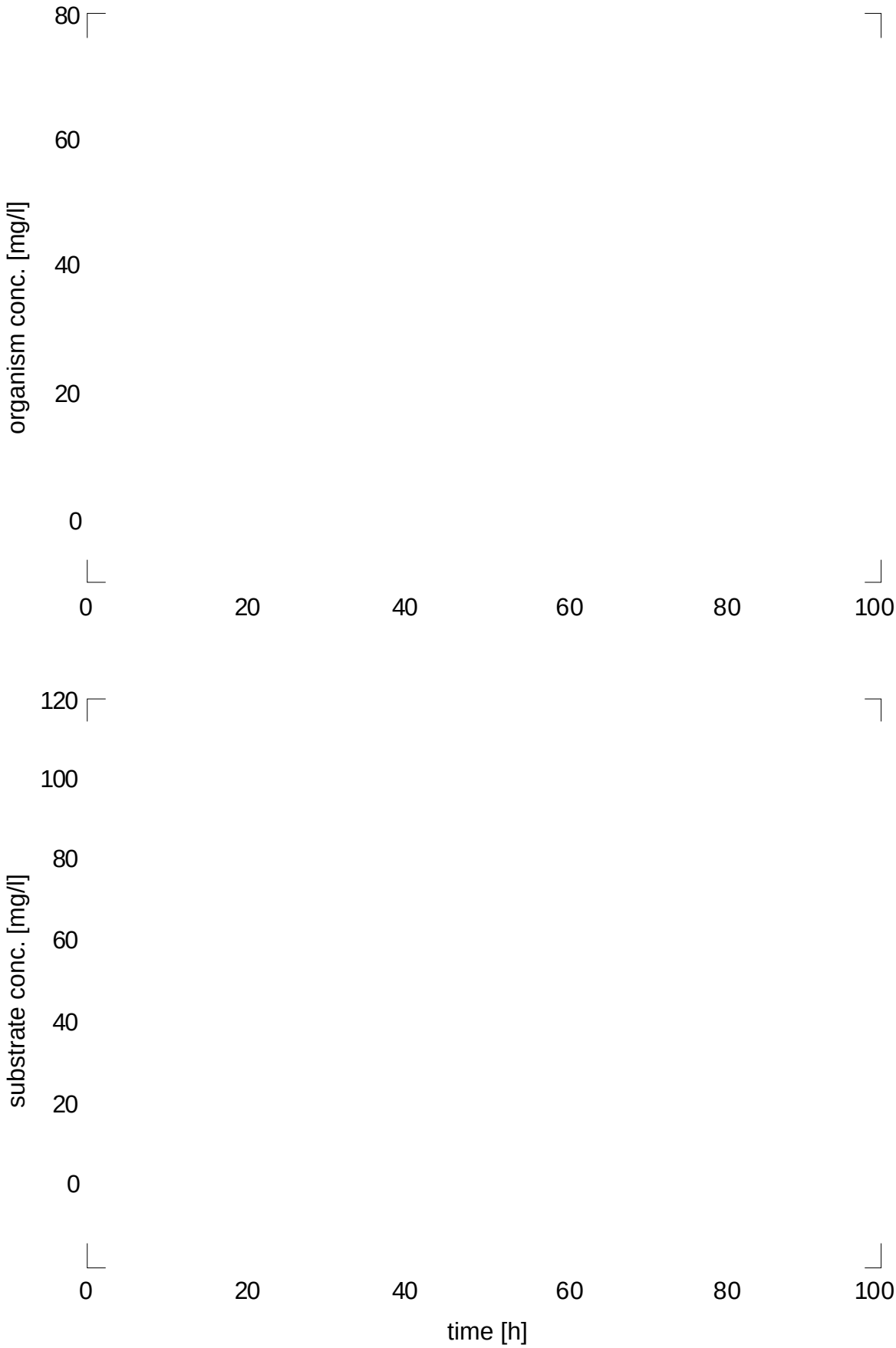

80
60
40
20
0
organism conc. [mg/l]
0
20
40
60
80
100
120
100
80
60
40
20
0
substrate conc. [mg/l]
0
20
40
60
80
100
time [h]

Figure 4.2 (Previous page) Simulation of system 4.1 with very different sets of model parameters compared to the noisy measurements of system 4.12 and 4.13 (dotted). For all simulations $X(0)$ = 2 mg/l and $S(0)$ = 100 mg/l. For the noisy system the parameter values are: μ = 6.0 day^{-1}, K_S = 10 mg/l, b = 0.48 day^{-1}, Y = 0.66.

Question: How is the system affected when noise is added?

Assume the same batch reactor system as earlier described but with noise added. The process noise (v) is Gaussian with a mean value of zero and a standard deviation of 10 % of the current state variable value. The measurement noise (ε) is specified in the same way together with an added Gaussian noise component with a mean value of zero and a standard deviation of 2 mg/l to reflect the difficulties of measuring very low concentrations accurately. It should be noted that the choosen noise level is not ver high when compared to real measurements (particularly for X, which is not possible to measure directly) and especially the selected mean noise value of zero, greatly simplifies matters.

Real measurements are normally affected by outliers, noise with non-zero mean and changing variance, trends, etc., and need to be adjusted with different kinds of filters which in turn affect the signals. These problems are, however, neglected and the following system results:

$$\begin{cases} \dfrac{dX}{dt} = \mu(S)\,X - bX + v_1 \\ \dfrac{dS}{dt} = -\dfrac{1}{Y}\mu(S)\,X + v_2 \end{cases} \qquad (4.12)$$

$$\begin{cases} X_{\text{measured}} = X + \varepsilon_1 \\ S_{\text{measured}} = S + \varepsilon_2 \end{cases} \qquad (4.13)$$

A simulation of the disturbed system (4.12 and 4.13) is shown in Figure 4.2. A theoretical analysis of how noise changes the behaviour of the system is not provided, instead computer simulations are used to point out some major problems.

The analytical approach described earlier is no longer applicable since the derivatives of the measured variables are no longer practically available due to the added noise. By simulating the above system for different sets of model parameters without noise, several sets can be found which model outputs are well within the noise deadband of the measurable signals. Some simulations with *extreme* parameter values (compared to the ones used under noisy conditions) are also presented in Figure 4.2, and are found to produce a similar system output. This indicates the importance of noise on the discussed system and the identifiability problem. The structure of the model leads to small variations of the measurable quantities even when the internal model parameters are changing significantly and the small output differences are easily dominated by noise.

To improve the outputs of the disturbed system (4.12 and 4.13) various means of filtering may be applied. An on-line filtered signal is always affected by an undesired time lag. If the signals are manipulated off-line this problem can be avoided. In this example a special lowpass filter with exactly zero phase distortion is used [Little and Shure, 1988] to transform the measurable data into a more applicable form (Figure 4.3). It should be noted that this is an ideal filtertype which can not be physically implemented and any real time filter would produce a poorer result. Since the applied noise in this example is so favourably chosen, filtering is not necessary for estimation purposes but is used to examplify some problems.

By applying an optimization algorithm to the filtered data, a set of model parameters which provides the best possible fit to the supplied data according to a certain criteria (normally minimizing the sum of squared residuals), can be found. In this example the two algorithms below were tested [Fletcher, 1987]:

- Nelder-Mead's algorithm (NM) - a simplex method which is very robust but requires a large number of iterations (Appendix D).
- Gauss-Newton's algorithm (GN) - a generalized least squares method with linear search, not as robust but faster convergence than *NM*.

In Figure 4.3 some results of the *NM* algorithm are shown, based on filtered noisy data. It should be noted that the found set of parameter values provide an excellent fit to the filtered data, though the actual values are quite far from the true set. Both methods are quite suited for off-line model calibration. The disturbed system sometimes show a rank deficiency if all four parameters are to be fitted simultaneously, which makes the results uncertain even though the algorithms often converge. By imposing a few simple restrictions on the algorithms the situation may be improved. Such restrictions include:

- predefining certain reasonable intervals within which the parameter values are to be found;
- realizing that measurements of concentrations can not have negative values;
- estimating the ratio μ/Y from equation (4.1b) instead of adjusting the two parameters separately.

The *GN* method is quite sensitive to the selected initial values of the parameters - significant differences between the true and guessed values causes divergence. The *NM* algorithm is more reliable for that type of problem which is why all presented results are based on this algorithm. However, it is difficult to determine whether the final parameter values are the best ones in a global sense or only a local optimum.

By running the algorithm for a large number of different initial values and comparing the final results, the global identifiability can be made plausible although it is not a theoretical proof. A very small change in the noise characteristics of the system or of the lowpass filter parameters will also greatly influence the best model parameter set found by the optimization algorithm. Especially the parameters μ and K_S are difficult to determine accurately and should therefore be considered to be uncertain (if not practically non-identifiable), whereas the estimates of b and Y seem to be more reliable (from this type of idealized batch experiment). Therefore situations where the parameters μ and K_S are used to characterize the biological process should be avoided if possible.

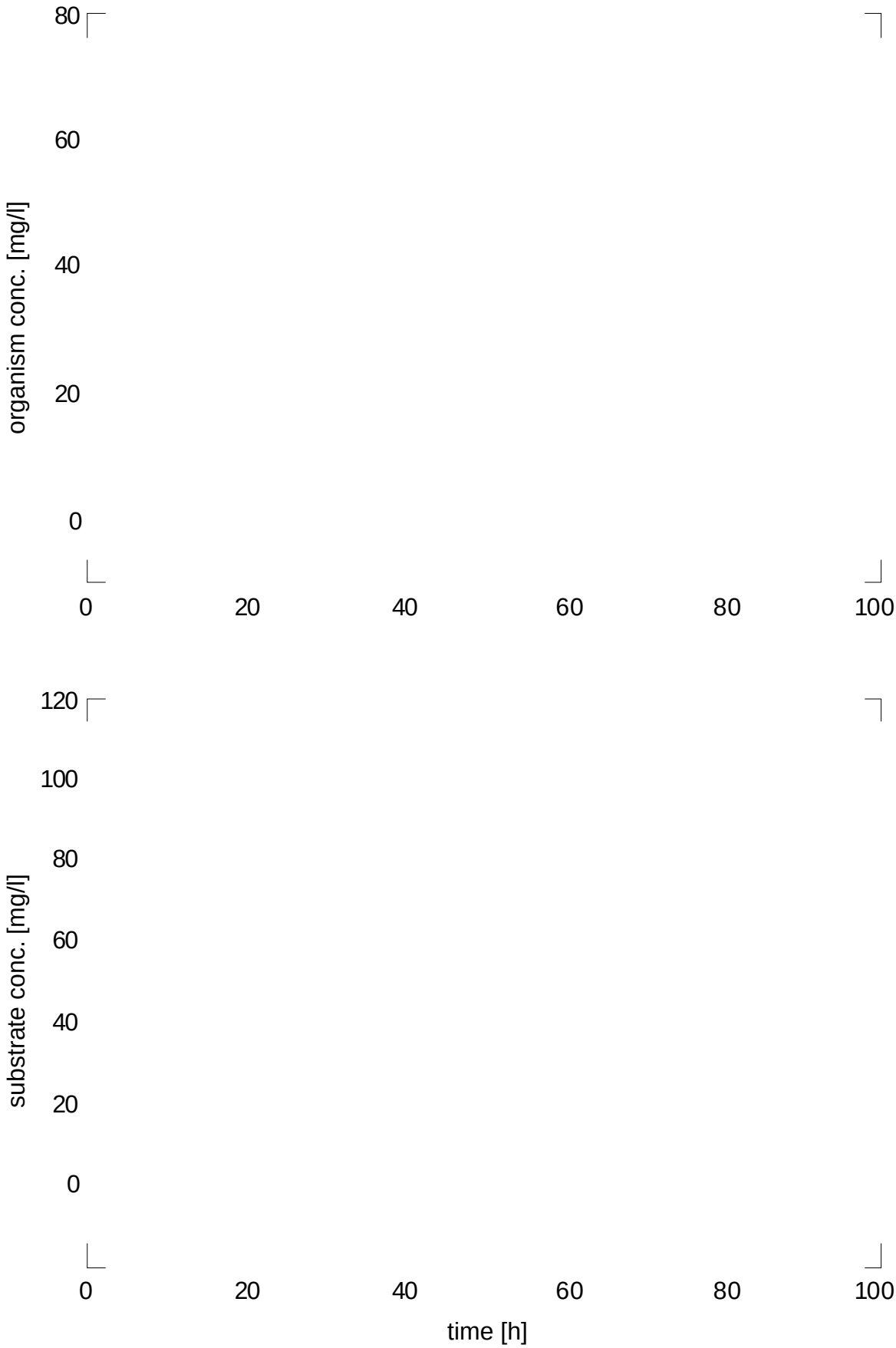

80
60
40
20
0
organism conc. [mg/l]
0
20
40
60
80
100
120
100
80
60
40
20
0
substrate conc. [mg/l]
0
20
40
60
80
100
time [h]

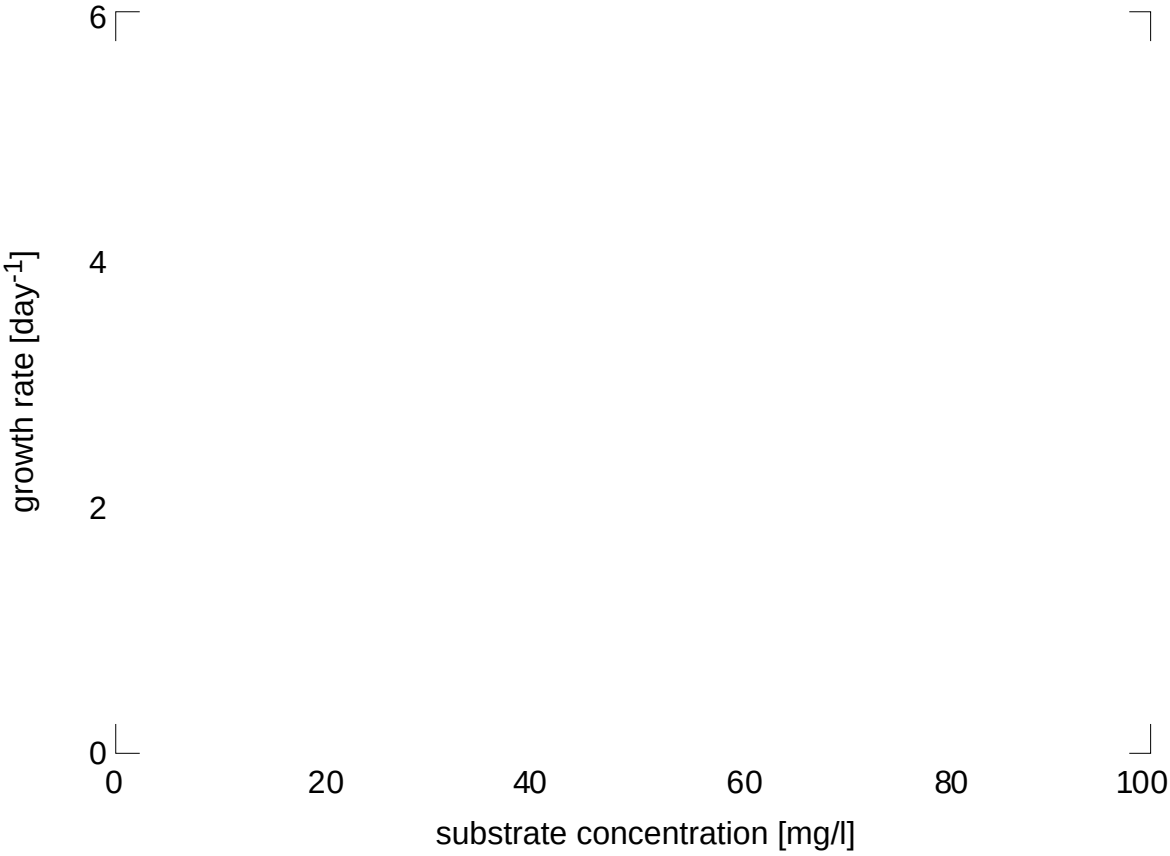
Figure 4.3 (Previous page) Simulation of system 4.1 with the optimum parameter set (solid) found with the *NM* algorithm based on filtered data, compared to the unfiltered (dotted) and filtered (dashed) noisy measurements. For all simulations $X(0) = 2$ mg/l and $S(0) = 100$ mg/l. For the noisy system the true parameter values are: $\mu = 6.0$ day^{-1}, $K_S = 10$ mg/l, $b = 0.48$ day^{-1}, $Y = 0.66$ and the found optimum parameter set is: $\mu = 6.6$ day^{-1}, $K_S = 19$ mg/l, $b = 0.47$ day^{-1}, $Y = 0.67$.

If no noise is added, both optimization algorithms detect the true parameter values practically independent of the selected initial values, i.e. the system is fully identifiable when perfect data is continuously available. This is not the case even under quite 'favourable' noise conditions.

The discussed example points out some of the problems which may appear, even for small models. Take into account that the model is actually only valid for single-substrate/single-organism batch processes (a traditional wastewater model involves several Monod type functions). The model parameters are usually dependent on other factors such as temperature, pH, time, etc. Measurements are seldom continuously available and determining the concentration of active biomass (X) is extremely difficult. It is easily recognized that the uncertainty of any estimated results are considerably increased when applied to real data.

Instead of trying to develop increasingly complex models an attempt will be made to take the simplification process one step further in order to develop a practically identifiable model. It is based on the fact that measurements are normally scarce and uncertain. A low complexity model is easier to calibrate (even possible to calibrate on-line) which may lead to more reliable results even though the biological and physical interpretation of some model parameters may be lost in the process.

One possible way to adjust the small model in this example is to use a simplified Monod formulation of the form,


6
4
2
0
growth rate [day-1]
0
20
40
60
80
100
substrate concentration [mg/l]

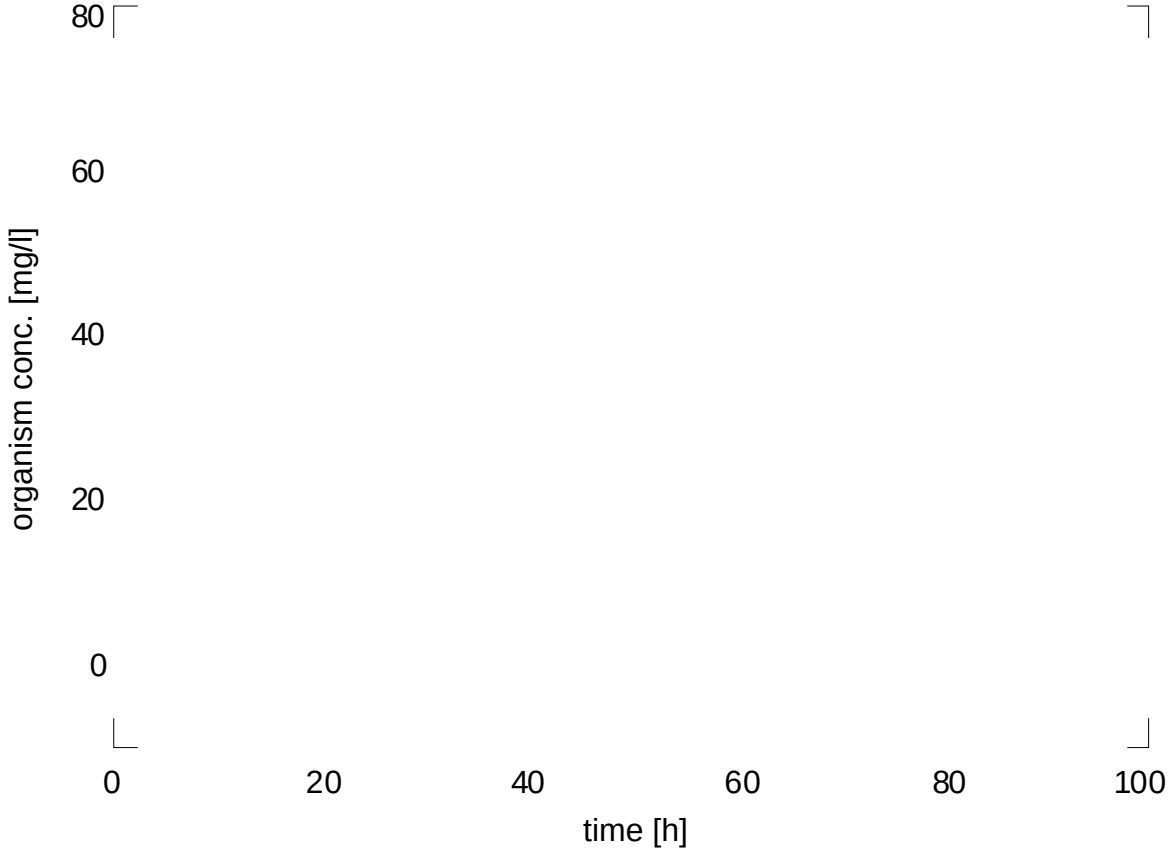

80
60
40
20
0
organism conc. [mg/l]
0
20
40
60
80
100
time [h]

Figure 4.4 (Previous page) a) Illustration of the standard (1.1, dashed) and simplified (4.14, solid) Monod equations. The parameters used are: μ = 6.1 day^{-1}, K_S = 11.6 mg/l, r = 0.22 l/(mg day), S_{sat} = 24.0 mg/l. b) Simulation of system 4.1 using the simplified (solid) and the standard Monod equations (dashed) with the optimum parameter sets found with the *NM* algorithm based on unfiltered, noisy data (4.12 and 4.13, dotted) with the standard Monod formulation. For all simulations $X(0)$ = 2 mg/l, $S(0)$ = 100 mg/l, and the noisy system uses: μ = 6.0 day^{-1}, K_S = 10.0 mg/l, b = 0.48 day^{-1}, Y = 0.66.

$$\mu(S) = \begin{cases} r\,S & \text{if } S < S_{sat} \\ r\,S_{sat} & \text{if } S \geq S_{sat} \end{cases} \qquad (4.14)$$

where: r = reaction rate factor [l·(mg day)$^{-1}$];
S_{sat} = growth saturation concentration [mg/l].

The expression is reduced to a first order rate expression for low substrate concentrations (the normal case) and a zero order expression for high substrate concentrations. Since the model in this case is to be used for a batch experiment both cases have to be included. It means that the number of parameters is not reduced (both r and S_{sat} have to be estimated), only the structure is simplified. However, if the model is applied to WWT plants for municipal wastewater, the substrate concentrations are usually sufficiently low to motivate the use of a first order reaction only which would provide a significant simplification.

In Figure 4.4a, a comparison of the traditional (1.1) and the simplified (4.14) Monod expressions are given. The model (4.1) is simulated with the simplified and standard Monod equations with the optimum parameter sets found with the *NM* algorithm based on noisy data and is also compared to the original disturbed system (4.12 and 4.13) using the standard Monod expression. The result for the organism concentration is presented in Figure 4.4b. The result is equally good for the substrate concentration. It is apparent that the results of the simplified model are very similar to that of the more complex model. If only a few measurements are available, the parameter r is much easier to identify than μ and K_S.

For the parameter estimations examplified above, the number of samples have been considered to be continuously available and distributed over the entire interesting range of substrate concentration. This is, however, often not the case. In order to show how *one* of these factors influences the results, two series of estimations are performed:

- sampling rate = 10 hour^{-1} (case A);
- sampling rate = 1 hour^{-1} (case B).

The only difference is the number of samples; in case A the sampling rate is ten times per hour and in case B one time every hour (which is more realistic). In both cases the standard Monod equation is used in combination with the batch model to generate three data sets on which the estimations are based - (1) no noise; (2) including noise (distributed as earlier described); (3) including noise and lowpass filtering (as in Figure 4.3).

The *NM* algorithm is then applied for the estimations and optimum sets of parameters are found for both the standard and simplified Monod equations based on the above data series. Various initial estimates are used as a rough investigation of global identifiability and the results are presented in Table 4.1. The impact of the number of samples is obvious when noise is present as well as the effects of filtering. It should be noted that the model fit to the available data is satisfactory in all cases (in the least squares sense) even though the estimated parameter values differ significantly. It should also be noted that the convergence rate for Y and b is in all cases much higher than for the other model parameters. In order to achieve good estimates of μ and K_S, several hundred iterations are required.

In this example it has been shown how difficult it is to globally determine the parameters μ and K_S of the Monod function when noise is added to the system. The difficulty is mainly due to nonlinear parameter interactions, especially obviuos when measurable data is not continuously available.

	Standard Monod	initial estimates				final estimates			
		$\hat{\mu}$	K_S	b	Y	$\hat{\mu}$	K_S	b	Y
Case A (1000 samples)	no noise	5.0 6.0 7.0	5 10 20	0.4 0.5 0.6	0.5 0.65 0.8	6.00 6.00 6.00	10.0 10.0 10.0	.480 .480 .480	.660 .660 .660
	including noise	5.0 6.0 7.0	5 10 20	0.4 0.5 0.6	0.5 0.65 0.8	6.13 6.13 6.13	11.6 11.6 11.6	.472 .472 .472	.664 .664 .664
	including noise and lowpass filtering	5.0 6.0 7.0	5 10 20	0.4 0.5 0.6	0.5 0.65 0.8	6.65 6.65 6.65	18.7 18.7 18.7	.473 .473 .473	.665 .665 .665
Case B (100 samples)	no noise	5.0 6.0 7.0	5 10 20	0.4 0.5 0.6	0.5 0.65 0.8	6.00 6.00 6.00	10.0 10.0 10.0	.480 .480 .480	.660 .660 .660
	including noise	5.0 6.0 7.0	5 10 20	0.4 0.5 0.6	0.5 0.65 0.8	7.07 7.07 7.07	23.2 23.2 23.2	.474 .474 .474	.677 .677 .677
	including noise and lowpass filtering	5.0 6.0 7.0	5 10 20	0.4 0.5 0.6	0.5 0.65 0.8	10.7 10.7 10.7	74.4 74.4 74.4	.488 .488 .488	.687 .687 .687
	Simplified Monod	r	S_{sat}	b	Y	r	S_{sat}	b	Y
Case A (1000 samples)	no noise	0.15 0.25 0.4	10 25 50	0.4 0.5 0.6	0.5 0.65 0.8	.224 .224 .224	23.7 23.7 23.7	.481 .481 .481	.660 .660 .660
	including noise	0.15 0.25 0.4	10 25 50	0.4 0.5 0.6	0.5 0.65 0.8	.222 .222 .222	24.0 24.0 24.0	.472 .472 .472	.663 .663 .663
	including noise and lowpass filtering	0.15 0.25 0.4	10 25 50	0.4 0.5 0.6	0.5 0.65 0.8	.161 .161 .161	33.4 33.4 33.4	.474 .474 .474	.666 .666 .666
Case B (100 samples)	no noise	0.15 0.25 0.4	10 25 50	0.4 0.5 0.6	0.5 0.65 0.8	.222 .222 .222	24.0 24.0 24.0	.481 .481 .481	.660 .660 .660
	including noise	0.15 0.25 0.4	10 25 50	0.4 0.5 0.6	0.5 0.65 0.8	.141 .141 .141	39.2 39.2 39.2	.476 .476 .476	.679 .679 .679
	including noise and lowpass filtering	0.15 0.25 0.4	10 25 50	0.4 0.5 0.6	0.5 0.65 0.8	.097 .097 .097	59.3 59.3 59.3	.491 .491 .491	.690 .690 .690

Tabel 4.1 Optimizations of the standard and simplified Monod equations.

Signal filtering also affects the estimations in a dramatic way. Several parameter sets provide a close fit to the measurable data but it is difficult to determine whether the found parameters are really the true ones when interpreted in a biological sense. Because of this, it is motivated to use simplified models when possible, in order to improve the identifiability and verifiability of the investigated system.

4.2 Model Development and Analysis

The basis for any reliable mathematical model development is a thorough understanding and experience of the true processes involved. To some small extent a stochastic (empiric) model may be achieved simply by using a fast computer, the proper software and a sufficiently large amount of experimental data. This is the area of system identification, which can be roughly described as multi-dimensional curve-fitting procedures.

True physical (deterministic) modelling is, however, an analytical approach where basic laws from physics, chemistry, etc., are used to describe the behaviour of a process. Based on this process knowledge a model suited for the required purposes is hypothesised. Its structure may then be analysed using the available tools (pole analysis, frequency analysis, sensitivity analysis, etc.) and stepwise further tested, adjusted, and verified.

One of the main difficulties when developing a model is usually to determine which reactions are the most significant ones and describe these in a simple, yet comprehensive manner. A good deterministic model should realistically mimic the true dynamic behaviour of the process but still contain a minimum number of equations, parameters, and variables without losing the physical interpretation of those.

Overall considerations

Activated sludge systems are described by deterministic mathematical models based on mass balance equations. These equations relate the changes of the state variables of the system (concentrations) to transport and transformation processes. Transport processes are often especially characteristic for the design of a system (reactor configuration, distribution of influent, mixing, excess sludge removal, etc.) but they leave the chemical structure of all materials unchanged. Transformation processes are governed by the local conditions and involve the change of the chemical structure of the components. A mass balance equation for a single component within a defined system boundary can be schematically described as:

$$input - output + reaction = accumulation$$

In the IAWPRC model (Appendix B), a total of seven dissolved and six particulate components are used to characterize the wastewater and the activated sludge (compare Table 2.1). These include seven fractions of organic matter (measured as COD) of which two are different types of biomass and four fractions of nitrogen. Eight transformation processes are considered which include a total of 19 stochiometric and kinetic model parameters (Table 2.2). The 'Activated Sludge Model No. 1' is particularly useful for the prediction of:

- biological degradation of organic material and denitrification;
- nitrification;
- the distribution of oxygen consumption along a 'plug flow' type reactor and in the course of diurnal variations;
- sludge production;
- variation of effluent quality under dynamic loading conditions.

In order for the above to be completely true there is normally a need to combine the biological model with accurate models for other parts of a wastewater treatment plant, for example the settler and the oxygen transfer mechanisms. In this work, however, the biological processes will be focused on and problems

due to other process units will not be regarded. Although the IAWPRC model comprises much of the current expert knowledge of the biological reactions (except biological phosphorus removal) in a WWT plant - and does so in a 'fairly simple' way - a number of drawbacks exist. Some of those are summarized below.

- Lacking identifiability - non-unique solutions for model parameters exist (e.g. the model may produce practically identical results for different model calibrations).
- Lacking verifiability - certain state variables and parameters are not directly measurable and therefore it is very difficult to experimentally verify the biological interpretations.
- Poor understanding and knowledge about some of the applied processes (e.g. the hydrolysis mechanism).
- Very troublesome practical characterization of the incoming wastewater, although essential for the model behaviour.
- Difficulties to estimate and update the varying model parameters (functions of time, pH, load, temperature, etc.) on-line.
- Troublesome nonlinearities (Monod, switching functions, etc.).
- Too high complexity for on-line control purposes.
- Expert knowledge required in order to understand all model interactions (complicated cause-effect relationships).
- Need for highly sophisticated instrumentation and laboratory facilities for calibration and verification purposes.

The goal of this work is to approach some of the problems listed above and develop reduced order models which can adequately represent both carbonaceous and nitrogeneous activities with a minimum number of variables and model parameters. In the same way as the IAWPRC model is considered too complex for on-line use, it is obvious that the model in the simple example (4.1) is overly simplified for describing the biological reactions of a WWT plant. The final result will be a compromise between these two extremes.

Instead of applying general methods for model reduction it is important to make use of physical insight of the process

behaviour. This means that the significant reactions should still be described in a physically reasonable manner, i.e. a mechanistic model structure should be maintained whenever possible. Therefore the selected basis to be used for the model simplifications is the standard IAWPRC model due to its widespread use, acceptance, and mechanistic structure.

Measurable variables

Models for on-line control purposes must naturally be based on quantities and variables which are possible to measure on-line. This is especially important as the model parameters are not constants but vary with time and operational conditions. Results from laboratory experiments and bench scale tests should of course be used for verifications and further improvements of the model output whenever possible but not be vital or even essential for the basic reliability and performance of the model.

The quantities which are possible to measure and quantify in the activated sludge process have been discussed in section 2.5. New instruments and measurement technologies are also constantly being developed. For this work the following set of quality variables are assumed to measurable on-line:

- biodegradable organic matter concentration;
- ammonia nitrogen concentration;
- nitrate nitrogen concentration;
- flow rates.

In some cases measurements of the oxygen uptake rate (OUR) or respiration rate are considered to be available as well (by methods discussed in section 2.5).

Cost, precision, accuracy, sensitivity, repeatability of the above measurements are not considered nor are the practical aspects of where to place the sensors, how data should be transfered to the computer systems, etc. All these questions are of great importance at a later stage of the work. Until the more basic and principal questions about the models have been thoroughly investigated the measurement quality questions are overlooked.

Simplifying assumptions

The assumptions for simplifying models of the activated sludge process are from a physical and biological viewpoint mainly based on a discussion how the following components are treated:

- dissolved oxygen;
- organic matter;
- nitrogen;
- microorganisms.

For the reduced models, measurements of the dissolved oxygen (DO) concentration is not considered although it is usually regarded as the most reliable on-line instrument for the activated sludge process. This is because the oxygen concentration is excluded as a state variable. It is assumed that the DO is controlled on a routine basis, so that corresponding growth expressions become independent of DO variations. Still the DO mass balance contains plenty of useful information. It is the basis for the estimation of the oxygen uptake rate, which is recognized as a fundamental information for further control.

As a result, the models describing the DO, oxygen transfer rate, blowers, etc., make up an important but separate modules of the whole control structure as shown in Figure 4.1. These sub-models need to be synchronized with the overall model of the plant. This approach makes it possible to separate the biological model from the oxygen model on the first level of control. It also allows a clearer borderline to be applied between the anoxic (with nitrate as the terminal electron acceptor) and aerobic (with oxygen as the terminal electron acceptor) environments from a modelling point of view. The existing DO control is assumed to provide a sufficient amount of oxygen to the aerobic part while minimizing the oxygen concentration in the anoxic reactor.

This type of separation allows models of lower complexity but may seem too rough a simplification. It must however be noted that the approach does not imply that the biological models are insensitive to the DO concentration. The effects are rather combined with other inhibitory circumstances and reflected as

changes in the estimated growth rate factors whereas the IAWPRC model applies a switching function to single out the effects of various DO concentrations.

The assumption of a constant DO concentration is valid only if the reactors are truly completely mixed. The situation is often different at real WWT plants. Experiments show that an aerobic (assumed completely mixed) reactor with a DO probe in the center of the tank and connected to a control system with a DO setpoint of 2 mg/l may have a true DO concentration of 0.5 mg/l close to the influent flow and 5 mg/l near the effluent flow. The expected value exists only close to the probe.

The reasons for this difference are mainly due to improper mixing and varying concentrations of available sustrate. There is no easy way of modelling these effects in a reasonable manner which can be verified (for example turbulence phenomenas are nowadays considered a chaotic process [Stewart, 1990]). A possibility often applied to improve matters is to increase the number of reactors (each completely mixed) to reduce the discrepancies, but practically, the number of reactors in a model is limited to approximately ten due to high complexity and computational problems. Another possibility is to use partial differential equations to describe the spatial distribution (as well as the distribution in time) of the concentration variations for all components including DO. Such an approach would however result in a highly complex model.

Compared to these problems, the assumption of truly anoxic and aerobic reactors is only a small additional simplification. The major simplification is actually to assume completely mixed reactors. In the IAWPRC model it is also possible to include the DO concentration in the model transport terms which may improve matters but many difficulties still remain.

The representation of organic matter presents the second considerable difference compared to the IAWPRC model. In the IAWPRC model four fractions of organic matter are considered:

- soluble inert organic matter (S_I);
- readily biodegradable substrate (S_S);
- particulate inert organic matter (X_I);
- slowly biodegradable substrate (X_S).

All these fractions are replaced by a single one in the reduced models (X_{COD}) which is considered to be made up of all biodegradable organic matter and also assumed to be measurable. The approach can be motivated in several ways.

The two inert fractions are not important from a biological point of view. The S_I fraction passes right through a WWT plant without any effect and the X_I fraction - together with a special fraction describing the particulate products arising from biomass decay (X_P) - is used to predict the total amount of sludge in the system in order to determine the wastage and recirculation rates. On the other hand, the two biodegradable fractions are of the utmost importance for the biological reactions. S_S is considered to be directly available for the microorganisms while X_S first has to be enzymatically broken down into S_S (the hydrolysis mechanism) before the organisms can use it for metabolism.

The hydrolysis process is, however, not very well known. The IAWPRC description of it is quite complex but still a simplification of the true reaction. Due to the uncertainty of the reaction it is not included in the reduced models.

Another reason for lumping the organic matter together is the difficulty of measuring the S_S and X_S fractions separately. In a laboratory scale experiment it is possible to monitor the oxygen uptake rate of a small batch reactor and thereby determining an average of the two fractions. To do this on-line at a full scale plant is more difficult. In practice, COD measurements on filtered samples of the wastewater is often considered to be equal to the amount of S_S and COD_{total} - $COD_{filtered}$ is used as the X_S fraction. Since there is no evidence that all soluble biodegradable matter is readily biodegradable and all particulate biodegradable matter is slowly biodegradable, this is not satisfactory. A further

complication is the fact that S indicates soluble and X indicates particulate matter in the IAWPRC nomenclature.

The organic matter which is received by a WWT plant includes all kinds of different molecular structures. Different organisms deal with different substrates in different time scales which makes it probable that an entire set of biodegradation processes with time constants ranging from fast to slow biodegradability is at work here. This has been made even more plausible as it has been recently proposed by some researchers to extend the number of fractions of organic matter to include very slowly biodegradable substrate and directly available substrate (in order to explain biological phosphorus removal).

Since there is no apparent upper limit to the number of substrates which would really need to be included, the opposite solution is suggested in this work. This is why the reduced models do not take rapid uptake phenomena into consideration. Rather it makes some averaging of biosorption and growth by combining soluble and stored organic substrate. Consequently fast dynamics (of the order less than an hour) is neglected. Together with the earlier discussed way of modelling the DO concentration, these simplifications make the models less stiff, i.e. the ratio between the smallest and the largest dynamical time constants is reduced. The complexity and the number of model parameters are naturally also significantly reduced and the possibility for achieving an identifiable model structure is increased.

The third major difference between the IAWPRC model and the reduced models concerns the nitrogen. In the IAWPRC model four fractions of nitrogen are considered:

- nitrate and nitrite nitrogen (S_{NO});
- ammonia nitrogen (S_{NH});
- soluble biodegradable organic nitrogen (S_{ND});
- particulate biodegradable organic nitrogen (X_{ND}).

The only two nitrogen fractions included in the reduced order models are the nitrate nitrogen and the ammonia nitrogen, which both are assumed to be measured on-line. The reasons for this is

firstly to reduce the model complexity and number of parameters. Secondly, the two organic nitrogen fractions S_{ND} and X_{ND} are mainly used to describe the formation of S_{NH} through hydrolysis and ammonification (the knowledge of both mechanisms is uncertain). In the reduced models, ammonia nitrogen is assumed measured and therefore its formation mechanism is not considered to be crucial for control purposes.

The two types of organisms from the IAWPRC model (and many others) are maintained in the reduced models:

- active heterotrophic biomass ($X_{B,H}$);
- active autotrophic biomass ($X_{B,A}$).

Heterotrophs are considered to grow under both anoxic and aerobic conditions whereas Autotrophs only grow in an aerobic environment. A kind of death-regeneration principle (section 2.2) is also applied for the organisms but in a somewhat different way. The decayed biomass is considered to form into available COD and ammonia nitrogen directly. In the IAWPRC model the decay material is suggested to be partly inert and partly formed into X_S and X_{ND} which after hydrolysis and ammonification become available as S_S and S_{NH}.

The reduced model

The simplifications discussed until now has reduced the number of state variables to five compared to the twelve variables of the IAWPRC model (alkalinity is not considered). The reaction mechanisms for hydrolysis of entrapped organics, hydrolysis of entrapped organic nitrogen, and ammonification of soluble organic nitrogen have also been left out, mainly due to measurement problems, uncertainties of the actual processes, and reduction of the overall model complexity.

It is possible to go one step further. As illustrated by the example in the previous section there are good reasons to reconsider the parametrization of the Monod and the similar switching functions. The basic idea is to approximate the Monod function by straight lines (Figure 4.4), i.e. a first order reaction followed

by a zero order reaction and observe that during most normal operations of WWT plants receiving municipal wastewater there is only need for the first order reaction.

Due to the assumed existing DO control, the DO influence on the switching functions is constant. The switching functions are regarded solely as functions describing growth limitation due to DO or various kinds of substrate. Therefore the estimated parameter of the first order rate equation includes both maximum specific growth rate and possible limitation by DO, nitrate, etc. The switching functions are removed.

The differential equations for the first reduced order model (model A) of the activated sludge process can now be formulated. It describes carbonaceous oxidation as well as nitrification and denitrification according to the simplifications discussed above. Altogether three summary reaction processes are proposed to describe the anoxic environment - growth of Heterotrophs, decay of Heterotrophs, and decay of Autotrophs - and four parameters have to be estimated - r_H, Y_H, b_H, and b_A - preferably on-line. Under anoxic conditions the following model is suggested (see also Appendix C):

$$\frac{dX_{COD}}{dt} = -\frac{1}{Y_H} r_H X_{COD} X_{B,H} + b_H X_{B,H} + b_A X_{B,A} \qquad (4.15)$$

$$\frac{dS_{NH}}{dt} = - i_{XB} \left(r_H X_{COD} X_{B,H} - b_H X_{B,H} - b_A X_{B,A} \right) \qquad (4.16)$$

$$\frac{dS_{NO}}{dt} = - \frac{1 - Y_H}{2.86\, Y_H} r_H X_{COD} X_{B,H} \qquad (4.17)$$

$$\frac{dX_{B,H}}{dt} = \left(r_H X_{COD} - b_H \right) X_{B,H} \qquad (4.18)$$

$$\frac{dX_{B,A}}{dt} = - b_A X_{B,A} \qquad (4.19)$$

In an aerobic environment four main reaction mechanisms are proposed - growth of Heterotrophs, growth of Autotrophs, decay of Heterotrophs, and decay of Autotrophs - and six parameters

need to be updated - r_H, r_A, Y_H, Y_A, b_H, and b_A. This suggests the following model for aerobic conditions (see also Appendix C):

$$\frac{dX_{COD}}{dt} = -\frac{1}{Y_H} r_H X_{COD} X_{B,H} + b_H X_{B,H} + b_A X_{B,A} \qquad (4.20)$$

$$\frac{dS_{NH}}{dt} = -i_{XB}\left(r_H X_{COD} X_{B,H} - b_H X_{B,H} - b_A X_{B,A}\right) - \left(i_{XB} + \frac{1}{Y_A}\right) r_A S_{NH} X_{B,A} \qquad (4.21)$$

$$\frac{dS_{NO}}{dt} = \frac{1}{Y_A} r_A S_{NH} X_{B,A} \qquad (4.22)$$

$$\frac{dX_{B,H}}{dt} = \left(r_H X_{COD} - b_H\right) X_{B,H} \qquad (4.23)$$

$$\frac{dX_{B,A}}{dt} = \left(r_A S_{NH} - b_A\right) X_{B,A} \qquad (4.24)$$

where: r_H = reaction rate factor for Heterotrophs [l·(mg day)$^{-1}$];
r_A = reaction rate factor for Autotrophs [l·(mg day)$^{-1}$];
Y_H = yield factor for Heterotrophs;
Y_A = yield factor for Autotrophs;
b_H = decay rate coefficient for Heterotrophs [day^{-1}];
b_A = decay rate coefficient for Autotrophs [day^{-1}];
i_{XB} = mass N/mass COD in biomass.

The factor 2.86 (see eq. 4.17) in the stochiometric coefficient for anoxic growth of heterotrophic biomass is the oxygen equivalence for conversion of nitrate nitrogen to nitrogen gas included to maintain consistent units. The parameter i_{XB} is considered to be a known constant with a value which equals 0.086 as suggested by the IAWPRC task group [Henze et al., 1987a]. The other parameters are considered unknown and need to be identified for proper model performance.

Note that the model parameter values are not assumed to be the same under anoxic and aerobic conditions and should therefore be separately estimated.

For purposes of comparison, the oxygen uptake rate is in some cases considered to be measurable (as discussed in section 2.5). It is basically modelled in the same manner as in the IAWPRC model apart from the simplified Monod equation and the lack of switching functions:

$$OUR = \frac{1 - Y_H}{Y_H} r_H X_{COD} X_{B,H} + \frac{4.57 - Y_A}{Y_A} r_A S_{NH} X_{B,A} \quad (4.25)$$

The factor 4.57 (see eq. 4.25) in the stochiometric coefficient for aerobic growth of Autotrophs is the theoretical oxygen demand associated with the oxidation of ammonia nitrogen to nitrate nitrogen.

Further simplifications

While maintaining the basic structure of 'model A' it is possible to impose some further simplifications in order to improve the identifiability. It is not unrealistic to assume the decay rates for Heterotrophs and Autotrophs to be equal under both anoxic and aerobic conditions. This will reduce the total number of parameters to be estimated from ten to eight.

Taking this approach one step further, b_H and b_A can be lumped together into one single decay rate for all conditions, b, reducing the number of unknown parameters to seven. The simplification is not crucial - the assumption of only two kinds of microorganisms representing many dozens of species is a more significant simplification. A more practical reason for such a suggestion is the difficulty to estimate decay rates during normal plant operations.

In the small example of section 4.1 the decay rate factor was easily estimated from a batch experiment. This is because during the final stage of such an experiment, the decay rate is the sole factor to influence the behaviour of the reaction (when all the substrate has been consumed). Its identifiability is therefore enhanced. From continuous reactors, however, the effects from growth and decay rates are difficult to separate, especially when

the oxygen consumption is not simultaneously monitored. This problem will be more closely investigated in Chapter 5.

The second proposed reduced order model (model B; see Appendix C) which will be investigated, contains the same basic equations as 'model A' (4.15 - 4.24) but with all parameters b_H and b_A replaced by a single decay rate, b, which is assumed to be identical under both anoxic and aerobic conditions.

In much the same way as outlined above, identifiability difficulties may arise when trying to estimate the yield and reaction rate factor simultaneously. An increased value for the yield and reaction rate factor may outbalance each other with respect to the substrate conversion rate. In the example (section 4.1), the situation was improved because both the substrate and organism concentration were assumed to be measurable plus the fact that a batch experiment excites all modes of the system. Unfortunately this is not true under normal plant operation. For this reason 'model B' might be even further reduced by assuming the same value for the heterotrophic yield, Y_H, under both anoxic and aerobic conditions (model C). This final simplification leaves a total of six model parameters to be estimated. This model will, however, not be further investigated.

As for the reaction rate factor of the Heterotrophs, r_H, it does not seem realistic to assume this parameter to be the same under both anoxic and aerobic conditions. Experiments have shown that either is only a fraction of the heterotrophic biomass able to function with nitrate as the terminal electron acceptor or is the maximum specific growth rate lower under anoxic conditions [Batchelor, 1982]. The reaction rate factor must therefore be separately identified in order to take these effects into account. The minimum realistic number of parameters to be updated would therefore be the six suggested for 'model C' - r_H (anoxic and aerobic), r_A, Y_H, Y_A, and b.

Summary of reduced order features

A basic reduced order model for the activated sludge process has been developed and analysed based on several assumptions and criteria; model complexity, model and control structure, measurable quantities, simplicity of measurements, wastewater characterization, significance and reliability of various reaction mechanisms, identifiability/verifiability, possibility for on-line applications, etc. Two slightly modified and further simplified versions of the model have been proposed due to more practical estimation problems.

Such simple models will naturally not show all the intricate details of the more complex models since many reactions and variables have been lumped together. The basic biological interpretations of the models, however, have been maintained when possible. The presented models serve a major purpose as experimental platforms in order to investigate how far the model reduction idea can be pursued without losing the possibility of adequately predicting the main phenomena of the true processes while simultaneously gaining the possibility to determine unique estimates and perform on-line model calibration.

4.3 Model Verification Methods

Model verification is concerned with determining whether an obtained model is adequate or not. It should always be seen in the light of the intended purpose of the model. Therefore the ultimate verification can only be performed by using the model in practice and monitoring the results. However, there are a number of ways which can be used to test if the model is likely to describe the system in a proper way (complexity vs. flexibility), before using the model effectively. The best one is often to use plots and common sense while others are usually based on statistical tests of the prediction errors (residuals). These may include:

- tests of whiteness of the residuals and of independence of the prediction errors and the input;
- tests for detecting a too complex model structure, for example by means of pole-zero cancellations and singular information matrices;
- tests on the values of the loss functions corresponding to different model structures, for example the χ^2-test [Söderström and Stoica, 1989].

It is also of importance to check the applicable *a priori* assumptions. A number of such tests are listed below.

- Test of linearity. If possible the experiment should be repeated with another amplitude (or variance) of the input signal in order to verify for what operating range a linear model is adequate.
- Test of time invariance. A convenient way of testing time invariance of a system is to use data from different experiments. The parameter estimates are determined in the usual way from the first set. Then the model output is computed for the second set of data using the parameter estimates obtained from the first set. If the process is time invariant the model should explain the process data equally well for both sets.
- Test for the existence of feedback.

Unfortunately, the activated sludge process is known to be both nonlinear and time variant which makes the reproducability of the system poor. Under normal operating conditions, a WWT plant is further exposed to several alternative feedbacks. All these factors create difficult problems when trying to apply traditional methods of verification and identification.

Overall considerations

Writing a mathematical model is generally easier than verifying it. The two aspects of modelling are however not meant to be separated and it is better to use an iterative model building

approach like the one principally outlined in Figure 4.5. Otherwise the tiring and costly verification work is often neglected and any resulting mathematical model must in such a case be treated with scepticism.

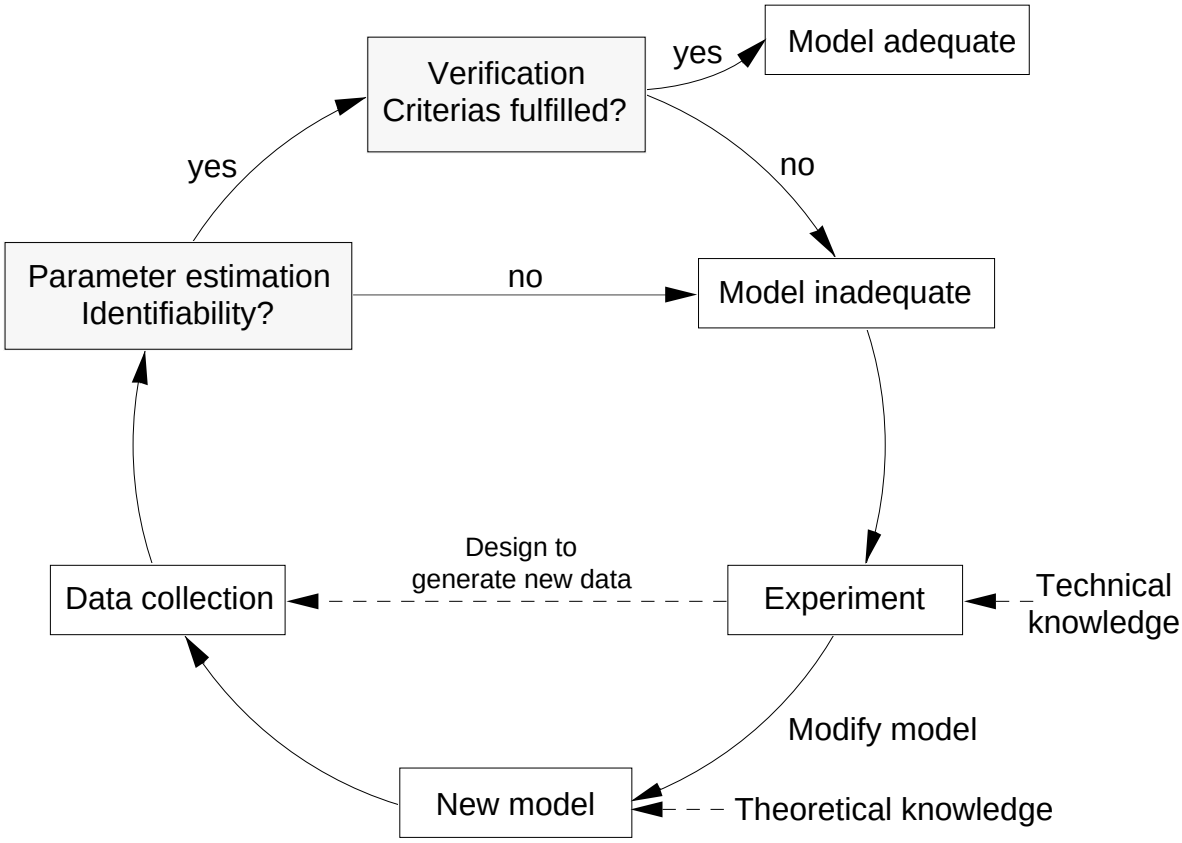


Figure 4.5 Adaptive model building approach.

At this early stage of the work, no verifications have been carried out on real data but only on simulated data mainly based on the IAWPRC model. Therefore it is not a true verification but still sufficient to reach some principle conclusions on the behaviour of the reduced order models. Future examinations must naturally also include real WWT plant data. Due to this fact the main verification principle applied has been based on comparisons of model outputs from the IAWPRC and simplified models to determine if they incorporate the same dynamical phenomena - both quantitatively and qualitatively - of importance in the actual time scales when subjected to the same type of influent conditions. A number of such comparisons will be presented and discussed in Chapter 5.

As for state and parameter estimation and the crucial question of identifiability (globally or locally) a more stringent analysis has been performed. It should be noted that such an investigation is of the utmost importance when determining the suitability of the reduced order models. The question of identifiability and verifiability should not be treated separately but be considered as two necessary aspects for reaching a credible conclusion. A highly complex model may very well be possible to fit excellently to a set of experimental data but is of limited use for predictions if the model parameters are not uniquely identifiable because its validity is restricted to the exact conditions under which it was calibrated.

Different sets of parameters may produce very similar results under certain conditions and then diverge completely when exposed to slightly different conditions. If the parameters also have a physical interpretation the problem is further emphasized. On the other hand, a model which is globally identifiable but does not provide a result which reasonably mimics the true process output is obviously of little practical use.

The result of the estimation problem depends on how the problem is formulated. The accuracy and reliability of an obtained model is significantly influenced by the amplitude and frequency content of the input signal. In order to identify a dynamical system it has to be sufficiently excited since it is from the variations that the most information are made available. Other factors which play important roles are the experimental conditions, available measurements, noise conditions, etc. There are also many different ways to organize the computations. Consequently, the number of available identification methods are large.

One broad distinction is between *on-line* methods and *off-line* methods. The on-line methods give estimates recursively as new measurements are obtained and are the only alternative if the identification is going to be used in an adaptive controller or if the process is strongly varying with time. In many cases the off-line methods give estimates with higher precision and are more reliable, for instance in terms of convergence. Depending on the time scale and amplitude of the parameter changes, an off-line

approach can provide a good result even for time varying systems if the estimation is repeated at sufficiently short intervals (the result is an optimum average parameter set over a specific time). The property of identifiability was defined in a strict mathematical way in section 3.4. It is related to the problem of determining if model parameters can be uniquely calculated from dynamical data. This means that both the states and parameters have to converge towards the same values, independently of the initial values of the estimates if the same set of data is used. Such a practical analysis is no guarantee for global identifiability (since not every possible combination can be tested) but it gives a hint. The method does, however, show if a model is *not* globally identifiable. A theoretical identifiability analysis like the one performed in the small example of section 4.1 is more or less impossible to perform due to the complexity of the models. In order for a model to be *practically* identifiable other factors have to be considered too. For example, the convergence has to be acceptable also for noisy measurements. Furthermore, numerical problems may appear.

Fortunately, it appears that the choice of identification method is not such a crucial issue. Moreover, the purpose of this study is primarily to verify that there is any convergence at all and not to determine which method is the optimum one. Therefore, one off-line optimization algorithm using a simplex search method and an on-line approach using a simplified extended Kalman filter were selected for the identification of the reduced order models. Both methods have been used in numerous applications and are well known.

General model representation

It is often unrealistic to assume that all the states of a system and the disturbances can be measured. If a mathematical model of the system is available, the states can often be computed from measured inputs and outputs - *state estimation*. For the basic theory refer to a textbook in control such as [Åström and Wittenmark, 1990 and Kuo, 1991]. In this work, the two reduced order models (A and B) are supposed to describe the dynamics of the process and can be schematically described in the format:

$$\begin{cases} \dfrac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, t) + \mathbf{g}(\mathbf{u}, t) \\ \mathbf{y} = \mathbf{h}(\mathbf{x}, t) \end{cases} \tag{4.26}$$

where $\mathbf{x}(t)$ is a vector that represents all the concentrations indicated in (4.15 - 4.24), i.e. five for each completely mixed subreactor and $\mathbf{y}(t)$ represents the vector of measurable concentrations (normally X_{COD}, S_{NH}, S_{NO}). The term $\mathbf{g}(\mathbf{u}, t)$ indicates the input from the influent flow to each reactor, respectively. The same system can also be described by a time discrete representation:

$$\begin{cases} \mathbf{x}(t_{k+1}) = \mathbf{F}(\mathbf{x}(t_k), t_k) + \mathbf{G}(\mathbf{u}(t_k), t_k) \\ \mathbf{y}(t_k) = \mathbf{H}(\mathbf{x}(t_k), t_k) \end{cases} \tag{4.27}$$

where t_k and t_{k+1} are consequtive measurement times. Alternatively, it may be written on the form (4.28) when it is linearized for *every* time step. Note that the Φ, Γ, and $\mathbf{C}$ matrices may change as $\mathbf{x}$ and $\mathbf{u}$ vary.

$$\begin{cases} \mathbf{x}(t_{k+1}) = \Phi\,\mathbf{x}(t_k) + \Gamma\,\mathbf{u}(t_k) \\ \mathbf{y}(t_k) = \mathbf{C}\,\mathbf{x}(t_k) \end{cases} \tag{4.28}$$

The relationship between the measurable quantities $\mathbf{y}$ and the state vector $\mathbf{x}$ is in this case very straightforward since all elements in $\mathbf{y}$ are also directly available in $\mathbf{x}$ (except in the special case when the *OUR* in the aerobic reactor is assumed measurable). Therefore $\mathbf{C}$ is normally an identity matrix.

The idea of estimation

If the model (4.28) is fully observable, the complete state vector can be directly calculated from the inputs and outputs. One disadvantage of such a method is that it may be sensitive to disturbances. But more important, a good result depends on the model being sufficiently accurate. As discussed in section 2.4, several model parameters (the Φ and Γ matrices) are time variant and it is essential to keep track of their values as the

conditions change over time. Therefore a direct method is not sufficient. It is possible, however, to use the dynamical model to reconstruct the state variables as well as performing *parameter estimation,* simultaneously.

Basicly, the method of reconstruction is based on the assumption that the true state **x** can be approximated by the state **x** of the model:

$$\hat{\mathbf{x}}(t_{k+1}) = \Phi\, \hat{\mathbf{x}}(t_k) + \Gamma\, \mathbf{u}(t_k) \qquad (4.29)$$

which has the same input as system (4.28). If the model (4.29) is perfect in the sense that the parameters are identical to those of system (4.28) and if the initial conditions of (4.28) and (4.29) are the same, then the state **x** will be identical to the state **x** of the true system. If the initial conditions are different, then **x** will converge to **x** only if system (4.28) is asymptotically stable.

The reconstruction in (4.29) does, however, not make use of the measured output **y**. Therefore the method can be improved by introducing the difference between the measured and estimated output as a feedback to obtain:

$$\hat{\mathbf{x}}(t_{k+1} \mid t_k) = \Phi\, \hat{\mathbf{x}}(t_k \mid t_{k-1}) + \Gamma\, \mathbf{u}(t_k) + \mathbf{K}\left[\mathbf{y}(t_k) - \mathbf{C}\, \hat{\mathbf{x}}(t_k \mid t_{k-1})\right] \qquad (4.30)$$

The system in (4.30) is called an observer and exists in many variations depending on how the **K** matrix is chosen. The notation $\mathbf{x}(t_{k+1} \mid t_k)$ is used to indicate that it is an estimate of $\mathbf{x}(t_{k+1})$ based on measurements available at time t_k. In order to use this method for simultaneous state and parameter estimation is has to be slightly modified and **x** will become a generalized state vector which contains not only the state variables but also the unknown parameters to be estimated.

Estimation as an optimization problem

Parameter estimation problems can also be formulated as an optimization problem where the best model is the one that best fits the data according to a given criterion. Such a criterion (J) is usually based on the difference between the real measurements **y**

and the model outputs **y** in the form:

$$J = \sum_{k=1}^{n} \left[\mathbf{y}(t_k) - \hat{\mathbf{y}}(t_k)\right]^T \mathbf{W} \left[\mathbf{y}(t_k) - \hat{\mathbf{y}}(t_k)\right] \qquad (4.31)$$

where n is the number of discrete time measurements and **W** is a weight matrix. J is a function of all the unknown model parameters. They are adjusted until J has obtained a minimum. If there is only one unique minimum for J then the system is defined as globally (or parameter) identifiable. The loss function in (4.31) is an example of a weighted least squares criterion.

In this work the weight factors (w_{ii}) have been chosen in such a way that a 10% difference between the measured value and the model estimate around the steady state behaviour gives approximately the same contribution for all measured quantities. This means that the measurements are considered to be of the same quality and that the model is capable of estimating all variables with the same accuracy. If this was not the case, less weight could have been given to some of the residuals. It is also possible to let the weight factors vary with time and the current value of the measurements. Finally, some other commonly used criteria are extended least squares, generalized least squares, and maximum-likelihood (which involves computation of the gradient of J with respect to the parameters, as well as a matrix of second partial derivatives).

The simplex method applied for this work is actually an *ad hoc* method and applies a type of random search by calculating and examining the function value (i.e. the loss function) at several points - together forming a simplex - and moving towards lower values until convergence. The method has already been used in the small example in section 4.1 with good results. The main advantage is the robustness of the algorithm and its insensitivity to noise, whereas the convergence rate is slow and the computational effort goes up rapidly (typically as 2^n) with the dimension (n) of the model. A more detailed description of the simplex optimization algorithm is given in Appendix D. Some results of the method applied to the reduced order models are presented in section 5.2.

The Kalman filter

The Kalman filter is based on the reconstruction algorithm of (4.30) [Åström and Wittenmark, 1990]. By updating the gain matrix **K** in a special way, the estimation of the states is optimal in the sense that the variance of the reconstruction error is minimized. The problem is that the disturbances and character of the noise have to be very well known. For the extended Kalman filter not only the states **x** are updated according to the measurements but also the unknown parameters. The filter algorithm can be divided into two phases, *prediction* and *correction*. The principle structure of the on-line identification procedure is illustrated in Figure 4.6.

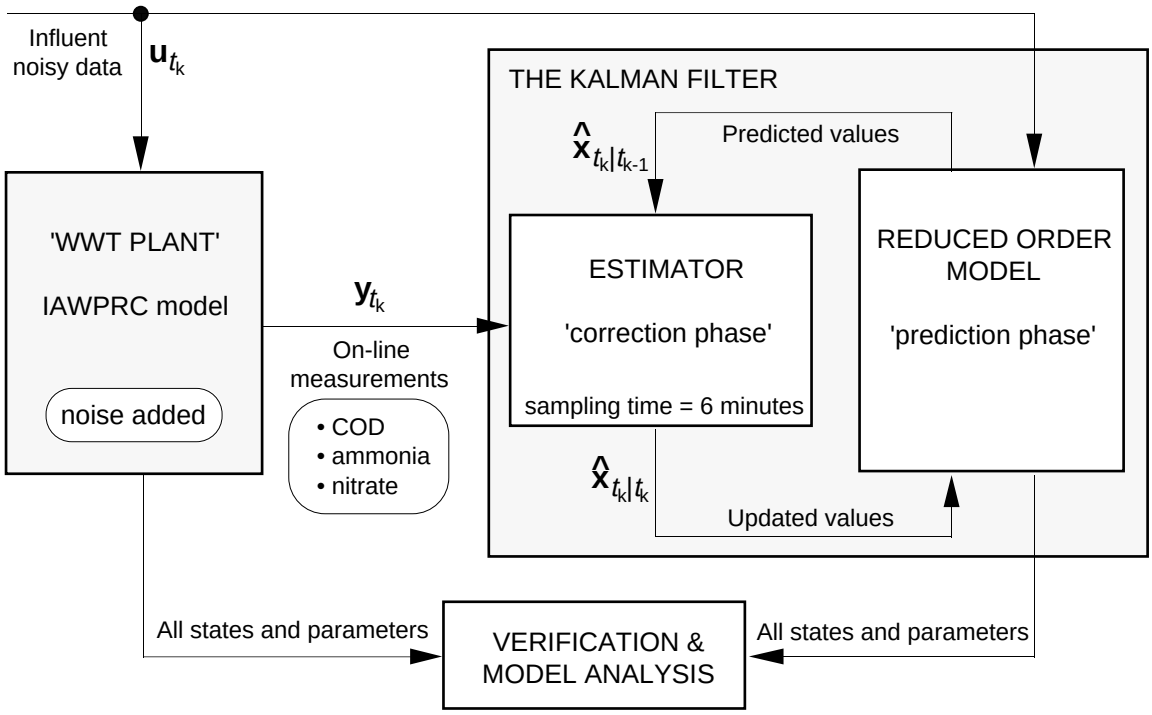


Figure 4.6 Structured identification using an extended Kalman filter.

In the prediction phase the dynamical equations of the model are integrated between two measurements, from time t_{k-1} to time t_k (using a smaller integration time) as shown below:

$$\hat{\mathbf{x}}(t_k \mid t_{k-1}) = \Phi\, \hat{\mathbf{x}}(t_{k-1} \mid t_{k-1}) + \Gamma \mathbf{u}(t_{k-1}) \qquad (4.32)$$

The states (i.e. the concentration variables) and the parameters are now based on measurements up until time t_{k-1}. As new measurements are acquired at time t_k, they are used to update the generalized state vector. The latter is called the correction phase and is based on the calculation:

$$\hat{\mathbf{x}}(t_k \mid t_k) = \hat{\mathbf{x}}(t_k \mid t_{k-1}) + \mathbf{K}\left[\mathbf{y}(t_k) - \mathbf{C}\,\hat{\mathbf{x}}(t_k \mid t_{k-1})\right] \qquad (4.33)$$

A Kalman filter is, however, based on the assumption that the dynamics are linear, which is not the case for this application. In order to calculate **K**, the dynamical equations are linearized around the existing operating point for each measurement instance. At the correction time, **K** is calculated from the linearized equations at time t_k and depends not only on the linearized states but also on the character of the noise that affects both states and measurements.

In this work the gain matrix **K** has been kept constant in order to simplify the computations. The value of **K** will influence the convergence speed of the parameters towards their final values. The chosen value of **K** has been calculated as the ultimate values which are obtained by the Kalman filter algorithm using the actual steady state values from the IAWPRC model as the operating point and the current noise conditions. In Appendix E, a more detailed description of the extended Kalman filter is given and some of the computational results are presented in section 5.3.

The different concentrations in the models have significantly different values, expressed in mg/l. In order to obtain reasonably accurate identification results, it is mandatory to *scale* the equations or *normalize* them to a reference point so that the concentration values are expressed in the same order of magnitude. This is even more important when the parameter values are considered. If scaling is not performed, the **K** matrix will contain so largely different elements that the estimation becomes numerically infeasible.

Sensitivity analysis

Sensitivity analysis is another important tool when analysing model characteristics. It expresses the influence of a small parameter change on the state variables and can therefore provide strong indications as to which parameters are most difficult to identify either because of their limited influence on the total system behaviour or due to the fact that several parameters compensate one another. A good example of the latter was the effect of μ and K_S in the example (section 4.1). A number of parameter sets gave approximately identical system behaviour because the effects of one parameter were compensated by another.

The method is also useful for experiment planning and model reduction. Since the aim is to identify the parameters of the reduced order models under normal operating conditions the analysis is performed under such conditions, i.e. varying influent flow rates and concentrations, both anoxic and aerobic zone active, changing internal recirculation and sludge recycle flow (conditions described in section 5.1). By simulating such a process using 'model A', changing one parameter slightly, rerun the simulation, etc. and storing the value of the loss function (4.31) for each simulation, a rough 'map' describing the influence of the parameters on the model behaviour is acquired. The relative change of the parameters are the same every time in order to get a fair comparison. The results of such an analysis are presented and discussed in section 5.2.

It is often more practical to do a sensitivity analysis on decoupled systems (only anoxic or aerobic part) under batch conditions and without feedback (recirculation). Results from this type of unit operation may show quite different results, which are not applicable for real plant operation since special modes which do not appear under normal operation are often emphasized. For example, the decay rate factor, b, in the small example was easily identified from a batch experiment because its effect was made so obvious when no substrate was available (as a sensitivity analysis will show). A situation like that would, however, not occur under normal operating condition and in a continuous reactor the identification would be much more difficult.

Therefore, since the purpose of the work is aimed at on-line parameter identification under normal operating conditions, it is important to perform the sensitivity analysis under the same type of conditions. On the other hand, if the purpose of the work is aimed at experiment design, the possibility of investigating how different conditions affect the identifiability properties may prove very useful.

Summary of verification methods

The number of principles and methods which may be applied for the identification and verification phase of a mathematical model is obviously large. Some of the factors which influence the results are listed below:

- type of identification method (on-line or off-line);
- noise conditions (amplitude, variance, character);
- process input perturbation;
- operational conditions (feedback, unit or coupled operation);
- physical outline and limitations of the process;
- measurable quantities (which ones, accuracy, sampling rate).

It is practically impossible to investigate all possible combinations of these factors. In Chapter 5 the results from a reasonable selection of cases are presented.

5

Computational Results

In this chapter some results based on computer simulations of the simplified models are presented using off-line as well as on-line state and parameter estimation methods. A short introduction describing the simulation conditions, physical configuration of the plant, influent conditions, and special assumptions applied is given in section 5.1. In the next section a comparison of the dynamic behaviour of the IAWPRC model and the reduced order models is performed for both unit and coupled operation based on an off-line optimization method under different assumptions of which measured quantities are available. The sensitivity of the reduced model to parameter changes under normal operating conditions by examining the loss function is also performed. Finally, an on-line estimation algorithm is tested under similar conditions. These results are discussed in section 5.3.

5.1 Plant Configuration

As discussed in section 4.3, all data used for the model identification and verification in this work have been based on simulations. There are several reasons for such an approach. Since the analysis is mainly theoretical, principal model weaknesses can be more thoroughly investigated using simulated data where the possibility exists to change the noise characteristics, repeat an experiment under identical conditions but for a change in one specific variable, control the input to the system, etc. Furthermore, the time and effort required to collect the data from a full scale plant can not be motivated at this early stage of the work. When the structural modelling problems have

been satisfactory solved, investigations based on real data are called for.

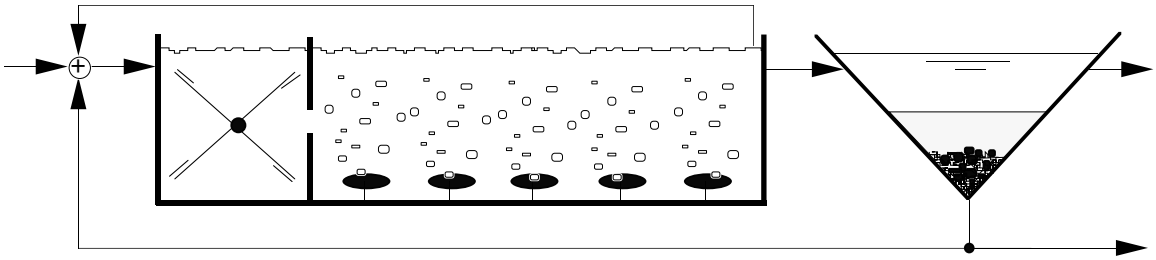
In order to test the reduced order models, the IAWPRC model has been used as the true 'plant'. The reason for choosing this model is mainly because of its widespread use and quite general acceptance. The physical outline of the plant includes a completely mixed anoxic reactor for pre-denitrification followed by a completely mixed aerobic reactor and a settler. The process involves an internal recirculation from the aerobic to the anoxic reactor as well as sludge recycling from the thickener to the anoxic reactor. All influent wastewater is fed into the anoxic reactor. The system is operated at a sludge age of ten days and a hydraulic retention time of ten hours. The default parameters for the IAWPRC model at 20 °C are used for the simulations. A more detailed description of the volumes, flow rates, influent wastewater characteristics, etc., is given in Figure 5.1.

The IAWPRC model does not, however, include any structured model of the sedimentation process. Since it is beyond the scope of this work to develope a new settler model, a fairly simple one was selected from the available literature. The thickener is modelled as a constant compaction ratio (γ) between the underflow sludge concentration and the average sludge concentration in the reactors with no internal biological activity [Olsson and Andrews, 1978]. Here a steady state relationship over the clarifier-thickener unit has been derived. From there the compaction ratio can be expressed in terms of flow rates and the sludge retention time as:

$$\gamma = \frac{Q_{in} + Q_r - \frac{V_{anox} + V_{air}}{\theta_x}}{Q_r} \qquad (5.1)$$

where: Q_{in} = influent flow rate to WWT plant;
Q_r = sludge recycle flow rate from settler;
θ_x = sludge retention time;
$V_{anox} + V_{air}$ = total reactor volume.

The hydraulic retention time of the settler is taken into account by a subsequent time lag. It should be noted that all the biodegradable organic matter of the reduced models (X_{COD}) is considered as part of the floc (which is why the IAWPRC nomenclature for particulate matter X, is used) and consequently settles. In the IAWPRC model, however, only the X_S fraction is treated in this way while the S_S fraction is considered soluble.



Operational variables			Model parameters [Henze et al.,1987a]		
Influent flow rate (Q_{in})	=	3000 m³/day	$\hat{\mu}_H$	=	6.0 day^{-1}
Recycle flow rate	=	0.5*Q_{in}	K_S	=	20 mg COD/l
Internal recycle flow rate	=	3*Q_{in}	$K_{O,H}$	=	0.2 mg O_2/l
Anoxic tank volume	=	250 m³	K_{NO}	=	0.5 mg NO_3-N/l
Aerobic tank volume	=	1000 m³	b_H	=	0.62 day^{-1}
Settler volume	=	1250 m³	$\hat{\mu}_A$	=	0.8 day^{-1}
Sludge age	=	10 days	K_{NH}	=	1.0 mg NH_3-N/l
Hydraulic retention time	=	10 hours	$K_{O,A}$	=	0.4 mg O_2/l
Influent wastewater characteristics			b_A	=	0.2 day^{-1}
S_S	=	30 mg COD/l	k_a	=	0.08 mg/(mg COD day)
X_S	=	70 mg COD/l	k_h	=	3.0 mg COD/(mg COD day)
X_I	=	10 mg COD/l	K_X	=	0.03 mg COD/(mg COD)
S_{NO}	=	2 mg N/l	η_g	=	0.8
S_{NH}	=	10 mg N/l	η_h	=	0.4
S_{ND}	=	1 mg N/l	Y_H	=	0.67 mg COD/(mg COD)
X_{ND}	=	0 mg N/l	Y_A	=	0.24 mg COD/(mg COD)
S_O	=	0 mg (-COD)/l	f_P	=	0.08
$X_{B,H}$	=	0 mg COD/l	i_{XB}	=	0.086 mg N/(mg COD)
$X_{B,A}$	=	0 mg COD/l	i_{XP}	=	0.06 mg N/(mg COD)
Setpoint for oxygen concentration in aerobic reactor = 2.0 mg O_2/l					

Figure 5.1 Configuration of the simulated WWT plant.

The impact of this difference on the process behaviour is not significant since the S_S concentration is normally very low by the time the wastewater reaches the settler in systems performing

nitrification, due to the long retention time. This difference has a larger impact when the total biodegradable substrate concentration (i.e. $S_S + X_S$ in the IAWPRC model) of the influent to the two models change. In the IAWPRC model only the S_S fraction is directly available for growth whereas the entire X_{COD} fraction in the reduced models is available for growth. The clarifier behaviour is hidden behind the definition of γ.

The importance of process excitation was discussed in section 4.3. For this work the system was perturbed using pulse disturbances of the following variables in the influent:

- flow rate;
- total biodegradable organic substrate concentration;
- ammonia concentration.

The reduced order models have a somewhat different behaviour than the IAWPRC model, especially for transients in an hourly time scale. The accuracy of the reduced models for such responses is not so good mainly due to the fact that the hydrolysis mechanisms are neglected. Thus, the pulse transients appear to be the most decisive model test. The type of disturbances applied is not optimal from an identification point of view but sudden changes of the input signal have the advantage that they excite several time modes of the system whereas sinusodial variations excite only one specific frequency.

At a real WWT plant it is usually quite easy to excite the influent flow rate in a steplike manner by varying the pumping capacity into the plant. It is more difficult to achieve a similar change for the biodegradable substrate concentration and the ammonia concentration. At certain times, abrupt sudden changes of the above concentrations may occur and can be used for identification purposes. Such natural variations of the influent variables should of course be used in order to achieve a good estimation result.

For the off-line optimization in section 5.2, the influent variables are usually perturbed according to Figure 5.2. The identification is based on measurements during a ten day period which begins

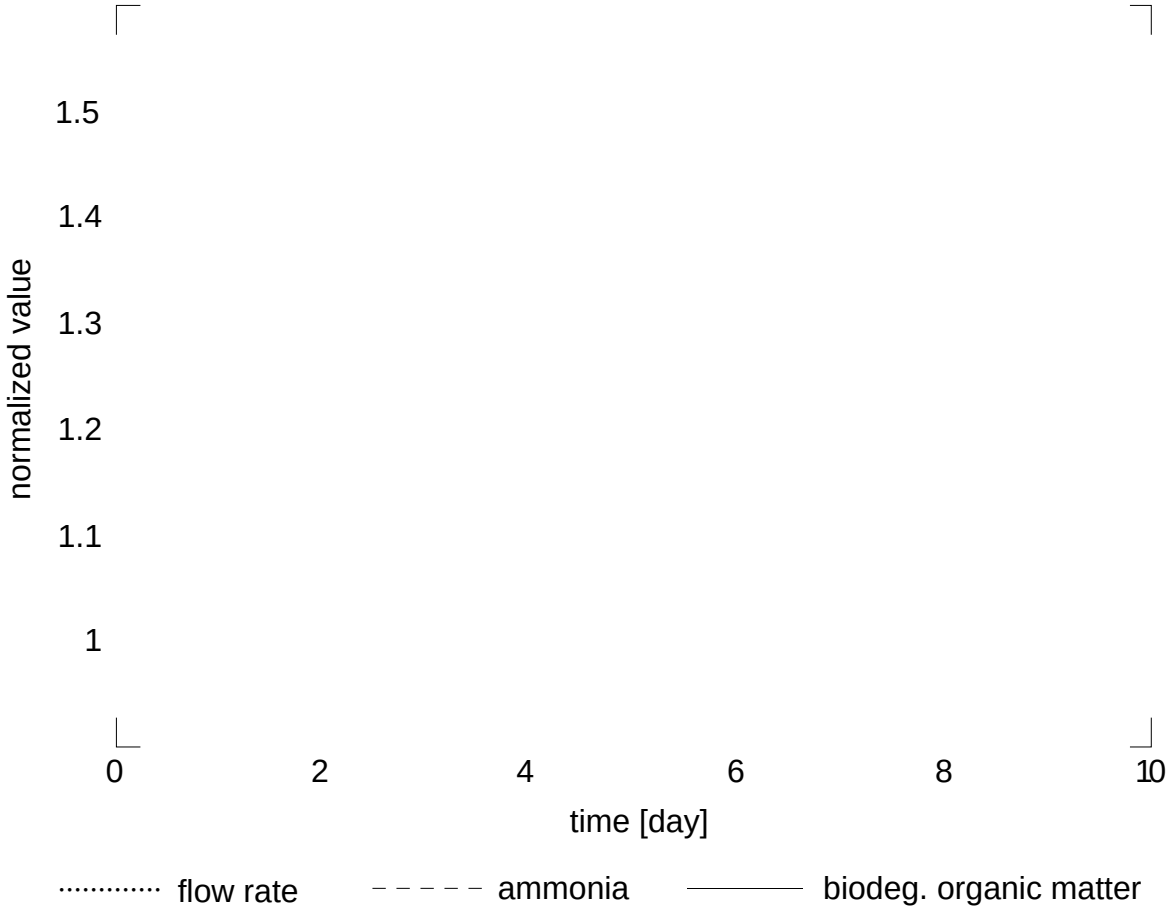


Figure 5.2 Pulse disturbances of the influent variables (normalized).

with a steady state period for three days followed by a 50% increase of the three quantities discussed above during a two day period (each disturbance lasting for one day) and finally a five day period to allow the transients to settle. Such a time series includes steady state behaviour, fast transients during the disturbance period, and slow transients as the system settles down towards steady state again. For the on-line identification in section 5.3, different types of disturbances have been applied.

The available measurements are assumed to be quite favourable. The basic case considers the following concentrations to be directly measurable on-line:

- biodegradable organic matter concentration;
- ammonia concentration;
- nitrate concentration.

The applied sampling time for the simulations have been six minutes. It is assumed that the variables are measured not only in the influent wastewater stream but also in both the anoxic and aerobic reactor. This is not unrealistic but requires a number of sophisticated and expensive sensors. Besides the concentrations above, the influent flow rate is assumed to be continuously available.

In the second case the oxygen uptake rate is also considered to be continuously available as a measurable quantity in the aerobic reactor. This is quite difficult to achieve at a true plant. Some methods aiming at this were discussed in section 2.5.

In the final case, in order to further investigate the identifiability properties of the reduced models, the concentration of organisms (both Heterotrophs and Autotrophs) are assumed to be measurable. Although this is an unrealistic assumption, the reason is to examine the behaviour of the models if all state variables are possible to measure directly (thereby locked). Then only the model parameters will determine a good model fit.

The applied noise conditions for the simulations vary significantly. For the off-line estimation, noise is usually not added because the main purpose is to investigate the basic identifiability properties. If the number of measuring points is sufficiently large then the same results will usually be reached whether noise is added or not (for the selected noise distributions of this work) as was illustrated in the example of section 4.1.

In the simulations where noise is added to the system (mainly for the on-line identification), this is done in two steps. *Input* noise is provided by adding Gaussian white noise to the variables in the input signal (influent flow, biodegradable organic matter, and ammonia concentration). The white noise has a mean value of zero and a standard deviation which is 10% of the actual value of the specific variable. This means that all variables are subjected to the same relative noise level. *Measurement* noise of the same character is added to the measurable quantities above in the same manner which implies that all measurable variables are considered to be equally uncertain. The selected noise level appears quite realistic although problems like 'outliers', trends, and uncalibrated sensors have not been considered.

All computer simulations in this work have been carried out using the simulation platforms Simnon™ [SSPA Systems, 1991] and Simulink™ [MathWorks, 1992]. A more detailed description of these two programs is provided in Appendix F.

5.2 Off-line Simulations

In order to perform a principle investigation of the behaviour of the reduced order models and the ability to identify the parameter sets from the type of data available from full scale WWT plant, a number of simulations are carried out. This kind of analysis does not provide certain proof whether a model is fully identifiable or not but it gives a strong indication of the major characteristics of the models and points out some of its potential weaknesses.

The assumed outline of the plant, the varying influent conditions, and the measurable variables were defined in the previous section together with the chosen values of the IAWPRC model parameters used to simulate the true WWT plant. The off-line optimization algorithm (a simplex method) and the type of loss function applied, have already been discussed in section 4.3 and the algorithm is described in more detail in Appendix D. In section 4.2 the reduced order models were developed and analysed from a biological/physical point of view and in this section they will be further investigated mainly from an identifiability/verifiability point of view. All the illustrated off-line estimations are based on simulated data without any noise added (compare with the example in section 4.1).

The results and conclusions are based on a large number of massive computations. Only a limited number of these are presented here. The selection of case studies to be discussed below are based on a combination of the following principle situations:

- different quantities assumed to be measurable;
- unit or coupled operation of the two reactor types;
- identification based on 'model A' or 'model B'.

An examination of the applied loss function and its sensitivity to parameter changes will also be performed for 'model A' under both unit and coupled operation. Such an analysis may explain and verify some results of the model optimization.

All results and plots are presented in the units milligram [mg], litre [l], and day [day]. During the actual computations the variables have been suitably scaled to satisfy numerical accuracy.

Case 1 – anoxic reactor, model A

The first case is an investigation of the *anoxic* reactor of 'model A'. Data is generated by simulating an IAWPRC model of a full scale plant according to the description in section 5.1. The variables of the total influent flow into the anoxic reactor as well as the internal variables of the reactor are stored to be used for the optimization. The anoxic reactor of 'model A' is then driven with the stored influent data.

The loss function on which the optimization is based, is calculated as the sum of weighted squares of the residuals. These residuals are the difference between the measurable quantities of the IAWPRC and the reduced model anoxic reactor. The simplex method finally suggests a new set of parameters for 'model A' and another iteration is started until an optimum solution is reached. The complete procedure is illustrated in Figure 5.3.

Two special cases are examined depending on which quantities are assumed measurable. In case 1A the X_{COD} (i.e. $S_S + X_S$ of the IAWPRC model), S_{NH}, and S_{NO} fractions of the anoxic reactor are assumed measurable and in case 1B the above variables plus the $X_{B,H}$ and $X_{B,A}$ fractions are assumed possible to measure. This difference affects the value of the loss function and thereby the achieved optimum. In both cases all five quantities are assumed available from the influent data (necessary if an optimization is

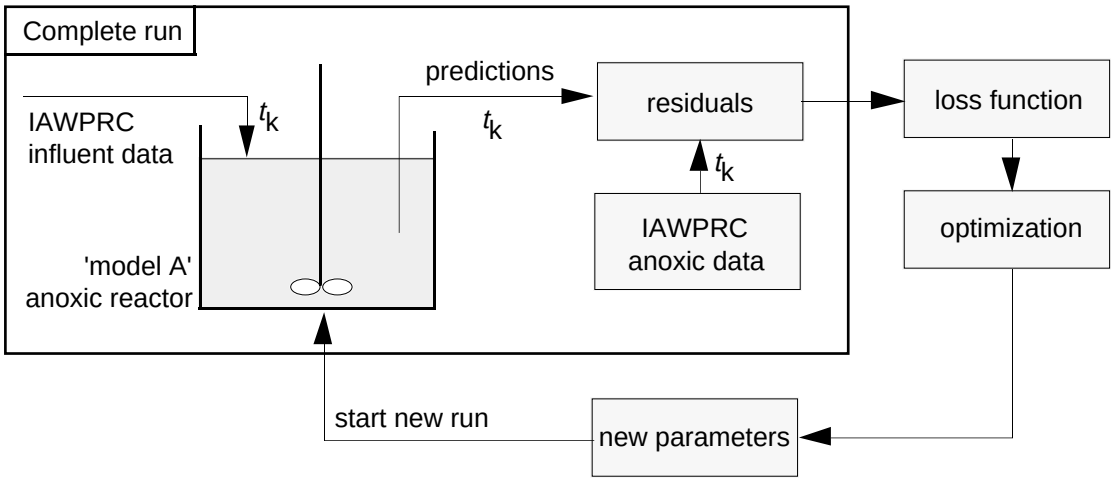


Figure 5.3 Optimization procedure of an anoxic reactor in unit operation.

to be performed on a single anoxic reactor, isolated from the rest of the plant).

It should be noted that the generated data are based on a simulation of a plant with both an anoxic and aerobic reactor, settler, sludge recirculation, etc., and not a special identification experiment with only an anoxic reactor. Such an approach would probably make the optimization easier but since the aim of the work is towards identification during normal operation and configuration, all data are generated from a complete plant. However, the situation is simplified because the anoxic reactor of 'model A' is simulated as a single unit operation although the data driving it is not. This means that the effects of the recirculation of 'model A' do not influence the results of these preliminary estimations.

It is important to determine whether the anoxic part of the reduced model is at all capable of mimicing the behaviour of the IAWPRC model (under the simplified assumption of unit operation). It is also a first rough test of the identifiability properties of 'model A' since several sets of initial values are used for the model parameters and different variables are assumed to be measurable for the different cases. The results of the optimization are presented in Table 5.1.

Optimization 'model A' anoxic part	initial estimates				final estimates				value of loss func.
	r_H	Y_H	b_H	b_A	r_H	Y_H	b_H	b_A	
Case 1A	.024	.35	.46	.06	.013	.498	.000	.014	19.3
(measured:	.046	.69	.94	.12	.013	.502	.000	.207	19.5
X_{COD}, S_{NH}, S_{NO})	.057	.86	1.18	.15	.025	.663	.326	.000	31.9
	.068	1.03	1.42	.18	.013	.498	.000	.000	19.3
Case 1B	.024	.35	.46	.06	.013	.498	.000	.014	19.5
(measured:	.046	.69	.94	.12	.013	.506	.000	.383	20.4
X_{COD}, S_{NH}, S_{NO},	.057	.86	1.18	.15	.013	.500	.000	.079	19.5
$X_{B,H}$, $X_{B,A}$)	.068	1.03	1.42	.18	.013	.502	.000	.159	19.6

Table 5.1 Results of the parameter optimization for case 1.

An analysis of the final estimates suggests some conclusions. First of all, the autotrophic decay rate is extremely difficult to estimate under the applied conditions. The concentration of Autotrophs does not change much due to reaction mechanisms in the anoxic reactor (see 4.19) but more due to the variations of the influent data (since an aerobic reactor was included to generate the data). Moreover the volume of the anoxic reactor is only 20% of the total reactor volume (Figure 5.1). Therefore b_A may assume practically any small value and its effect will be negligible. The situation is emphasized by the fact that the estimations are not improved in case 1B when $X_{B,A}$ is considered measurable. Its effect on X_{COD} and S_{NH} through the transformation of dead microorganisms is also small due to the low $X_{B,A}$ concentration. Part of the above holdsalso for b_H and its effect may be compensated for by the value of r_H under the current conditions.

Secondly, the parameters r_H and Y_H determine the main behaviour of the investigated system. In most situations these two parameters converge more or less globally (when the initial estimates are 'reasonable') for both cases 1A and 1B. The effect of including $X_{B,H}$ and $X_{B,A}$ in the loss function is small (the same optimum is reached for both cases) which could be expected since the effects of the recirculation are neglected. However, for the initial estimates of case 1A(row 3), a completely different optimum is reached which shows that the system is not globally

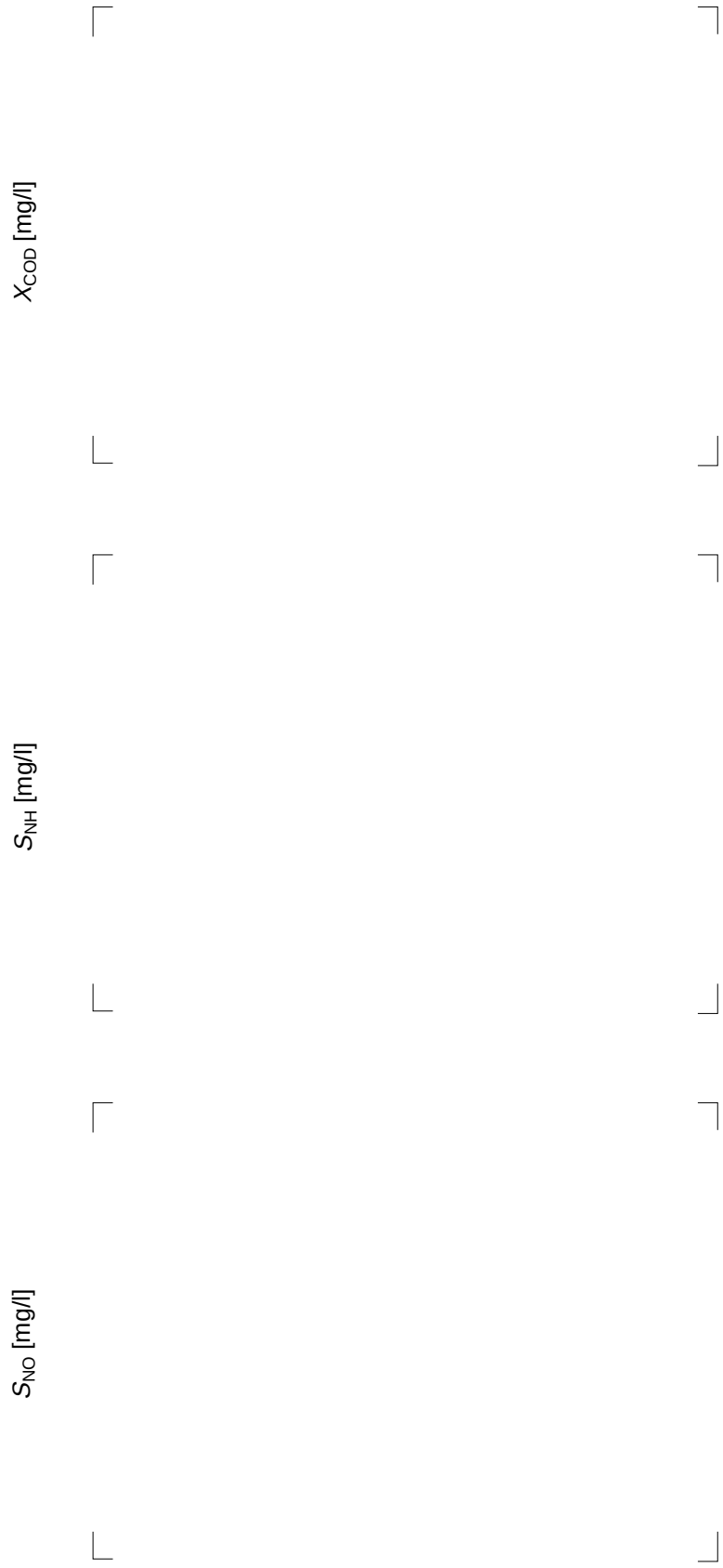
identifiable under those circumstances. When $X_{B,H}$ and $X_{B,A}$ are included in the loss function, their influence is significant enough to draw the optimization algorithm away from this local optimum and towards the truly best parameter set even when the values from the local optimum are used as initial seeds.

The reason why both optimums produce a similar output (Figure 5.4) is mainly due to the fact that the numerical values of the reaction rate expressions ($r_H \cdot X_{COD} - b_H$) and ($r_H \cdot X_{COD}/Y_H + b_H$) are practically the same for both parameter sets. By optimizing the system for several initial parameter sets and examine the value of the loss function it is possible to detect such difficulties.

Unfortunately it is not realistic to assume $X_{B,H}$ and $X_{B,A}$ to be measurable. The situation may be improved if instead measurements of the denitrification rate are included in the optimization. Such measurements can be to performed with reasonable accuracy but the possibility has not been tested in this work.

In order to verify the behaviour of the anoxic reactor in 'model A' after the optimization, the system is simulated with the obtained parameter sets and the values of the internal state variables are compared to those of the IAWPRC model when exposed to the same set of influent data. Such a comparison is illustrated in Figure 5.4.

It is obvious from the graphs that the anoxic part of the reduced model is capable of producing a result close to that of the IAWPRC model, both during transient and steady state operations. The main difference is observed for the biodegradable organic substrate since it is treated quite differently in the two models (as discussed in section 5.1). The discrepancy, however, is not very significant.


X_{COD} [mg/l]
S_{NH} [mg/l]
S_{NO} [mg/l]
time [day]

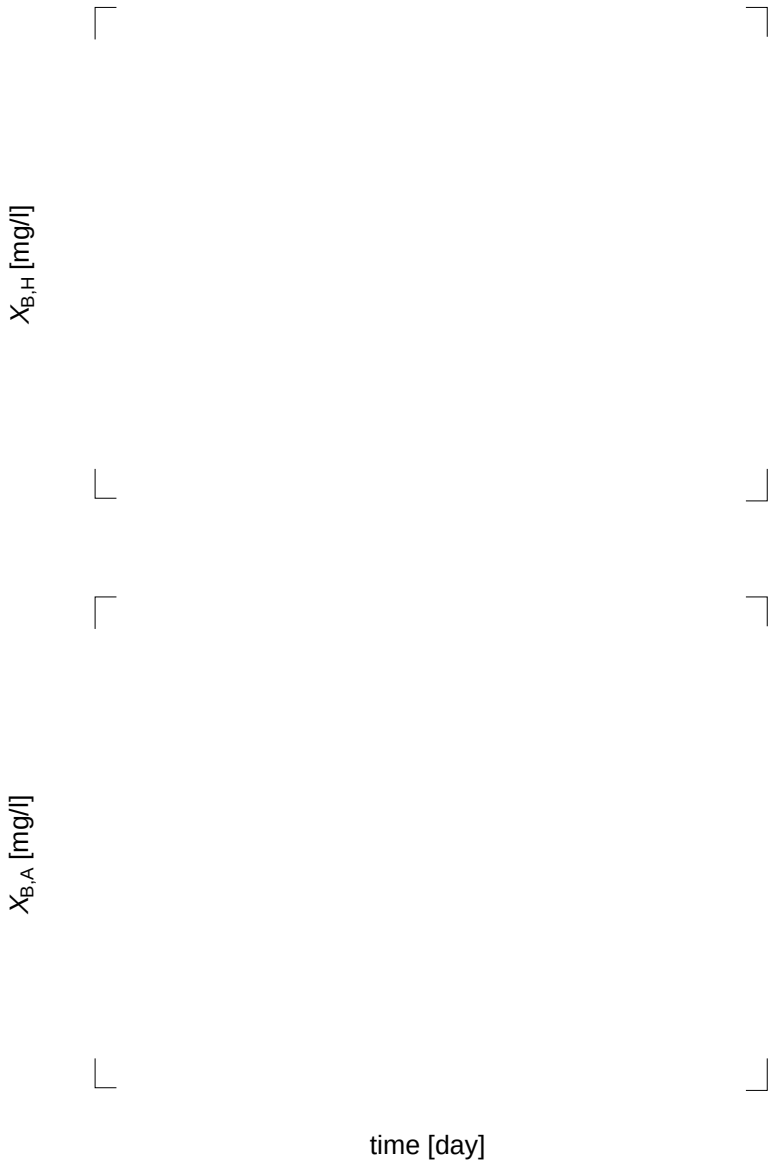


Figure 5.4 Behaviour of the anoxic reactor of 'model A' (dashed - case 1A(row 4), dotted - case 1A(row 2)) compared to the IAWPRC model (solid) simulated with identical influent characteristics.

Case 2 – aerobic reactor, model A

The second case is an investigation of the *aerobic* part of 'model A'. It is carried out in the same principal way as described for case 1. Data is generated by simulating an IAWPRC model of a full scale plant according to section 5.1. The variables of the total influent flow into the aerobic reactor as well as the internal variables of the reactor are stored to be used for the optimization. The aerobic reactor of 'model A' is then 'driven' with the stored

influent data. The loss function on which the optimization is based, is calculated as the sum of weighted squares of the residuals. These residuals are the difference between the measurable quantities of the IAWPRC and the reduced model aerobic reactor. The iteration is proceeded until an optimum parameter set is achieved.

Four special cases are examined depending on which quantities are assumed measurable from the aerobic reactor. These are:

- case 2A: measurements of X_{COD}, S_{NH}, S_{NO};
- case 2B: measurements of X_{COD}, S_{NH}, S_{NO}, OUR;
- case 2C: measurements of X_{COD}, S_{NH}, S_{NO}, $X_{B,H}$, $X_{B,A}$;
- case 2D: measurements of X_{COD}, S_{NH}, S_{NO}, $X_{B,H}$, $X_{B,A}$, OUR.

Depending on which quantities are considered measurable, the loss function assumes different values and various optima may be achieved. In the described cases, all five fractions are assumed available from the influent data.

This analysis shows whether the aerobic part of the reduced model is capable of mimicing the basic behaviour of the IAWPRC model or not. It is also a preliminary test of the identifiability properties of the aerobic part of 'model A' since several sets of initial values are investigated for the model parameters and different variables are assumed to be measurable for the various cases. Some results of the optimization are presented in Table 5.2.

Two principal situations may be observed - when the OUR is included in the optimization and when it is not. This leads to similar results for cases 2A and 2C on one hand and for cases 2B and 2D on the other. The effect of assuming the concentration of microorganisms ($X_{B,H}$ and $X_{B,A}$) to be measurable is small. As in case 1 it is extremely difficult to estimate the autotrophic decay rate factor b_A in a satisfactory way under the applied conditions. The impact of this parameter is negligible and easily compensated for by other parameters.

Optimization 'model A' aerobic part												value of loss func.
initial estimates						final estimates						
r_H	r_A	Y_H	Y_A	b_H	b_A	r_H	r_A	Y_H	Y_A	b_H	b_A	
Case 2A (measured: X_{COD}, S_{NH}, S_{NO})												
.046	.111	.33	.10	.42	.064	.029	.004	.494	.004	.172	.000	376
.093	.222	.66	.20	.86	.126	.035	.066	.558	.075	.227	.000	376
.116	.278	.83	.25	1.08	.157	.034	.000	.552	.000	.213	.072	369
.140	.333	.99	.30	1.30	.188	.032	.000	.528	.000	.219	.096	375
Case 2B (measured: X_{COD}, S_{NH}, S_{NO}, *OUR*)												
.046	.111	.33	.10	.42	.064	.044	.043	.582	.045	.358	.000	472
.093	.222	.66	.20	.86	.126	.044	.000	.588	.000	.364	.153	468
.116	.278	.83	.25	1.08	.157	.044	.001	.583	.000	.357	.018	466
.140	.333	.99	.30	1.30	.188	.043	.000	.581	.000	.342	.309	469
Case 2C (measured: X_{COD}, S_{NH}, S_{NO}, $X_{B,H}$, $X_{B,A}$)												
.046	.111	.33	.10	.42	.064	.033	.012	.554	.013	.207	.116	380
.093	.222	.66	.20	.86	.126	.034	.000	.554	.000	.208	.042	375
.116	.278	.83	.25	1.08	.157	.034	.000	.552	.000	.210	.080	377
.140	.333	.99	.30	1.30	.188	.033	.025	.544	.028	.209	.000	374
Case 2D (measured: X_{COD}, S_{NH}, S_{NO}, $X_{B,H}$, $X_{B,A}$, *OUR*)												
.046	.111	.33	.10	.42	.064	.044	.042	.583	.043	.362	.000	474
.093	.222	.66	.20	.86	.126	.044	.000	.584	.000	.359	.009	470
.116	.278	.83	.25	1.08	.157	.044	.000	.589	.000	.368	.000	472
.140	.333	.99	.30	1.30	.188	.043	.024	.579	.025	.352	.000	472

Table 5.2 Results of the parameter optimization for case 2.

It is interesting to observe that many parameters appear to converge more or less globally - r_H, Y_H, and b_H. For example, it was not possible to determine b_H realistically in case 1 (converged towards zero in most cases). The found optimum parameter set is quite different when the *OUR* is included in the calculations due to the added information.

Moreover, the parameters r_A and Y_A do not appear to converge globally. However, the ratio r_A/Y_A does always converge towards practically the same values (0.89 for cases 2A and 2C, 0.98 for cases 2B and 2D). This implies that the small difference in the $X_{B,A}$ concentration is not sufficient to separate the effects of the two parameters (see 4.24) even when the quantity is assumed measurable - especially as b_A varies as well. Only the combined effect of the parameters is possible to determine under the above conditions.

A change of the weight factors defined for the loss function may improve matters slightly. It should also be noted that for case 2A there exist a number of local minima close to each other depending on the strong correlation between r_H and b_H. It is quite natural since in this case, the optimization is based on the smallest amount of information (fewest number of measured quantities assumed).

The behaviour of the aerobic part of the reduced model is quite good when compared to the IAWPRC model for both steady state and transient situations (Figure 5.5). The differences are, however, larger than for the anoxic part of the model (amplified by the fact that the aerobic reactor volume is 80% of the total reactor volume). This is clear when the values of the loss functions for cases 1 and 2 are compared. For case 2 the values are about ten times larger, although the weight factors for both cases are chosen to show approximately the same impact for the same relative value of the squared residuals.

The behaviour of the aerobic reactor of the reduced order model is verified against the data from the IAWPRC model by simulating the system with the obtained parameter sets for identical influent conditions. In Figure 5.5 such a comparison is illustrated for two measurable quantities, namely S_{NH} and OUR. For the other variables, the results are quite similar to the ones already presented in Figure 5.4.

It is clear from the graphs in Figure 5.5 that the discrepancy for the S_{NH} fraction is larger for the aerobic reactor than it was for the anoxic one when compared to the IAWPRC model. This is among other things explained by the fact that the

ammonification and hydrolysis of organic nitrogen is neglected in the reduced order model.

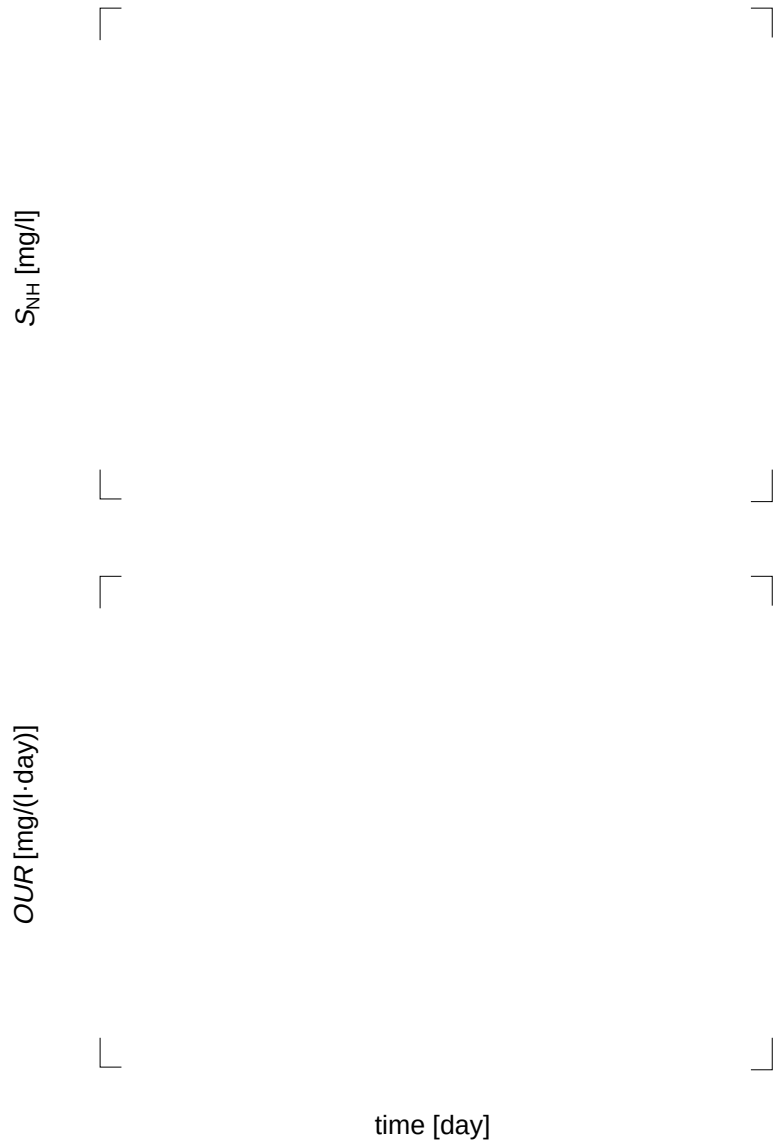


Figure 5.5 Behaviour of the aerobic reactor of 'model A' (dashed - case 2A(row 3), dotted - case 2D(row 3)) compared to the IAWPRC model (solid) simulated with identical influent characteristics.

The amount of readily biodegradable substrate (S_S) is also highly dependent on the hydrolysis of slowly biodegradable substrate (X_S) in the IAWPRC model because the influent S_S has already been consumed in the anoxic reactor. Variations of the S_S/X_S ratio is therefore an important factor in the IAWPRC model whereas

the reduced model does not respond to such changes as long as the S_S+X_S concentration is constant. For the applied conditions this effect is more prominent in the aerobic reactor than in the anoxic part of case 1. It emphasizes the impact of hydrolysis of organic matter in the IAWPRC model. Abrupt changes in the ratio of organic matter fractions are therefore troublesome to mimic (especially in an off-line estimation approach) and problems occur because the reduced order model does not include a hydrolysis mechanism. Due to the above reasons, a discrepancy also exists for the X_{COD} fraction and thereby an effect on the other measurable quantities is observable.

As for the *OUR* it shows that the reduced model produces a reasonable result even when the optimization is not based on information of this quantity. It is naturally improved when included in the loss function, especially in the steady state region. The discussed differences in the X_{COD} and S_{NH} fractions though, lead to a result which is not in complete agreement.

If the optimization of cases 1 and 2 are performed on stationary data or data with a 'small' perturbation, the parameters will not converge towards the values found in this work. The result in such a case will to a large degree depend on the initial values of the parameters. The information from that type of data is not sufficient to draw the algorithm towards the true optimum set.

Loss function analysis of cases 1 and 2

In cases 1 and 2, the anoxic and aerobic reactor of 'model A' were examined for their identifiability properties. Investigations were performed using different sets of initial parameter estimates as well as assuming various variables to be measurable. The reactor models were identified in unit operation and the effect of recirculation was therefore not included (although the data driving the single reactors was taken from a complete WWT plant simulated with the IAWPRC model). It was clear that the reduced model was capable of mimicing the behaviour of the IAWPRC model under these simple conditions although it proved troublesome to identify all the parameters - especially b_H and b_A but also to some extent r_A and Y_A - in a global sense.

In order to investigate the sensitivity of the model when exposed to parameter variations, the value of the applied loss function is analysed. This is not to be considered as a complete sensitivity analysis of the model but since the result of the optimization algorithm is based on the how the value of the loss function changes, it may serve as an indication of convergence problems.

The actual case chosen to be examined is when the earlier described weighted loss function is based on measurements of only X_{COD}, S_{NH} and S_{NO} (i.e. case 1A and 2A). By simulating each reactor repeatedly with identical influent data, introducing a small change of the value of a parameter for each run and storing the value of the loss function, an 'image' of how this value varies is achieved. The initial sets of parameters are the optimums found in case 1A(row 2) and case 2A(row 2) and each parameter varies ±20% around its initial value in steps of 2% for each run. This means that the relative change of every parameter is the same.

The situation is first illustrated in Figure 5.6 for the anoxic part of 'model A'. The model is simulated under the same basic conditions as described in case 1. Contour plots show how the value of the loss function is affected as the parameters change.

In the first plot in Figure 5.6 the parameters b_H and b_A are held constant at the optimum values and in the second plot r_H and Y_H are held constant. Therefore not all parameter interactions are shown (only two parameters change at a time). To show the complete effect of the parameter variations would require a four dimensional plot which would be awkward to interpret.

A small complication occurs when the sensitivity of b_H and b_A is to be examined. Since the optimum values found in case 1A for these parameters are so close to zero, the effect of a 20% change would be negligible. Because of this fact, the values for b_H and b_A are set to vary from 0.0 to 0.4 day^{-1} (quite realistic values) for the second plot in Figure 5.6 while the other two parameters remain at the optimum value. To be able to compare the plots, two contour lines next to each other indicate that the value of the loss function has changed 10 units (also in Figure 5.7).

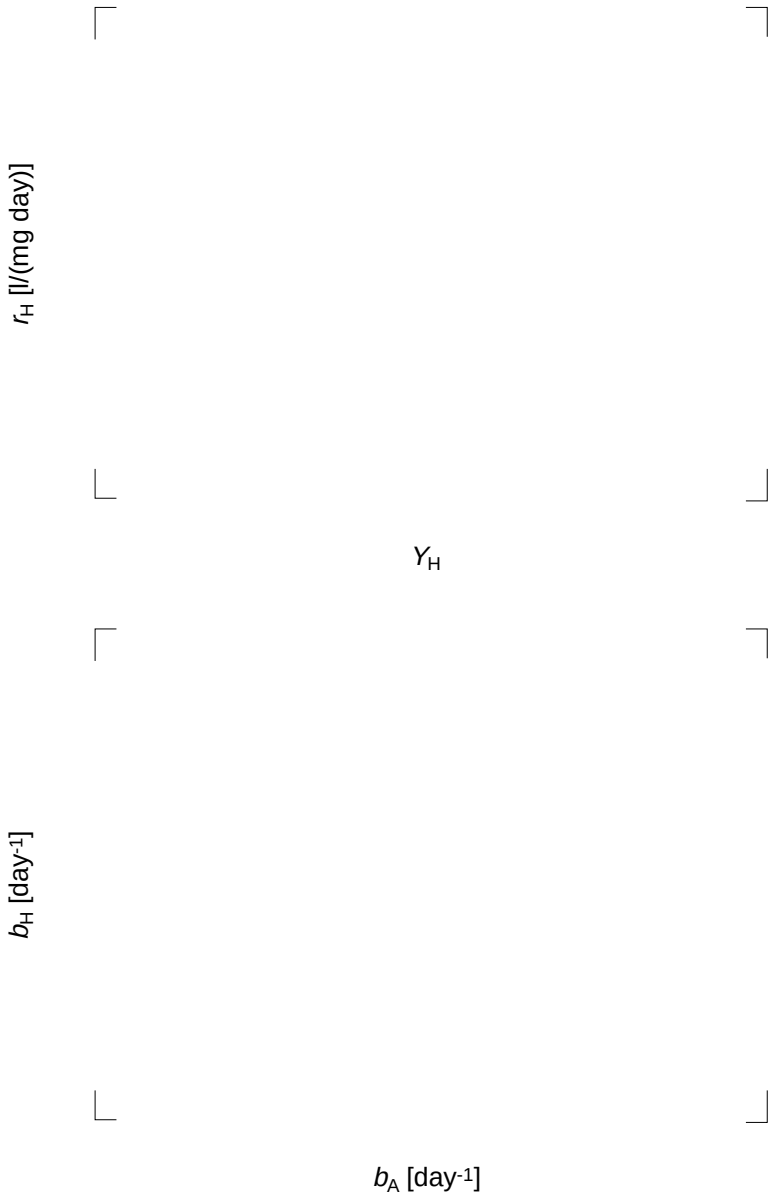


Figure 5.6 Contour plots of the loss function for the anoxic reactor, case 1.

The first plot above shows that the value of the loss function is sensitive to small changes in both r_H and Y_H and the direction of the gradient is obvious. This implies a good convergence although a problem occurs as the optimum is approached. Two different optima appear, very close to each other. The values of the parameters are almost the same for both optima but it is an indication that the information on which the loss function is based, is not sufficient or that there is a structural problem in the model. If the analysis was extended, another optimum would

show for the parameter set found in case 1A(row 3), though not as distinctly. It is also obvious that the ratio between the parameters are of importance for the model behaviour. Within the 'valley' - indicating a low value of the loss function - this ratio is quite constant.

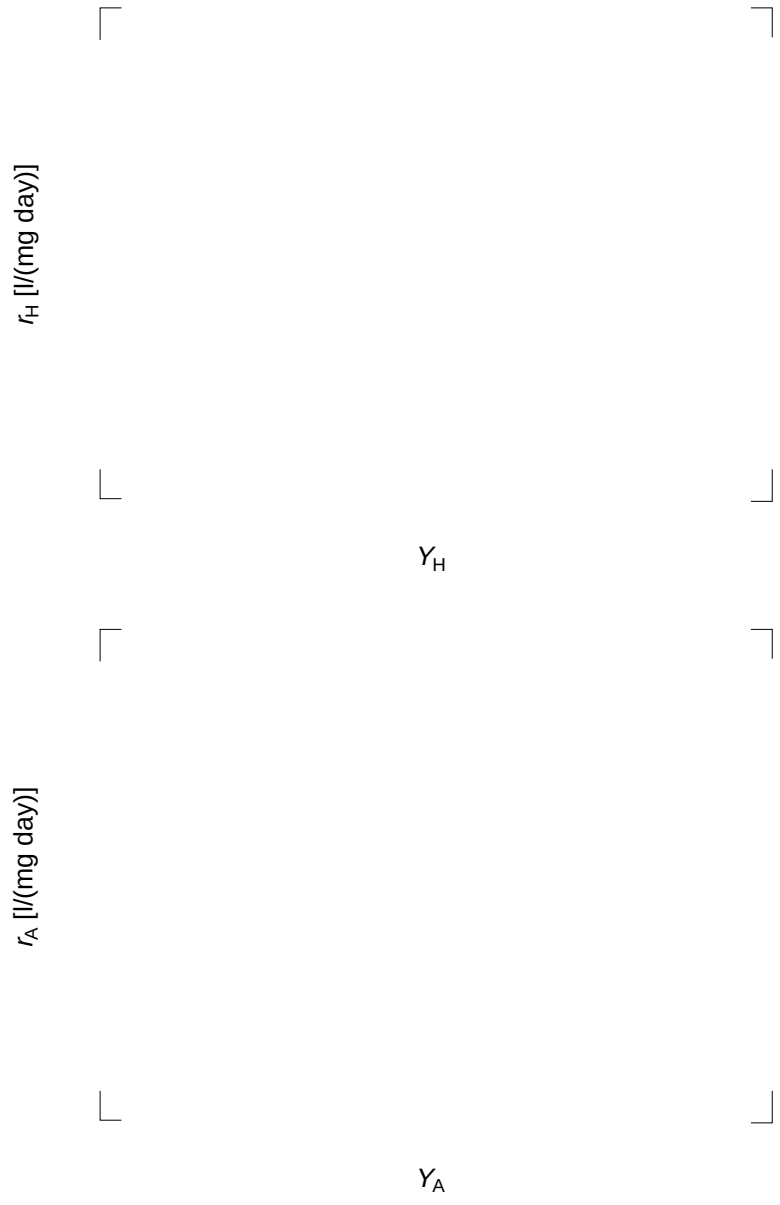
The second plot illustrates the previously discussed difficulty of identifying b_A. The gradient of the loss function clearly indicates how the optimization algorithm would change the value of b_H to improve the result though a significant change in b_A does practically not affect the loss function at all. The model is very insensitive to this parameter under the applied conditions which explains some of the results from case 1. In Table 5.1 it is also shown that the sensitivity is not notably improved when $X_{B,H}$ and $X_{B,A}$ are included in the calculations of the loss function (case 1B).

In Figure 5.7, results from the same type of analysis as described above are shown for the aerobic part of 'model A'. The initial parameter values are the optimum found in case 2A(row 2) and the simulations are carried out under the same conditions as described there. In order to detect any influence whatsoever when changing the b_A parameter, its value is set to vary between 0.0 and 0.4 day^{-1} in the last plot (for reasons discussed above).

The first plot in Figure 5.7 illustrates the strong sensitivity of the model to the two parameters r_H and Y_H. A clear optimum is achieved although from the results in Table 5.2 it is clear that the best parameter set is also dependant on the convergence of the other model parameters.

The second plot shows that the loss function is also sensitive to the parameters r_A and Y_A. However, their individual values are obviously of practically no significance under the applied conditions; it is only the ratio of the two that matters. The 'valley' is extremely long and narrow which means that the optimization algorithm has no problems determining the best ratio but no possibility to find the right individual values of the parameters. This explains some of the results discussed in case 2.

The final plot indicates the much smaller influence of b_H and b_A on the model behaviour. A variation of the value of b_A has a very limited effect. It is also notable that the optimum for b_A in this case appears to be a negative value. However, a deliberate restriction in the optimization algorithm hinders any of the model parameters to assume negative values even if it would further lower the value of the loss function.



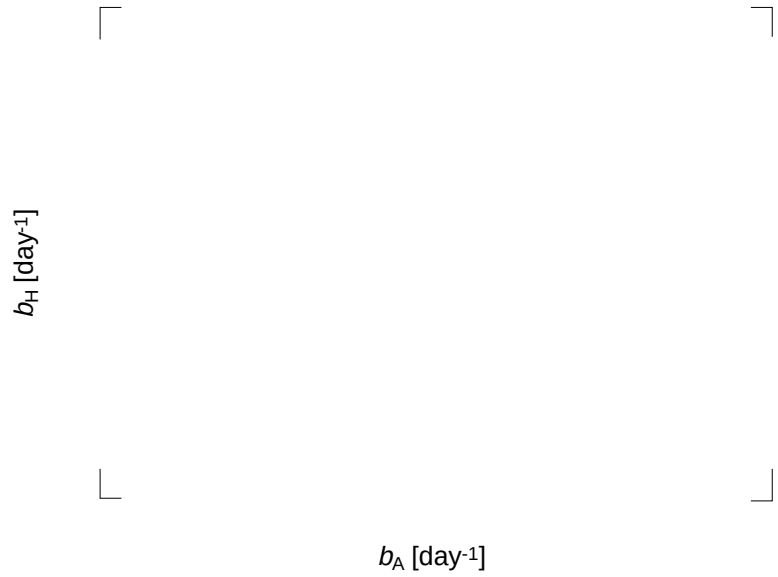


Figure 5.7 Contour plots of the loss function for the aerobic reactor, case 2.

Case 3 – anoxic/aerobic reactor combination, model A

Somewhat strengthened by the results already discussed, the optimization is now generalized. A natural extension would be to investigate the anoxic and aerobic part of 'model A' separately but under conditions where the effect of the recirculation is included in a direct way. However, the number of different cases to examine have to be limited for practical reasons and therefore the algorithm is applied to the complete model (coupled operation of the two reactor types) and all parameters are simultaneously estimated. The principle of the optimization is shown in Figure 5.8.

The required data is generated by simulating the IAWPRC model (the 'true' WWT plant) as earlier discussed. The influent data to the plant is identical to what was used in the earlier cases (see Figure 5.2). The reactor volumes, recirculation rates, settler model, etc., are also the same for both the IAWPRC and the reduced model (see Figure 5.1). Measurements are assumed to be available from the influent wastewater, the anoxic reactor, and the aerobic reactor.

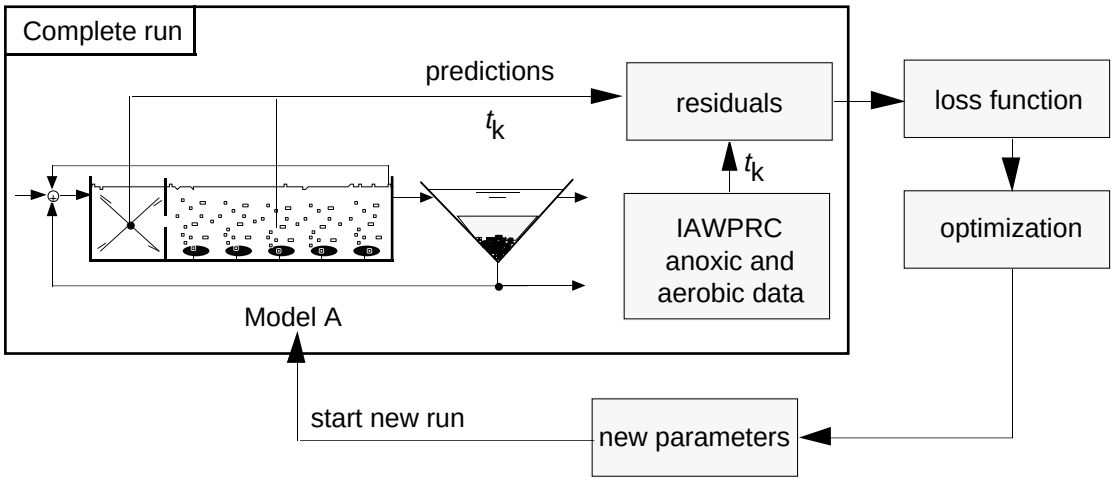


Figure 5.8 Optimization procedure of the complete 'model A'.

From the influent wastewater four quantities are always considered measurable - X_{COD}, S_{NH}, S_{NO}, and Q (the concentration of microorganisms in the influent is assumed negligible). Depending on which quantities are considered measurable in the two reactors, five interesting cases are examined:

- case 3A: measurements of X_{COD}, S_{NH}, S_{NO} in both the anoxic and aerobic reactor;
- case 3B: measurements of X_{COD}, S_{NH}, S_{NO} in both reactors plus OUR in the aerobic reactor;
- case 3C: measurements of X_{COD}, S_{NH}, S_{NO} in both reactors plus OUR, $X_{B,H}$, $X_{B,A}$ in the aerobic reactor;
- case 3D: measurements of X_{COD}, S_{NH}, S_{NO} in both reactors plus $X_{B,H}$, $X_{B,A}$ in the anoxic reactor;
- case 3E: measurements of X_{COD}, S_{NH}, S_{NO}, $X_{B,H}$, $X_{B,A}$ in both reactors plus OUR in the aerobic reactor.

The results of the optimization are presented in Tables 5.3 and 5.4 for different sets of initial estimates. The two tables show the anoxic and aerobic part respectively but the results should be interpreted simultaneously. They are separated only to make the results easier to read. This means that the value of the loss function in Table 5.4 represents the combined loss function for both the anoxic and aerobic part of the model.

Optimization	initial estimates				final estimates			
'model A' anoxic part	r_H	Y_H	b_H	b_A	r_H	Y_H	b_H	b_A
Case 3A	.024	.35	.46	.06	.023	.723	.372	.111
(measured:	.046	.69	.94	.12	.014	.726	.131	.071
X_{COD}, S_{NH}, S_{NO})	.057	.86	1.18	.15	.008	.711	.015	.073
	.068	1.03	1.42	.18	.030	.626	.579	.040
Case 3B	.024	.35	.46	.06	.050	.872	1.10	.090
(measured:	.046	.69	.94	.12	.013	.660	.117	.057
X_{COD}, S_{NH}, S_{NO})	.057	.86	1.18	.15	.021	.743	.302	.086
	.068	1.03	1.42	.18	.047	.862	1.02	.158
Case 3C	.024	.35	.46	.06	.043	.744	.830	.120
(measured:	.046	.69	.94	.12	.032	.691	.521	.139
X_{COD}, S_{NH}, S_{NO})	.057	.86	1.18	.15	.019	.573	.172	.110
	.068	1.03	1.42	.18	.029	.671	.452	.101
Case 3D	.024	.35	.46	.06	.031	.675	.510	.116
(measured:	.046	.69	.94	.12	.037	.711	.644	.128
X_{COD}, S_{NH}, S_{NO},	.057	.86	1.18	.15	.020	.566	.185	.070
$X_{B,H}$, $X_{B,A}$)	.068	1.03	1.42	.18	.041	.733	.782	.136
Case 3E	.024	.35	.46	.06	.027	.647	.360	.209
(measured:	.046	.69	.94	.12	.028	.652	.340	.075
X_{COD}, S_{NH}, S_{NO},	.057	.86	1.18	.15	.025	.630	.335	.163
$X_{B,H}$, $X_{B,A}$)	.068	1.03	1.42	.18	.027	.646	.380	.286

Table 5.3 Results of the parameter optimization for case 3 (anoxic part).

It is clear that the value of the loss function is much larger for case 3 than would be expected if the results from cases 1 and 2 could simply be added together for coupled operation. This is because the influent characteristics to the unit reactors in cases 1 and 2 were identical to the applied influent of the IAWPRC model. In case 3 only the influent wastewater into the plant is the same. An error introduced in the anoxic reactor is propagated into the aerobic reactor, amplified, and propagated into the anoxic zone again through the recirculation. This emphasizes the differences between the reduced and the IAWPRC model.

Optimization 'model A' aerobic part												value of loss func.
initial estimates						final estimates						
r_H	r_A	Y_H	Y_A	b_H	b_A	r_H	r_A	Y_H	Y_A	b_H	b_A	
Case 3A (measured: X_{COD}, S_{NH}, S_{NO})												
.046	.111	.33	.10	.42	.064	.078	.156	.766	.202	.780	.042	1430
.093	.222	.66	.20	.86	.126	.041	.180	.728	.059	.385	.097	1238
.116	.278	.83	.25	1.08	.157	.019	.250	.643	.716	.123	.174	979
.140	.333	.99	.30	1.30	.188	.042	.162	.478	.052	.385	.069	1499
Case 3B (measured: X_{COD}, S_{NH}, S_{NO}, *OUR*)												
.046	.111	.33	.10	.42	.064	.025	.142	.555	.129	.203	.030	1361
.093	.222	.66	.20	.86	.126	.028	.192	.602	.141	.234	.107	1112
.116	.278	.83	.25	1.08	.157	.029	.196	.576	.707	.248	.105	1200
.140	.333	.99	.30	1.30	.188	.028	.214	.568	.500	.239	.108	1370
Case 3C (measured: X_{COD}, S_{NH}, S_{NO}, $X_{B,H}$, $X_{B,A}$, *OUR*)												
.046	.111	.33	.10	.42	.064	.091	.232	.714	.208	.944	.138	2359
.093	.222	.66	.20	.86	.126	.041	.223	.539	.204	.406	.124	2146
.116	.278	.83	.25	1.08	.157	.050	.238	.586	.219	.503	.152	2121
.140	.333	.99	.30	1.30	.188	.044	.244	.556	.222	.439	.160	2150
Case 3D (measured: X_{COD}, S_{NH}, S_{NO})												
.046	.111	.33	.10	.42	.064	.042	.141	.528	.128	.409	.023	1369
.093	.222	.66	.20	.86	.126	.038	.208	.501	.188	.374	.107	1413
.116	.278	.83	.25	1.08	.157	.043	.221	.530	.199	.421	.138	1376
.140	.333	.99	.30	1.30	.188	.033	.201	.466	.180	.315	.095	1429
Case 3E (measured: X_{COD}, S_{NH}, S_{NO}, $X_{B,H}$, $X_{B,A}$, *OUR*)												
.046	.111	.33	.10	.42	.064	.044	.148	.548	.136	.443	.021	2149
.093	.222	.66	.20	.86	.126	.041	.189	.526	.173	.401	.097	2169
.116	.278	.83	.25	1.08	.157	.044	.189	.549	.172	.442	.075	2156
.140	.333	.99	.30	1.30	.188	.042	.166	.536	.152	.418	.015	2153

Table 5.4 Results of the parameter optimization for case 3 (aerobic part).

The different way the organic matter is treated by the settler model (as discussed in section 5.1) also has an influence on the behaviour. On an average more biodegradable organic matter is recirculated in the reduced model ($\gamma \cdot X_{COD}$) than in the IAWPRC model ($S_S + \gamma X_S$) and fed into the anoxic reactor. In cases 1 and 2 this discrepancy had no impact.

In order to avoid the influence of the initial transients when the parameters are adjusted during the optimization, the reduced model is simulated towards steady state for each new parameter set before the actual optimization is started.

The results from case 3A show large variations. Ten parameters are optimized based on six measurable quantities and several local optima are detected. Not even the most sensitive parameters r_H and Y_H converge towards the same values. However, a strong correlation between r_H and b_H is apparent in both reactors. A high value of r_H is always coupled to a high value of b_H. This is because the algorithm attempts to minimize the residuals of X_{COD}. A high r_H indicates a high growth rate and a large consumption of organic matter. This is compensated by a high b_H which leads to a high conversion rate of decayed material into X_{COD}. The mathematical relationship between the two parameters is not clear due to the different values of Y_H (among other things). However, the model sensitivity to the b_H parameter is increased due to the recirculation when compared to cases 1 and 2.

The same basic correlation seems to exist for r_A and b_A in the aerobic reactor due to the minimization of the residuals of S_{NH}. This conclusion is, however, more uncertain since a large part of the S_{NH} formed from decayed material originates from the Heterotrophs (i.e. b_H). The heterotrophic yield coefficient in the anoxic reactor appears quite stable whereas the other yield coefficients and the anoxic b_A show very large variations.

Another important factor which also has to be considered does not show in the tables. Since the concentrations of microorganisms are not assumed measurable, the different

parameter sets lead to very different values of $X_{B,H}$ and $X_{B,A}$. Based on the available measurements in case 3A, the optimization algorithm can not determine whether the true system involves a high concentration of organisms with a low reaction rate or vice versa. This fact affects all parameter values and is the major reason why the estimated sets are so different.

In an attempt to restrict the behaviour of the simplified model and enhance the optimization, *OUR* (4.25) is included as a measurable quantity in case 3B. The effect is apparent on the parameters r_H, Y_H, and b_H in the aerobic zone which now converge towards approximately the same values independent of the initial estimates. Therefore the predicted value of $X_{B,H}$ is much more stable (although not the same as the concentration suggested by the IAWPRC model). Since the concentration of Autotrophs is much smaller than $X_{B,H}$ its effect on the *OUR* is quite small. Consequently, $X_{B,A}$ converges towards different values and the estimates of the autotrophic parameters r_A, Y_A, and b_A are not significantly improved by the extra information. The same behaviour could be observed for case 2B. In the anoxic reactor, the earlier discussed correlation between r_H and b_H still holds although the actual results are not improved when compared to case 3A.

In case 3C the concentrations of Heterotrophs and Autotrophs in the aerobic reactor are also assumed measurable. This will force $X_{B,H}$ and $X_{B,A}$ in the reduced model towards the values of the IAWPRC model and thereby eliminating one of the problems discussed for cases A and B. Note that the concentration of microorganisms will be approximately the same in both reactors due to the recirculation although it is only assumed measurable in the aerobic reactor. The result of row 1 is clearly a special case where the algorithm has converged towards a local optimum quite far from the best one. The other three examples reach almost identical values of the loss function *but* with quite different parameter sets although the autotrophic parameters converge remarkably consistent. In case 3B the situation was almost the opposite. There are a number of reasons for this.

Apart from the earlier discussed correlation between r_H and b_H in the two reactors a new correlation occurs between the reactors. A low r_H in the aerobic zone leads to a high r_H in the anoxic zone and vice versa (the same holds for the ratio r_H/Y_H). In case 3A the opposite relationship could be observed. This implies that the model deals with the dynamic disturbances differently in the two reactors which can be observed in simulations but not in the value of the loss function where all errors are lumped together. The expression $(1-Y_H)\cdot r_H/Y_H$ in the aerobic reactor is also practically the same for all examples of case 3C due to the *OUR* measurements and the fact that $X_{B,H}$ is almost identical.

A further problem is that the measurements of the organism concentrations are in conflict with the *OUR* measurements. In order to reach the best estimation of the *OUR* behaviour, the reduced model requires a set of parameters which leads to a significantly higher $X_{B,H}$ whereas the measurements of this variable force the optimization algorithm away from this set to achieve the low concentration predicted by the IAWPRC model. Therefore the values of the loss function are much higher in cases C and E when compared to cases B and D.

In case 3D the concentrations of microorganisms in the anoxic reactor are considered measurable. Only the three basic quantities are assumed measurable in the aerobic zone. The results are very similar to case 3C. The values of the loss function are almost identical for all examples although much lower than in C. The parameter values of cases C and D are also within the same region which is an indication that the weight factor for the *OUR* residuals should maybe be increased in order to make the effect of these measurements more prominent. However, it is clear that the information of $X_{B,A}$ is essential for identifying many of the autotrophic parameters.

In the final case all internal state variables of 'model A' are assumed measurable in both reactors. Furthermore, the *OUR* is available for more information. In practise it is unrealistic to have so much data available but the case may be used for model analysis. Parameters which do not converge globally in this case will probably never be possible to estimate from this type of experiments based on full scale plant operations. The conclusion

would be that the influence of certain parameters is either negligible or the model suffers from basic structural problems.

The results from case 3E clearly indicate that most model parameters converge practically globally when this much information is available. The slightly different parameter values are mainly due to the large number of iterations required and in some cases the optimization has been stopped somewhat early. However, the autotrophic decay rate coefficient does not converge globally under the applied conditions in neither the anoxic nor the aerobic reactor (which to a small extent also influence the values of r_A and Y_A though the ratio r_A/Y_A is perfectly constant). A low b_A value in the anoxic reactor is compensated by a higher value in the aerobic reactor and vice versa. The recirculation then equalizes the differences in the $X_{B,A}$ concentration between the reactors.

The low concentration of Autotrophs (2-10%) when compared to Heterotrophs, implies that b_A, from the model point of view, is basically only important to keep the $X_{B,A}$ concentration at the proper level (if it is assumed measurable). It has no real significance on the process of transforming decayed material into X_{COD} and S_{NH}. This is a basic problem and motivates the modification of 'model A' into 'model B' which is further discussed in case 4. A small correlation is still apparent between the values of r_H and b_H.

An investigation of the behaviour of the loss function for the complete 'model A' during coupled operation has also been performed. By varying the model parameters (the same principle as for cases 1 and 2) the model sensitivity to parameter changes is examined. Although not presented here, the results of this analysis further motivate the conclusions discussed above.

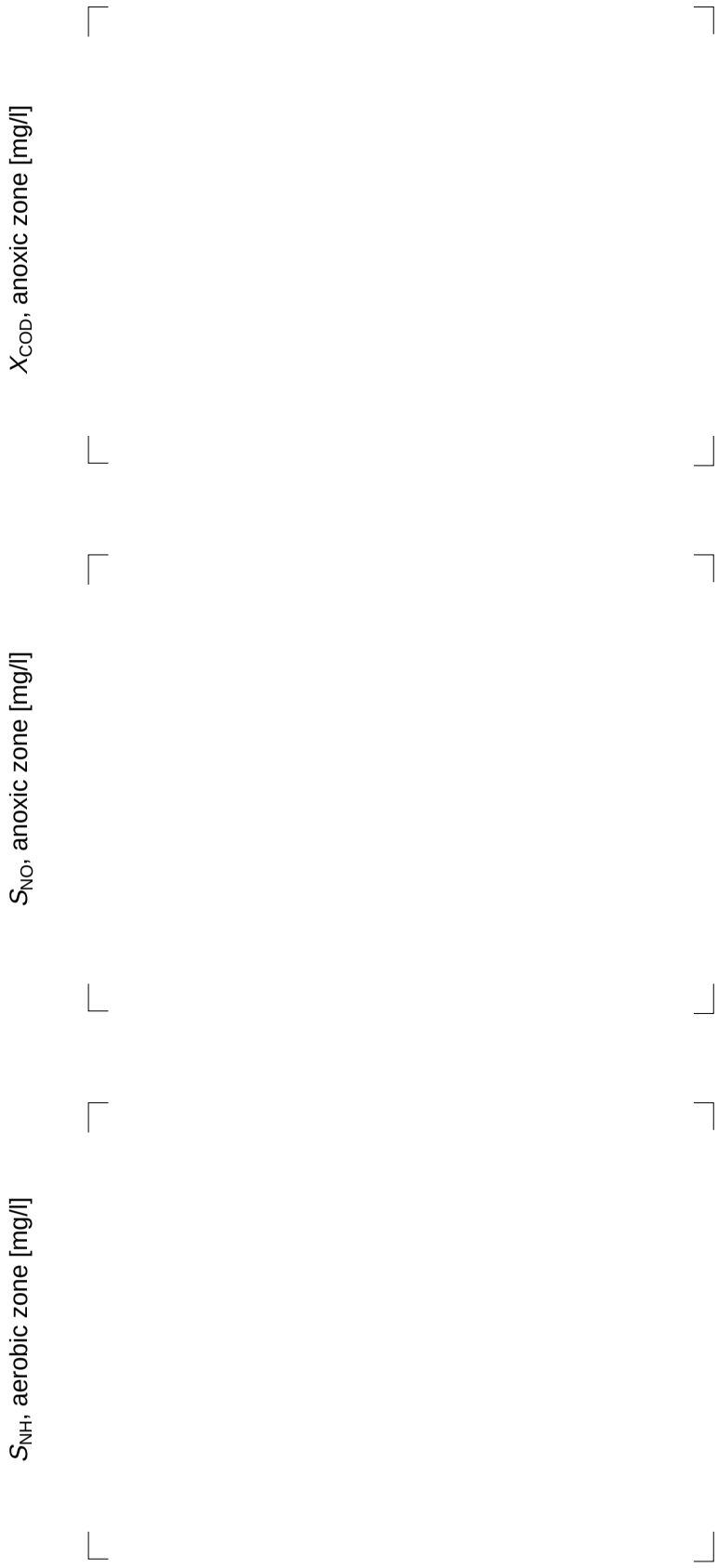
The behaviour of 'model A' is verified against data from the IAWPRC model by simulating the systems separately but with identical influent wastewater characteristics, flow rates, recirculation rates, etc. In Figure 5.9 such a comparison is illustrated for some of the state variables.

All plots show that the qualitative behaviour of the two models are quite similar. The first three plots also indicate that the optimization in regard to X_{COD}, S_{NH}, and S_{NO} is practically identical for both cases 3A and 3E although based on different amounts of information. The reason for the discrepancy of X_{COD} between the IAWPRC and reduced model has already been commented (the hydrolysis process, the changing ratio of S_S and X_S, and the settler behaviour).

The third plot shows a significant difference between the two models in the predicted ammonia concentration, especially around day five. This is mainly because the ammonia half-saturation parameter of the IAWPRC model is set to 1 mg/l. As the S_{NH} concentration is increased from 1 to 3 mg/l (the most nonlinear part of the applied Monod expression) in the aerobic basin it is practically impossible to achieve the same result with the first order approximation of the nitrification process in 'model A'. Therefore the nitrification rate increases more rapidly in the simplified model when the ammonia concentration goes up and the high peak predicted by the IAWPRC model is flattened. Due to this problem a slightly higher concentration of nitrate in the anoxic reactor is also predicted by the reduced model during the same time period (second plot).

The fourth plot illustrates how differently the quantitative concentrations of microorganisms are predicted for cases A and E. This is naturally a major reason for the different sets of parameters found. If instead the concentrations are normalized around the steady state values, the results are almost inseparable.

Finally, the *OUR* is shown in the fifth plot. It appears remarkable that the predictions from case 3A are better than from case 3E although the *OUR* was not considered measurable for the first case. The large error for case E is due to the simultaneous optimization of the organism concentrations. While the predictions of $X_{B,H}$ and $X_{B,A}$ are improved, the *OUR* is deteriorating. There is a conflict between these quantities.


X_{COD}, anoxic zone [mg/l]
S_{NO}, anoxic zone [mg/l]
S_{NH}, aerobic zone [mg/l]
time [day]

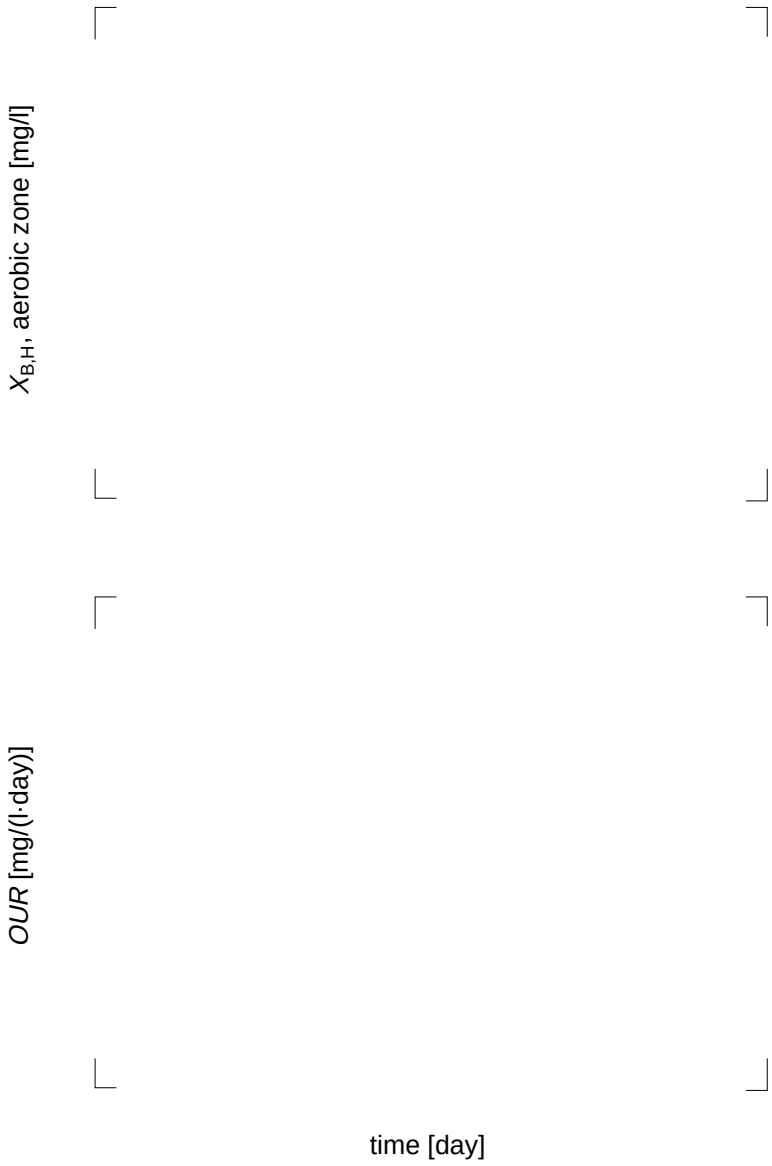


Figure 5.9 Behaviour of a complete WWT plant simulated with 'model A' (dashed - case 3A(row 2), dotted - case 3E(row 3)) and compared to simulations of the IAWPRC model (solid) with identical influent wastewater characteristics (see Figure 5.2).

If the *OUR* for case 3B would have been presented, it would show a perfect agreement with the IAWPRC results (as the value of the loss function indicates). On the other hand, the heterotrophic concentration would have a stationary value close to 800 mg/l.

Case 4 – anoxic/aerobic reactor combination, model B

The results from case 3 show that the problems concerning global identifiability are apparent. Even when a large number of quantities are assumed measurable it is practically impossible to obtain an optimum global set of parameters. This is mainly caused by the effects of recirculation and the varying time constants of the system. However, the problem is also due to the high model complexity. Although 'model A' is a quite straightforward model for the activated sludge process when compared to the IAWPRC model, there is still ten parameters to be identified simultaneously (if one anoxic and one aerobic reactor are assumed for the model) during standard operating conditions (no specialized identification experiments) for a WWT plant.

In an attempt to further reduce the number of degrees of freedom for the model and improve its global identifiability, 'model A' is further reduced into 'model B'. This is done simply by assuming the decay rate factors b_H and b_A to be identical, both for anoxic and aerobic conditions. The number of model parameters are hereby reduced from ten to seven. Moreover, the difficulty of identifying b_H and especially b_A, which has been discussed for the previous cases, may also be reduced since the effect of each of the four parameters on the model behaviour are now combined into one. Therefore the possibility for the optimization algorithm to detect significant changes in the value of the loss function is improved, i.e. the model sensitivity to that parameter is enhanced.

From a biological point of view there is no motive to differentiate the two decay rate factors depending on the applied condition (anoxic or aerobic) since the microorganisms are circulated through the plant and exposed to both situations during their life cycle. As for the assumption that the decay rates for both Heterotrophs and Autotrophs are the same, this is more debatable. However, both b_H and b_A are rough average values because Heterotrophs and Autotrophs are large groups which consist of many different organism species with individual variations. Moreover, when exposed to the same operational

conditions and identical influent wastewater there is no obvious microbiological reason to assume different decay rate coefficients for the two groups.

It should be noted that the quite different values for b_H and b_A suggested in the IAWPRC model (see Table 2.2) mainly depends on the fact that b_A represents a traditional decay rate coefficient whereas b_H does not. This follows from the fact that the recycling of organic matter that results from decay (the death-regeneration hypothesis, see section 2.2) occurs through the activity of the heterotrophic biomass and not the autotrophic biomass [Henze et al., 1987a]. By applying the formula:

$$b_H = \frac{b_{H,trad}}{1 - Y_H(1 - f_P)} \tag{5.2}$$

where: $b_{H,trad}$ = traditional decay rate factor;

with default values for the other parameters and calculating $b_{H,trad}$, the difference between b_H and b_A are no longer very significant but more a result of the interpretation of the model.

In most AS models the decay coefficient is important in order to predict the sludge production and the oxygen consumption. This is not the purpose of the simplified models presented in this work. Since the reduced models only deal with active biomass and all material which results from decay is transformed into X_{COD} and S_{NH} directly, the so called specific decay rate factor in this case actually represents a transformation rate coefficient, describing this process. To compare identified values of b_H and b_A from the reduced models with values normally applied to traditional AS models may therefore prove without any real relevance. This is a major drawback of reduced models.

The reasons discussed above motivates the simplification of 'model A' into 'model B'. It is also considered important to apply a model structure which enhances parameter identification and automatic calibration from full scale plant operation. The possibility of determining the traditional b_H from lab experiments by monitoring the *OUR* exists although such an experiment

usually requires several days and is based on a small sample of sludge. To practically determine b_H of the IAWPRC model is much more complicated (unless Y_H and f_P are assumed perfectly known) since it includes the transformation of decayed material into organic matter. Also b_A is difficult to determine from actual experiments.

The optimization algorithm is now applied to 'model B' under exactly the same conditions as were described in case 3. The selection of cases to investigate (A to E), depending on which quantities are assumed measurable, are also identical. In Table 5.5 the results of the optimization are presented for different sets of initial estimates.

When the results are compared to case 3, there are three important observations to be made immediately. The first is the fact that 'model B' is capable of mimicing the behaviour of the IAWPRC model practically as well as 'model A' although the number of model parameters have been reduced from ten to seven. A small increase in the value of the loss function (10-15%) is noticeable for the cases where the microorganism concentrations are considered measurable (C, D, and E) whereas the results for cases A and B are actually improved.

Secondly, the variation of the values for the optimized sets of parameters are significantly smaller for 'model B' when compared to 'model A'. This also leads to a smaller variation in the values of the loss function for each case which more than adequately compensates for the somewhat higher values for some cases. The possibility of finding the truly best optimum (or an optimum very close to it) is thereby greatly enhanced.

Finally, the algorithm appears to converge towards a practically global optimum parameter set not only for case E but also for case B. This is a dramatic improvement since the assumption of being able to monitor the different organism concentrations on-line (case E) is not realistic. On the other hand, the measurements assumed for case B are not impossible to achieve.

Optimization 'model B'														value of loss func.
initial estimates (anoxic)							final estimates (anoxic)							
r_H	Y_H	r_H	r_A	Y_H	Y_A	b	r_H	Y_H	r_H	r_A	Y_H	Y_A	b	
Case 4A (measured: X_{COD}, S_{NH}, S_{NO})														
.024	.35	.046	.111	.33	.10	.20	.011	.772	.018	.209	.635	.577	.114	992
.046	.69	.093	.222	.66	.20	.40	.011	.754	.017	.197	.615	.160	.101	958
.057	.86	.116	.278	.83	.25	.50	.014	.700	.012	.167	.434	.226	.073	1171
.068	1.03	.140	.333	.99	.30	.60	.015	.780	.027	.302	.679	.188	.209	1075
Case 4B (measured: X_{COD}, S_{NH}, S_{NO}, *OUR*)														
.024	.35	.046	.111	.33	.10	.20	.016	.697	.024	.282	.564	.236	.189	1192
.046	.69	.093	.222	.66	.20	.40	.016	.697	.025	.295	.568	.209	.202	1193
.057	.86	.116	.278	.83	.25	.50	.016	.697	.021	.257	.539	.055	.163	1185
.068	1.03	.140	.333	.99	.30	.60	.016	.700	.025	.287	.568	.206	.195	1189
Case 4C (measured: X_{COD}, S_{NH}, S_{NO} plus $X_{B,H}$, $X_{B,A}$, *OUR* in aerobic reactor)														
.024	.35	.046	.111	.33	.10	.20	.018	.631	.033	.393	.476	.363	.309	2363
.046	.69	.093	.222	.66	.20	.40	.015	.567	.023	.290	.403	.258	.196	2507
.057	.86	.116	.278	.83	.25	.50	.018	.564	.028	.330	.443	.299	.238	2431
.068	1.03	.140	.333	.99	.30	.60	.034	.699	.052	.609	.586	.550	.533	2509
Case 4D (measured: X_{COD}, S_{NH}, S_{NO} plus $X_{B,H}$, $X_{B,A}$ in anoxic reactor)														
.024	.35	.046	.111	.33	.10	.20	.022	.600	.027	.343	.418	.310	.253	1585
.046	.69	.093	.222	.66	.20	.40	.021	.588	.031	.374	.447	.330	.281	1604
.057	.86	.116	.278	.83	.25	.50	.017	.531	.019	.244	.330	.216	.147	1659
.068	1.03	.140	.333	.99	.30	.60	.020	.572	.021	.274	.356	.248	.181	1614
Case 4E (measured: X_{COD}, S_{NH}, S_{NO}, $X_{B,H}$, $X_{B,A}$, *OUR*)														
.024	.35	.046	.111	.33	.10	.20	.026	.633	.032	.398	.465	.364	.314	2450
.046	.69	.093	.222	.66	.20	.40	.024	.622	.034	.408	.479	.372	.323	2448
.057	.86	.116	.278	.83	.25	.50	.025	.628	.034	.408	.474	.371	.323	2447
.068	1.03	.140	.333	.99	.30	.60	.025	.631	.035	.425	.487	.385	.340	2449

Table 5.5 Results of the parameter optimization for case 4.

When the five investigated cases are more closely examined, the major part of the conclusions from case 3 still hold. In case 4A a strong correlation between b and r_H, r_A in the aerobic reactor is clear. Since the organism concentrations converge towards quite different values the yield coefficients also vary significantly.

The inclusion of *OUR* in the calculation of the loss function in case 4B leads to an almost global set of optimum parameters. In case 3B only the convergence of the heterotrophic parameters in the aerated reactor were improved to this extent. The sole major discrepancy is seen in row 3 for the autotrophic parameters. In this case the concentration of $X_{B,A}$ is about one fourth of what is predicted by the other rows and consequently, Y_A is four times smaller. The main reason for this is that the autotrophic parameters are less sensitive to the *OUR* because the concentration of $X_{B,A}$ is very low compared to $X_{B,H}$. Apart from this the results are remarkably consistent and is clearly the greatest advantage of 'model B'.

Case C shows a number of local optima with approximately the same values of the loss function but with very different parameter sets. Row 2 and 4 illustrate this very well. For the parameters in row 2, the *OUR* is somewhat better predicted whereas the $X_{B,H}$ concentration is more accurately predicted by row 4. The differences are, however, quite small and the total value of the loss function is the same. The parameter values of row 4 is clearly a local optimum whereas row 1, 2, and 3 show much more consistent values. The increased loss function when compared to cases A, B, and D is an indication of the difficulty of a simultaneous good prediction of both the organism concentrations and the *OUR* for the reduced 'model B'. The same fact was observed for 'model A'. The main reason for this is once again the different description of the organic substrate in the reduced models and the IAWPRC model. Note that the ratio of r_A and Y_A is identical for all examples in cases C, D, and E (when $X_{B,A}$ is considered measurable) and the correlation between b and r_H, r_A in the aerobic reactor is apparent.

In case D the *OUR* is excluded from the optimization. This leads to that also the ratio of r_H and Y_H converges towards almost the same values for all examples in both the anoxic and the aerobic reactor.

Finally, a global optimum is reached for case E. Due to the assumed measurements of $X_{B,H}$ and $X_{B,A}$ in both reactors the effect of the *OUR* measurements is reduced and the convergence is improved when compared to case C.

The behaviour of 'model B' is verified against data from the IAWPRC model by simulating the systems separately but with identical influent wastewater characteristics, flow rates, recirculation rates, etc. In Figure 5.10 such a comparison is illustrated for some of the state variables. The behaviour of the other variables (organic substrate, ammonia, and nitrate) are very similar to the ones presented in Figure 5.9.

The first two plots show how the concentration of Heterotrophs is overestimated and the concentration of Autotrophs is underestimated when the *OUR* is used for the optimization and measurements of the organisms are assumed unavailable (case B). On the other hand, the predictions of the *OUR* are for this case quite accurate (third plot).

In order to decrease the predicted concentration of Heterotrophs for case E the yield factor (Y_H) is reduced (among other things). Consequently, to increase the predicted concentration of Autotrophs, the yield factor (Y_A) is almost doubled (see Table 5.5). This has an immediate effect on the *OUR* . The factor $(1-Y_H)/Y_H$ in the expression (4.25) is doubled which however is compensated by the fact that the $X_{B,H}$ concentration is reduced with almost 50%. But since r_H is simultaneously increased to maintain the good predictions of X_{COD} and the denitrification process the predicted oxygen uptake rate reaches too high values. The *OUR* for the Autotrophs are on the whole quite constant for both cases B and E since the factor $(4.57-Y_A)/Y_A$ is decreased by 50% while the factor $r_A \cdot X_{B,A}$ is increased by the same amount and compensates for the very different yield factor.

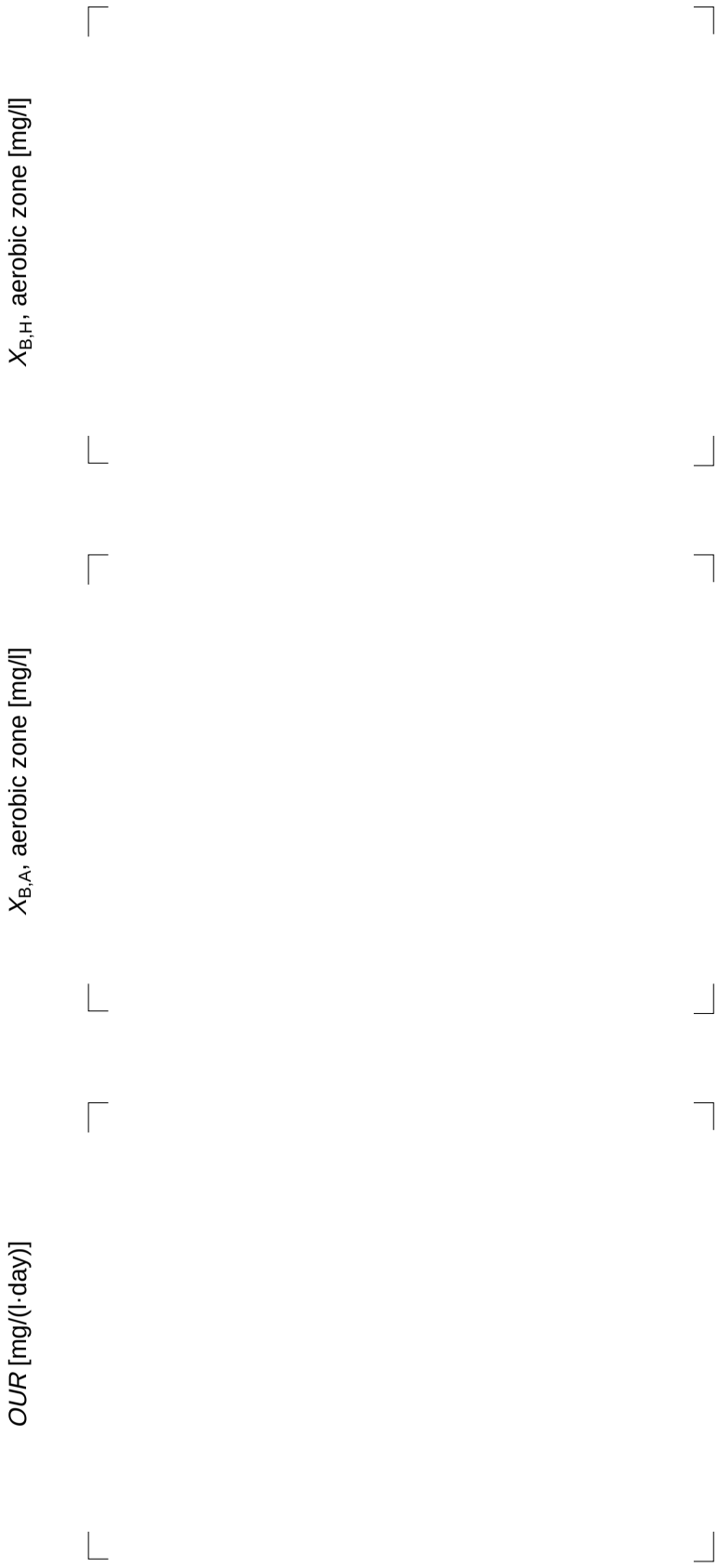

$X_{B,H}$, aerobic zone [mg/l]
$X_{B,A}$, aerobic zone [mg/l]
OUR [mg/(l·day)]
time [day]

Figure 5.10 (Previous page) Behaviour of a complete WWT plant simulated with 'model B' (dashed - case 4B(row 1), dotted - case 4E(row 3)) and compared to simulations of the IAWPRC model (solid) with identical influent wastewater characteristics.

Summary discussion

A number of case studies have been presented in this section to illustrate the behaviour and identifiability properties of the reduced models in regard to operational conditions, assumed measurable quantities, initial parameter values, etc. Naturally the investigations do not cover all important aspects of the models but provide a basis for further analysis.

In cases 1 and 2 it was shown that the reduced model was capable of mimicing the behaviour of the IAWPRC model with reasonable accuracy in unit operation (only one type of reactor). The main differences were due to the fact that the reduced models do not include a hydrolysis process for the organic matter which affects the time constants of the models. Difficulties to determine global optimum parameter sets for the simplified model were also clearly indicated.

This difficulty was further investigated by examining the behaviour of the loss function when the model parameters were changed. The models proved especially insensitive to variations of the decay rate coefficients but also r_A and Y_A were impossible to determine separately under the applied conditions and available measurements.

In case 3 the complete 'model A' was used to simulate a WWT plant with both anoxic and aerobic reactors and the results were compared to simulations of an identical plant simulated with the IAWPRC model. The recirculation introduced new problems when trying to detect a global optimum parameter set by means of optimization. Not even under extremely favourable conditions could the autotrophic decay rate factor be uniquely determined.

This motivated the use of a yet further simplified model - 'model B'. Here all the decay rate factors of 'model A' were identical. The approach significantly improved the optimization result without

deteriorating the basic behaviour of the model, as was shown in case 4. What appeared to be an almost global parameter set could be determined without assuming the concentrations of microorganisms to measurable.

However, it seems troublesome to uniquely identify the parameters of the models based on measurements of X_{COD}, S_{NH}, and S_{NO} alone. In many cases, when measurements of the *OUR* was considered available the results were significantly improved, especially for parameters describing the heterotrophic biomass in case 3 and for all parameters in case 4. New technology like on-line respirometers which are currently being developed (see section 2.5) may therefore provide a very useful tool for identification and verification of the AS process. A main reason for this is that the *OUR* provides direct information about the activity of the process while traditional measurements of concentrations only give indirect knowledge.

In some cases it was assumed possible to measure the concentration of active heterotrophic and autotrophic biomass. This is unrealistic but useful when principal properties of a model is investigated. When $X_{B,H}$ and $X_{B,A}$ are not measurable, these quantities may converge towards very different values when the model is optimized, depending on the initial parameter estimates. This leads to non-unique parameter sets because high concentrations of microorganisms with a low activity result in approximately the same model behaviour as low concentrations of microorganisms with a high activity.

Since it is practically impossible to measure the concentration of *active* biomass directly, it may prove useful to base an AS model on estimations of actual reaction rates instead of rate coefficients. Knowledge of the reaction rates are necessary for process control whereas the rate factors are more important for biological interpretations concerning the state of the process. For example, estimations of $r_H \cdot X_{B,H}$ as a combined parameter and not as two separate variables could improve the identifiability of a model especially since the $X_{B,H}$ concentration is not verifiable. All parameters r_H, r_A, b_H, and b_A of the reduced models do always appear in multiplicative combinations with $X_{B,H}$ or $X_{B,A}$ and therefore an identification may be global in the sense that the

combinations converge towards identical values but assume practically any value if all parameters are analysed separately. Simple measurements like the suspended solids concentration or the total COD content may also prove useful since they contain a certain amount of information about the microorganism concentration (though as a complicated combination of several quantities). Future work will include the development and analysis of an AS model which applies the concept of direct estimation of reaction rates.

For all case studies presented, the assumed measurements were perfect (i.e. no noise). An analysis of the impact of noise (Gaussian noise within reasonable limits) shows that the effect is negligible for this type of optimization since it is based on such a large number of measurements. The results of the optimization have also been verified against other types of dynamical behaviour of the influent wastewater than presented here (sine waves, ramps, etc.). The different behaviours of the reduced models and the IAWPRC model were, however, most prominent for steplike variations, which motivated the selection.

No thorough analysis has been performed on how the system should be perturbed in order to enhance the model identifiability further. Influent variations syncronized with the time constants of the model, repeated rapid changes of high amplitude, etc., may improve matters. But since the possibilities to control the influent wastewater to most WWT plants are quite limited, such an investigation would probably lead to results which are not practically feasible, though theoretically interesting.

As a final remark the computational effort is commented. In order to determine *one* set of optimized parameters the system of differential equations describing the dynamics of the plant has to be simulated between 500 and 1000 times (the simplex algorithm converges slowly). This means that the required CPU time for one optimization on a 'standard' workstation should be measured in hours and days rather than seconds and minutes. The work presented in this section is based on several hundred such optimizations. For practical use the algorithm would have to be modified and the error tolerance increased.

5.3 On-line Simulations

The main disadvantage of the optimization approach is that the resulting model is fitted to certain operational and influent conditions. The model is capable of predicting the behaviour of the real process as long as these circumstances are not dramatically changed. Over longer periods of time, smaller variations will accumulate in the plant and the behaviour of the microorganisms will be affected. Therefore it may prove necessary to update the model parameters on a regular basis. One possibility is to make a new optimization when needed but better still is to automatically update the model and track the parameters on-line as new measurements become available. It will guarantee that any required model predictions are as reliable as possible since the model is always calibrated to the most current conditions. In this section some results from such an approach are illustrated and discussed.

The off-line and the on-line methods should produce approximately the same results under identical operating conditions if the algorithms are applied for a sufficient amount of time. However, since the simplex method is more robust it has been chosen for the principal model investigations reported in 5.2. Although the Kalman filter (see Appendix E) is one of the more robust on-line estimation algorithms, difficulties may occur. This is mainly due to the recirculation and the long time delays of the activated sludge system. When a parameter is adjusted the effect on the system is only partly 'instantaneously' measurable. After a period of time (several hours) the input is also changed through the recirculation, which is interpreted as a new disturbance of the system and consequently may lead to new parameter adjustments. In certain cases this can lead to stability problems where the parameters are over-compensated back and forth. Especially the effect on the microorganism concentrations is very slow and may take days to become significant. To some extent it is possible to modify the estimation algorithms to take the correlation between a changing parameter and a later change in the recirculated material into account. Such modifications, however, have not been tested in this work. Instead the

maximum rate of change for the model parameters is set quite low (by adjusting the Kalman gain matrix).

The reasons given above motivate a combination of the two investigated methods. By first applying an off-line optimization algorithm, a set of reasonable parameter values can be determined. They may be used as initial seeds for the on-line estimation methods, thereby reducing the risc of divergence and unwanted oscillations.

When performing on-line identification on real data it is also more important to know about the noise characteristics of the process and the measurements. Whereas the off-line methods are based on a very large number of data points and the effect of noise is reduced by the averaging calculations, the on-line methods react practically immediately to sudden changes. By a thorough investigation of the noise character of the process this information can be included in the identification algorithm in order to prevent the estimated parameters to vary rapidly back and forth in an unrealistic manner due to noise. Furthermore the measurements should be 'logically' analysed before the data is used for on-line identification. The reason for this is to detect trends, outliers, drastic sudden changes, etc., which may indicate that a certain sensor is not properly calibrated or does not function as supposed to, before the identification procedures are applied and produce an erroneous result. This would be one important function of the top level expertsystem in Figure 4.1.

A number of cases will be presented below in order to determine the principle model behaviour when used for on-line parameter identification. The basic conditions will be those already described in section 5.1 and previously applied for the off-line optimization. The following situations will be more closely investigated:

- a case study of a modified 'model A' for various assumed available measurements and initial estimates as in section 5.2 (cases 3A, 3B, and 3E (without the *OUR*));
- the effect of both measurement and input noise on the identification results;
- the effect of a change in one of the IAWPRC parameters on the reduced model behaviour during on-line estimation.

Results from Tables 5.3 and 5.4 indicate that the decay rates (*b*) may assume almost any values especially when only X_{COD}, S_{NH}, and S_{NO} measurements are available. If all ten parameters of 'model A' are identified simultaneously with the extended Kalman filter this is manifested as divergence. The correlation between the decay rate factors and the other parameters often lead to a trend-like behaviour of the estimates, i.e. one parameter keeps changing at a certain rate and another variable changes with the same relative rate and the total effect on the measurable state variables is negligible. Finally, the model identification process collapses.

To avoid this problem the elements of the Kalman gain matrix related to the decay parameters are set to zero, e.g. the decay rate coefficients are kept constant. The more variables that are measurable the more parameters may be estimated successfully. If the *OUR* or the organism concentrations are assumed to be measurable as well, one or two of the decay parameters may be estimated too. All ten parameters, however, can not be estimated simultaneously for those cases either. In order to allow the results of the different estimations to be compared, all decay parameters are kept constant in this section. The model applied for the on-line estimations is consequently a simplification of models A and B since only six parameters are assumed to vary.

The basic problem is the one discussed in section 5.2. It is 'quite easy' to estimate the net reaction rates for the organisms but much more difficult to determine the growth rates and the decay rates separately. Future models will be developed taking this fact into account.

In Figure 5.11 the results from an on-line estimation are shown for some key variables. The measurements assumed available for the identification are the ones described for case 3A (X_{COD}, S_{NH}, and S_{NO} i both reactors) but the process is now stationary. In order to start the on-line estimation with reasonable initial parameter values, the off-line results from case 3A(row 2) are used (solid line). This parameter set is then increased (dashed) and decreased (dotted) by 15% (not the decay rate coefficients) and the identification is repeated for the new sets to check for 'global' convergence.

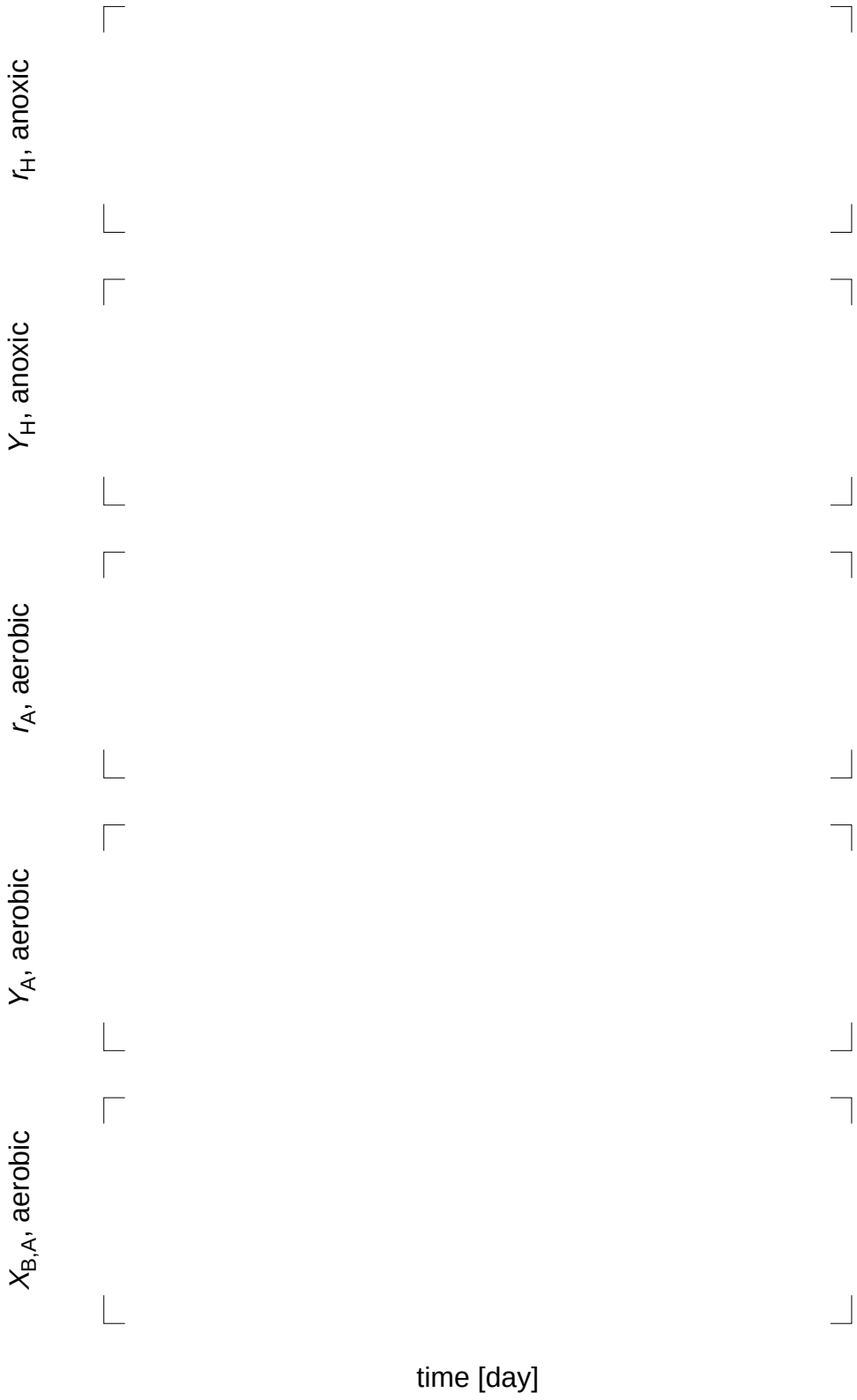


Figure 5.11 On-line state and parameter estimation of a simplified 'model A' based on assumed measurements of the organic substrate, ammonia, and nitrate concentrations from a stationary AS process.

The graphs illustrate that the heterotrophic parameters (shown for the anoxic zone) converge towards the same values independent of the initial estimates. This is possible since the decay rate factors are identical for all simulations (if not, the identification algorithm would be forced to find different estimates). For the autotrophic parameters, however, these measurements are not sufficient. The autotrophic reaction rate converges towards the same value whereas the yield factor reaches quite different values (it does converge though). Since $X_{B,A}$ is not assumed measurable, this variable may compensate for the different yield values. A high value of the yield factor leads to a high concentration of Autotrophs and vice versa. The ratio of $X_{B,A}$ to Y_A converges towards the same value but the Kalman filter can not determine a global estimate for the two variables separately, especially not under steady state conditions. The entire model is also less sensitive to variations of the autotrophic parameters due to the low concentration of Autotrophs when compared to the Heterotrophs.

The low sensitivity is a major reason why the convergence rate is considerably slower for the autotrophic parameters, which is clearly illustrated in Figure 5.11. This is natural since the time constants for changes of the autotrophic biomass are larger than for other reactions in the activated sludge process. Furthermore, the convergence rates for all parameters are very low in the first examples because of the steady state data which leads to very small residuals and consequently a slow convergence. The applied simplified Kalman filter (constant gain matrix) further enhances this. The rate of convergence is also correlated to the amount of available measurements (see Figure 5.13) and the number of parameters to be estimated. Comparisons of the convergence rates for the different parameters and cases are informative.

The estimation procedure described above is repeated to obtain the results shown in Figure 5.12. In this case the identification is based on the assumption that the *OUR* is also measurable. The first initial parameter set is taken from case 3B(row 2) and then varied ±15%. Apart from this, the same data from the IAWPRC model simulating a complete WWT plant under stationary conditions is applied.

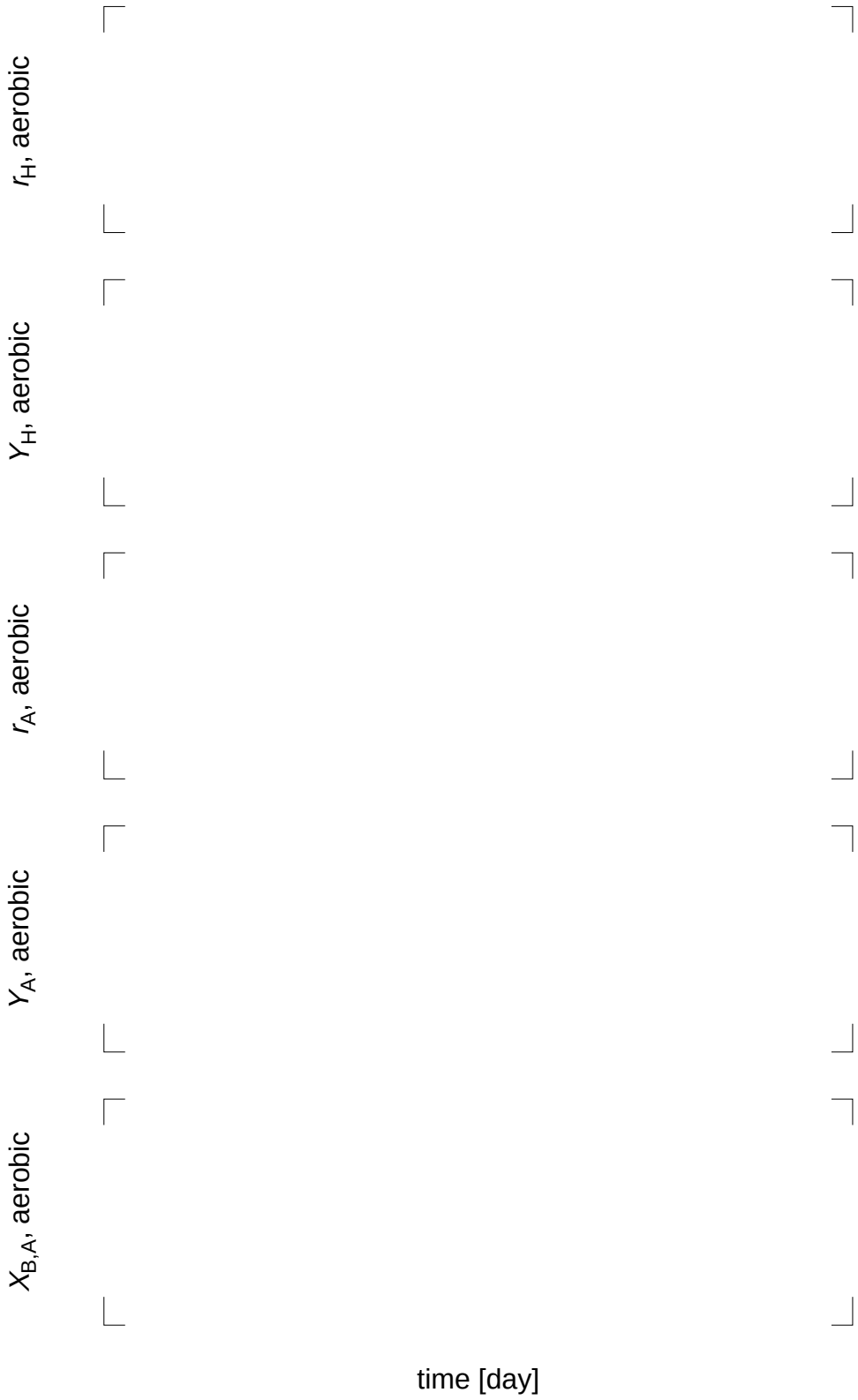


Figure 5.12 On-line state and parameter estimation of a simplified 'model A' based on assumed measurements of the organic substrate, ammonia, and nitrate concentrations plus the *OUR* from a stationary AS process.

The graphs of Figure 5.12 show that the additional information from the *OUR* measurements is not enough for the identification algorithm to determine a global set of the estimated variables. It was already discussed in section 5.2 that the effect of the *OUR* is most apparent on the heterotrophic parameters because of the low concentration of Autotrophs. The autotrophic yield factor and the concentration of autotrophic biomass still converge towards different values depending on the initial set. The situation is however somewhat different when compared to the previously investigated case. The ratio of $X_{B,A}$ to Y_A does not converge globally. Rather the weighted mean value of this ratio and the expression $(4.57\text{-}Y_A)\cdot X_{B,A}/Y_A$ which is part of the calculation of the oxygen uptake rate (4.25) converge in a global sense. Thereby the reduced model may provide a good fit to the data from the IAWPRC model for several sets of estimates also when the *OUR* is available.

In Figure 5.12 no estimations of the heterotrophic parameters in the anoxic reactor are shown. For this zone the *OUR* does however not provide any new information and the estimation results are therefore very similar to what was illustrated in Figure 5.11. A comparison of r_H and Y_H in the two figures indicate that the rate of convergence is considerably higher for the anoxic reactor. The main reason for this is the smaller number of variables to be estimated simultaneously in the anoxic zone. A similar examination of the autotrophic variables in Figures 5.11 and 5.12 show a higher rate of convergence when the *OUR* is measurable. Its effect is not enough to determine a global set of estimates but is still a help for the algorithm. The more information available, the faster and more accurate is the convergence.

Finally, the identification procedure is tested on a new case. Now the *OUR* is not considered measurable; instead the organism concentrations ($X_{B,H}$ and $X_{B,A}$) for both reactors are assumed available. Some results are presented in Figure 5.13. The initial parameter set is taken from case 3E(row 3). Note that two different time scales are used in the diagrams. Neither of the scales are the same as those applied in Figures 5.11 and 5.12.

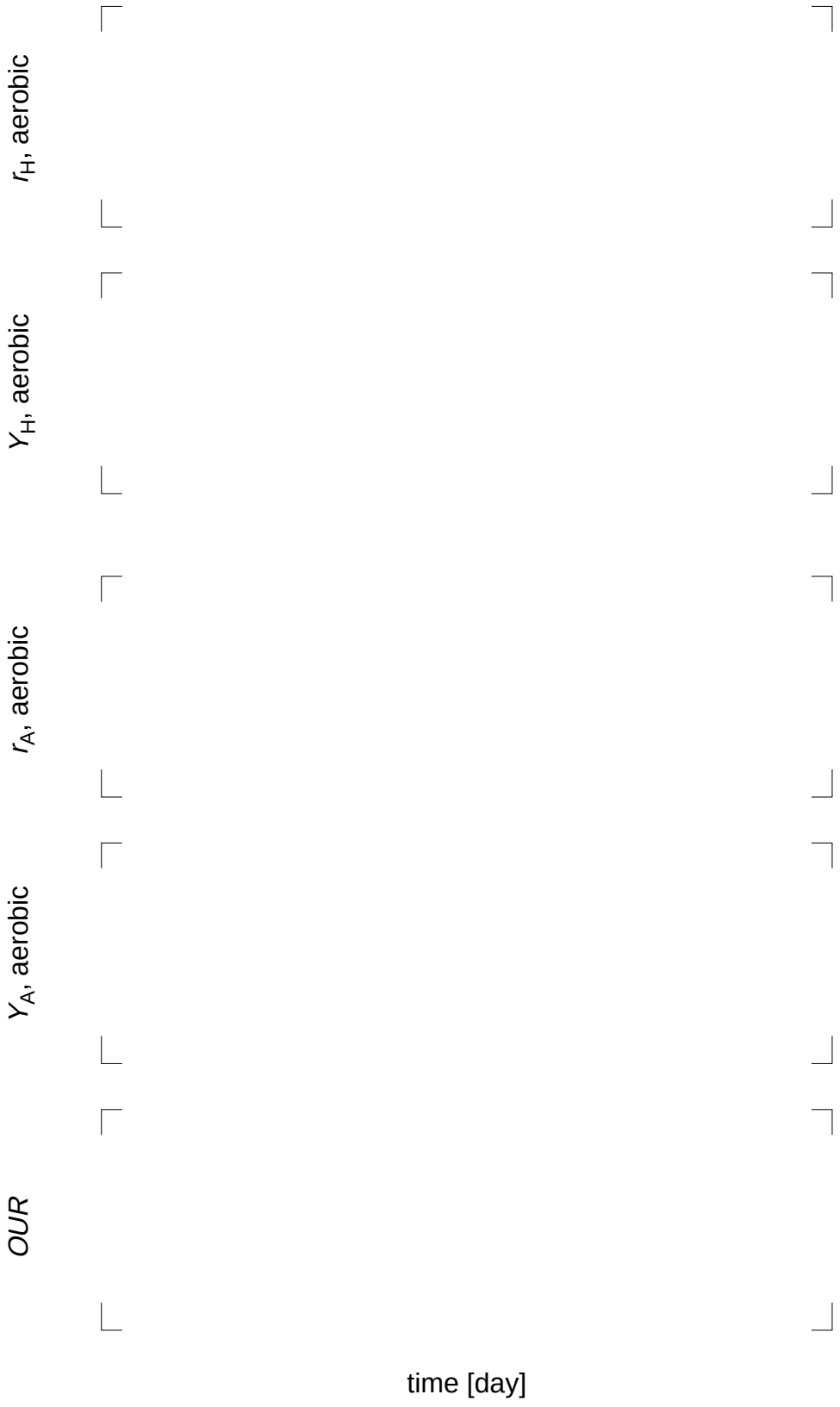


Figure 5.13 On-line parameter estimation of a simplified 'model A' based on measurements of the organic substrate, ammonia, nitrate, and organism concentrations from a stationary AS process.

The graphs illustrate how all parameters converge globally for the different initial sets. The rate of convergence is also significantly higher than for the earlier shown cases. In order to determine a global set of estimates it would actually be sufficient to assume measurements of $X_{B,A}$ from the aerobic reactor. Possibly, assumed measurements of the total organism concentration ($X_{B,H} + X_{B,A}$) would also be enough instead of requiring data from both types of organisms separately. This would, however, lead to slower convergence. The prediction of the oxygen uptake rate based on the estimated parameters also converges and, moreover, towards the same value as predicted by the IAWPRC model (244 mg O_2/(l·day)).

The actual values of the parameters determined in this section are not identical when compared to the optimum sets found in section 5.2. The reason for this difference is that the scaling of the residuals performed by the gain matrix of the Kalman filter is not exactly the same as for the weighted loss function used by the simplex method. The off-line optimization was also based on data generated during varying influent conditions whereas the on-line identification uses steady state data.

The large initial variations for some of the estimates shown in the figures are not an error produced by the identification algorithm. They are the result of the very large residuals which appear at the early stage of the estimation. This is because the reduced model is simulated towards steady state prior to the identification and the stationary values are used as initial predictions of the simplified model. Consequently, they may be quite different when compared to the assumed measured variables of the IAWPRC model. The use of a constant gain matrix for the Kalman filter also emphasizes the large initial variations.

In order to investigate how noise affects the results of the estimations the assumed measurements are corrupted by Gaussian white noise with a mean value of zero and a standard deviation which is 10% of the actual value of the specific variable. Noise of the same type is also added to the flow rate, organic substrate concentration, and ammonia concentration of the influent wastewater to the IAWPRC model to further complicate

the situation. Otherwise the conditions are identical to those that have already been described. Note that the noise added to the different variables is uncorrelated.

To illustrate the effect of the applied noise level to the IAWPRC model and to examplify the difficulties for the identification algorithm, two of the corrupted variables are shown in Figure 5.14.

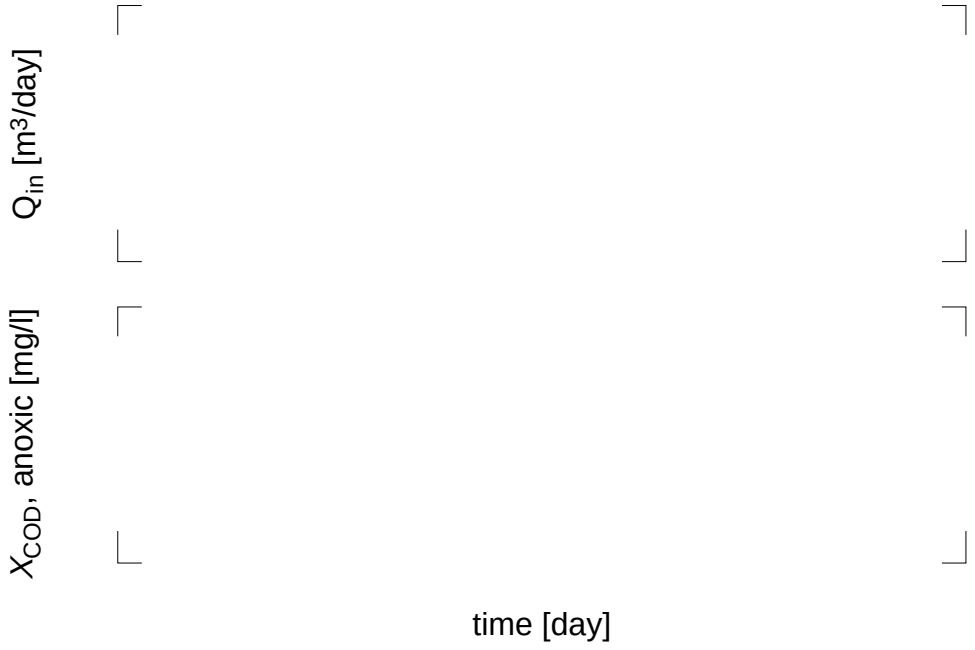


Figure 5.14 The effect of the applied noise level on the influent flow rate and measurements of the organic substrate concentration in the anoxic reactor.

Noise creates a problem for the identification algorithm. The residuals on which the Kalman filter determines how the variables are to be updated are dramatically affected by noise and may very well hide the basic underlying trend and thereby cause the identification to fail. This difficulty is most apparent when the process is stationary (small residuals) and the situation is normally improved by exciting the system in a suitable manner. Since the influent wastewater to a real WWT plant is only to a small extent controllable, such perturbations may be difficult to produce in reality.

For this reason the identification is performed using noisy data under steady state conditions and thereby giving the algorithm every possibility to fail. Investigations were performed on all three basic cases (depending on which quantities are assumed measurable and the selected set of initial estimates) as were illustrated in Figures 5.11, 5.12, and 5.13. For all cases the behaviour of the estimates was approximately the same as in those figures. The identification algorithm converge towards the same values as when the data was free from noise, independent of the chosen initial values. The estimates are naturally partly corrupted by the noise and not constant though the Kalman filter reduces the effect of the noise. The result would be significantly improved if a time variable extended Kalman filter was used since this would allow the gain matrix to be updated on-line and thereby reduce the variations of the estimates further.

A few results are illustrated in Figure 5.15. The assumed measurable variables and the initial estimates are here the same as was used for one of the estimations shown in Figure 5.13. In order to allow for a closer comparison with the previous results, the estimates are filtered using a traditional low pass filter and are shown together with the estimates produced when no noise was present, for some key variables in Figure 5.16.

The last example of this section demonstrates the capability of the reduced model to track and update its parameters on-line to maintain a good fit to the measurable variables as conditions change. In order to simulate this the maximum specific hydrolysis rate (k_h) of the IAWPRC model is increased by 50% over a period of one day starting at t=1. The hydrolysis rate is selected as a suitable parameter because of its significant influence on the complete IAWPRC model behaviour. In reality this change would reflect that the incoming slowly biodegradable organic substrate was of a new character which was more easily broken down into readily biodegradable substrate. A sudden increase of the incoming flow rate (50%) lasting for one day and starting at t=6 is also simulated to investigate the behaviour of the reduced model to such a change. Note that no noise has been added to the measurements in this case.

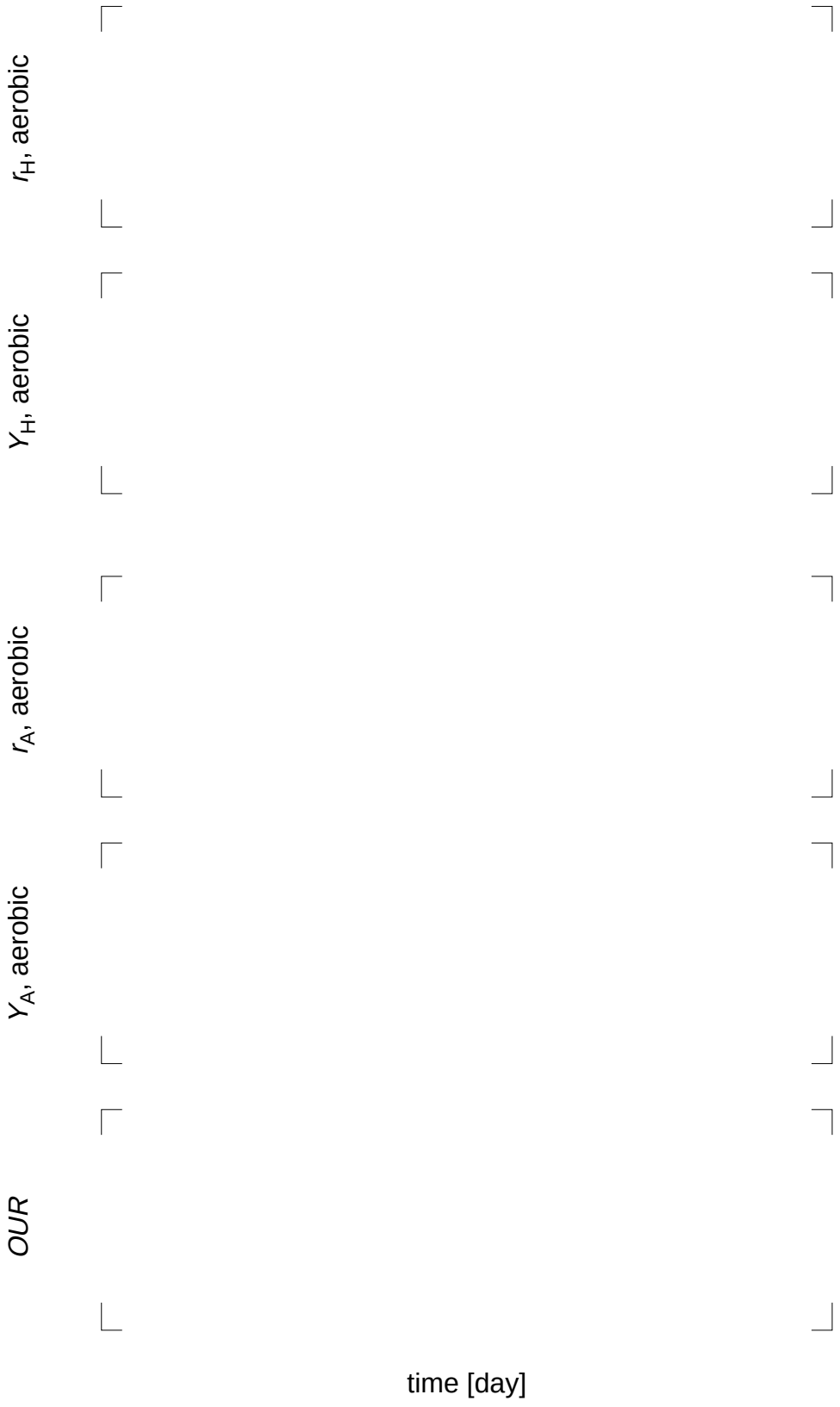


Figure 5.15 On-line parameter estimation of a simplified 'model A' based on noisy measurements of the organic substrate, ammonia, nitrate, and organism concentrations from a stationary AS process.

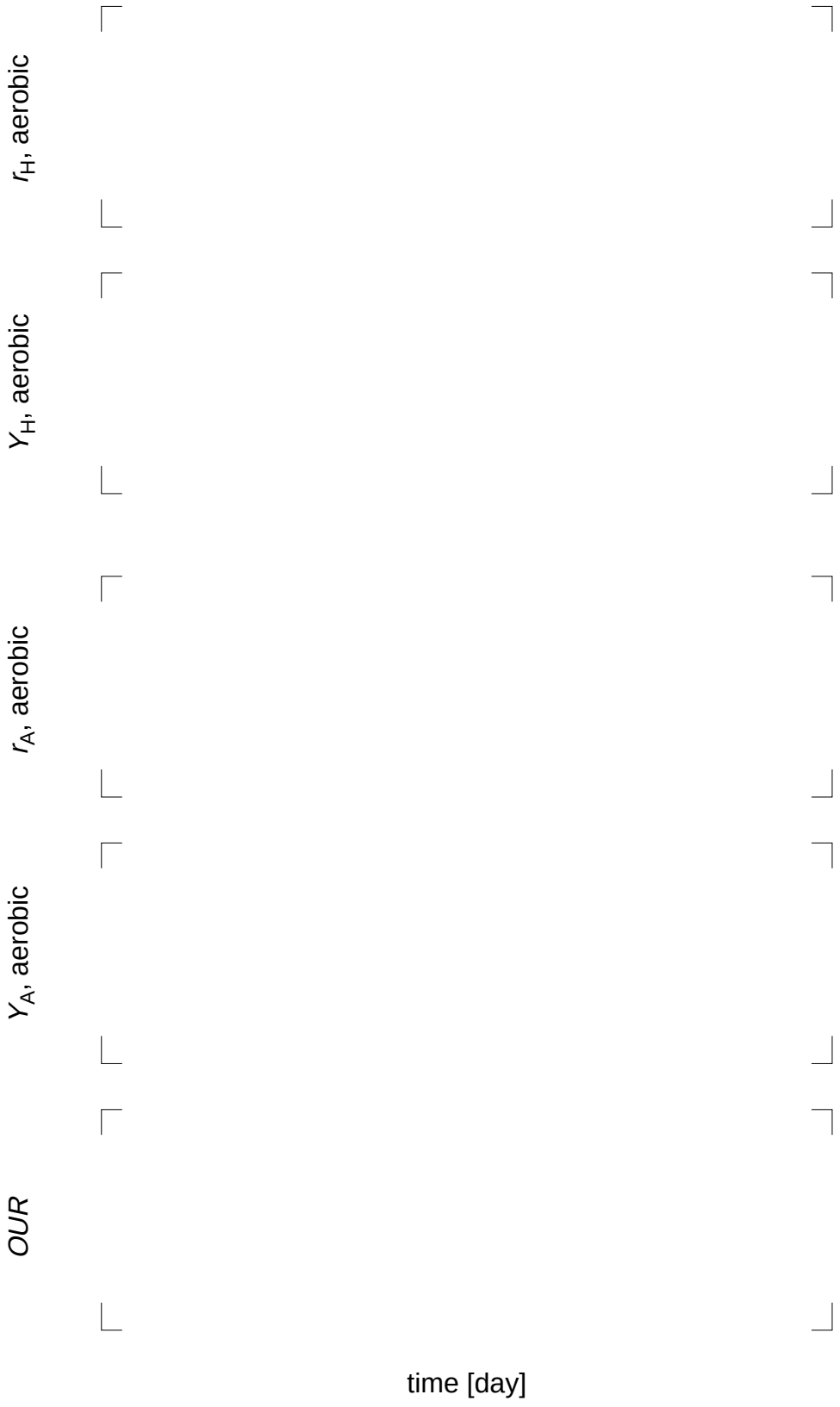


Figure 5.16 Lowpass filtered estimates (solid) from Figure 5.15 compared to the same estimates determined when no noise was present (dashed) from Figure 5.13.

The effects on the behaviour of the IAWPRC model when the hydrolysis rate is increased are mainly a considerable reduction of the amount of slowly biodegradable substrate in both reactors (in this case about 50% in steady state) and a lower concentration of nitrate in both reactors since the extra produced readily biodegradable substrate improves the denitrification. The concentration of Heterotrophs is also slightly increased. The effect on the ammonia concentration and the autotrophic biomass is negligible.

All three cases earlier described (depending on the available measurements) have been investigated. It is interesting to note that the model parameters are qualitatively updated in the same way whatever measurements are considered available, as seen in Figure 5.17. Note that the convergence rate is not a problem for any of the cases because the process is now partly in a transient state due to the changes of the hydrolysis rate and the incoming flow rate; consequently the rate of convergence is much higher. Prior to the time shown in the graphs, the Kalman filter has been applied to steady state data and the parameters of the reduced model have converged in order to avoid initial transients.

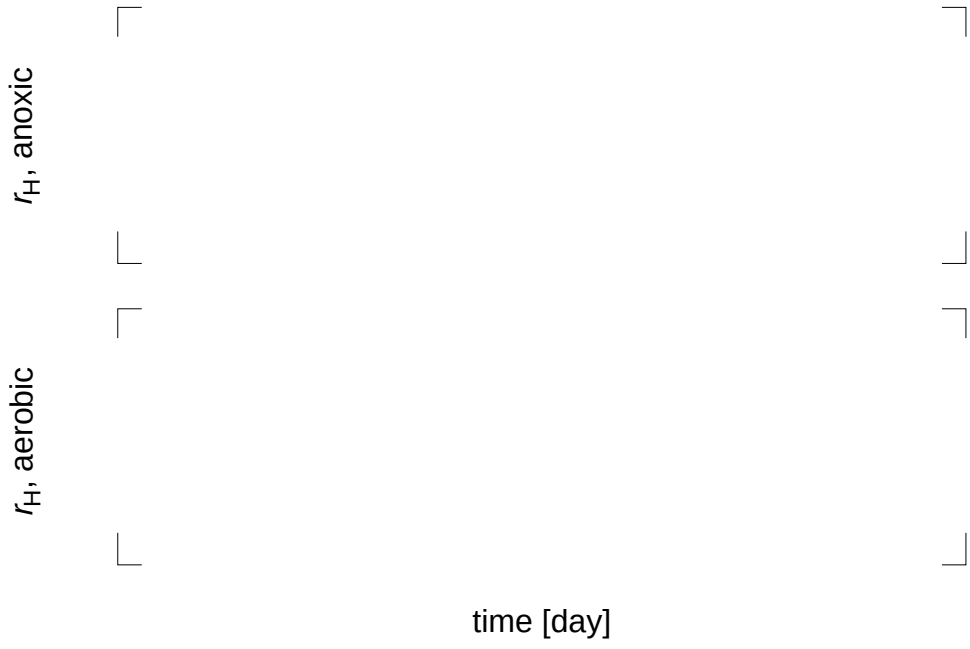


Figure 5.17 On-line parameter tracking based on different available measurements – X_{COD}, S_{NH}, S_{NO} (dashed), plus OUR (dotted), plus $X_{B,H}$ and $X_{B,A}$ (solid). The hydrolysis rate is increased from t=1 to 2 and the influent flow rate is increased at t=6 (pulse).

Figure 5.18 illustrates some results of the on-line parameter estimation when X_{COD}, S_{NH}, S_{NO}, $X_{B,H}$, and $X_{B,A}$ are assumed measurable. Predictions of the oxygen uptake rate is also shown and compared to the result from the IAWPRC model (dotted). As the hydrolysis rate of the IAWPRC model increases, the total amount of biodegradable organic substrate ($S_S + X_S$) goes down and consequently measurements of X_{COD} is reduced. As earlier discussed, changes of the ratio of S_S to X_S has a significant influence on the behaviour of the reduced model. As an effect r_H increases rapidly (in both reactors) but during the most transient stage (from t=1 until 2) the heterotropic yield is also changed. This is a way for the reduced model to maintain the same concentration of Heterotrophs as more X_{COD} is consumed - r_H and Y_H are correlated. As the process settles at the higher hydrolysis rate, r_H reaches an optimum value and Y_H returns to approximately the original value which is realistic since the concentration of active biomass has not changed in any significant manner. Measurements of the *OUR* would reduce the variations of Y_H. The rise of the *OUR* predicted by the reduced model contrary to the IAWPRC model is namely caused by these changes. The effects on the autotrophic parameters caused by the varying hydrolysis rate are quite small as would be expected.

The influent flow rate disturbance is introduced at t=6. If the reduced model was a perfect replica of the process (in this case the IAWPRC model) this perturbance would not require any model parameters to be updated since both model and process would react in exactly the same way. This is almost true for the heterotrophic parameters where only small adjustments (≈2-3%) are required to maintain a good fit of the model. For the autotrophic parameters larger adjustments are needed (≈10%). The reaction rate factor r_A is decreased mainly because the autotrophic growth rate of the IAWPRC model is in its most nonlinear region for the concentrations applied in this example. The sudden increase of the ammoina concentration caused by the increased flow rate leads to an overly large predicted growth rate for the Autotrophs and consequently the value of r_A is reduced. Y_A is increased basicly for the same reason. Note that for the flow rate disturbance, the prediction of the *OUR* is quite accurate when compared to the IAWPRC model (dotted).

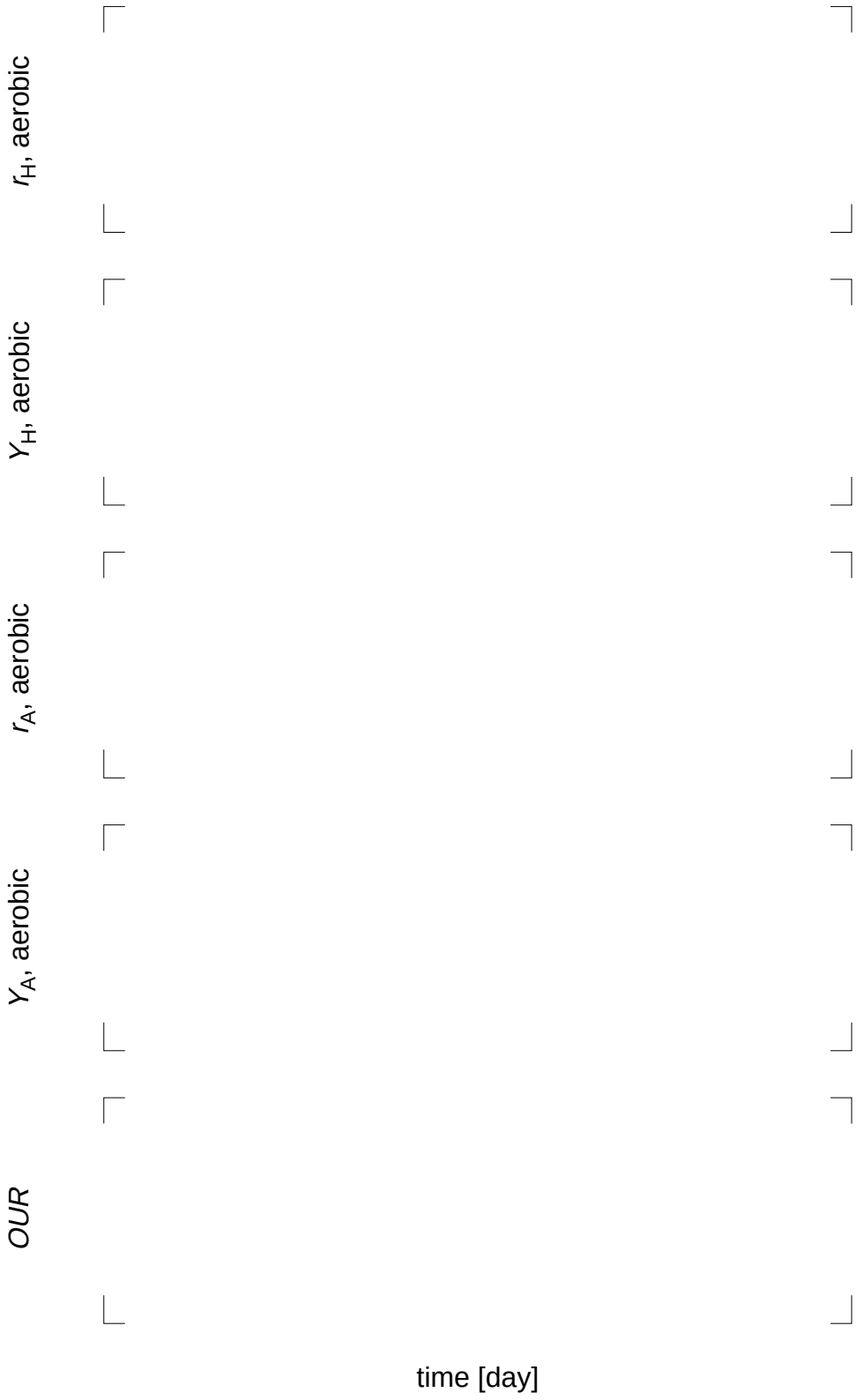


Figure 5.18 On-line parameter estimation based on assumed measurements of X_{COD}, S_{NH}, S_{NO}, $X_{B,H}$, and $X_{B,A}$ during a change of the maximum hydrolysis rate (starting at t=1) and a pulse disturbance of the influent flow rate (starting at t=6).

Note that if the initial and final parameter values of Figure 5.18 are compared, only the heterotrophic reaction rate factor has increased (for both reactors) with any real significance. All other parameters are approximately the same. This is a realistic consequence of the imposed variation of the process; a higher rate of the hydrolysis mechanism produces more readily biodegradable substrate and this substrate is consequently consumed at a higher rate by the active heterotrophic biomass - altogether an increased reaction rate for the Heterotrophs.

Summary discussion

The examples presented in this section illustrate the behaviour of the reduced model during on-line state and parameter identification using an extended Kalman filter. The model was investigated for various assumed measurable quantities, different sets of initial parameter sets, sensitivity to applied noise, etc. At an early stage it was clear that all ten parameters of 'model A' could not be estimated simulataneously under the applied conditions (caused divergence) and therefore a further simplified model with constant decay rate factors was used. A total of six parameters and a number of state variables was updated on-line.

The estimates of the model converge for all tested cases even when data from a stationary process is used. However, in order for all parameters to converge in a global sense some kind of knowledge of the microorganism concentration is required. Otherwise, $X_{B,A}$ and Y_A can not be uniquely determined; instead the ratio of the two variables converges globally. The rate of convergence is also significantly higher when measurements of the organism concentrations are considered available.

Even when a significant amount of noise is added to the measurements and the simulated process, the identified parameters converge in approximately the same way as without noise for all cases. An on-line parameter tracking system applied when the real process is exposed to changing conditions and disturbances of the influent wastewater also show reasonable and accurate results.

6

Conclusions

The activated sludge process is a very complex system and there is a wide range in the degree of complexity of the models which can be used to describe it. Some of the more detailed models are so complex that they are inappropriate for operational use, especially for automatic control applications. However, in many situations the activated sludge process can actually be described satisfactorily by fairly simple models. In order to assure that the model is as simple as possible, each new constant or element of complexity added, should be shown to be essential and its omission should result in failure to describe some important feature of the process.

Unfortunately, determining a suitable model for the intended purpose only solves part of the problem. Due to the biologically adaptive nature of the AS process and the large variation of the character of the influent wastewater over time, any reliable model will have to update its parameters on a regular basis. This imposes further restrictions on the structure of the applied models concerning the concept of system and parameter identifiability.

In this work the problem of over-parametrization in complex AS models has been approached. Existing complex models of the activated sludge system dynamics do not have a unique set of parameters which can explain a certain behaviour. An attempt has been made to derive a reduced order model with less number of states and parameters which is capable of adequately describing the major dynamical behaviour of both the carbonaceous and nitrogeneous activities of the AS process. Still, the basic physical interpretation of the modelled reactions has been retained when possible. Furthermore, the lack of available instrumentation emphasizes the need for a realistic complexity of models for operational purposes.
A thorough investigation of the identifiability of the proposed

reduced order models has been performed using both off-line (simplex) and on-line (extended Kalman filter) algorithms. Results have been presented for a large number of different cases depending on which quantities were assumed measurable. Under certain conditions the simplified models were shown to be globally identifiable, even when the data was corrupted by a significant amount of noise. Simulation comparisons between the IAWPRC model and the reduced models have verified that the main features of the dynamics have been retained.

Correlations between different model parameters under various conditions have been investigated. One such correlation exists between the reaction rate factor and the decay rate factor. The difficulty of estimating these parameters separately in a global sense compared to the combined net reaction rate was shown. Finally, a sensitivity analysis of the reduced model to parameter changes has been performed.

The reduced order model is primarily aimed as an on-line tool for supervision and control as an integral part of a hierarchical control structure. Since its parameters can be gradually updated from on-line measurements, any deviation between the real plant and the model predictions can be used as an early warning system for process diagnosis purposes.

Topics for future work

In the course of the work presented in this thesis, several types of problems and questions have been encountered which deserve more attention in the future. Some of these are summarized below.

- How may the reduced models of this work be further simplified? For identifiability reasons, the simplest model should be based on net reaction rates and the explicit use of active heterotrophic and autotrophic biomass as state variables ought to be eliminated. Such a model would require fewer measurements to calibrate and update. It would also be easier to verify.
- Exactly what measurements, with what accuracy, how often, and performed where, would be the minima required in order

to guarantee global identifiability of the reduced order models? Could easily available measurements like SS (suspended solids), total COD content, etc., improve matters?

- How should the simplified models be applied for control in practice? The possibility to develop control algorithms based on feed forward and adaptive control principles should be investigated once the simple, identifiable models are available. Due to the long time constants of the AS process an optimal control scheme should be based on measurements of the influent wastewater and actions taken before the problems actually occur (feed forward). On the other hand, adaptive control is the proper way of controlling processes that change their behaviour over time, such as the activated sludge system.

- How could the the model verification be performed at a full scale WWT plant during normal operations? The possibilities of temporarily changing flow rates and flow schemes, adding reject water of high concentrations at specific points, use the largest natural variations of the influent wastewater for step feed, etc., in order to achieve more information from the transient behaviour of the process should be investigated. These perturbances should, however, not affect the total plant performance.

- How should a hierarchical control structure be implemented? The different low level control modules need to be syncronized and the partly contradictory contol criteria of the overall control strategy have to formulated to allow optimal performance.

- How may an expertsystem for diagnosis and logical reasoning improve plant performance and how should it be implemented?

7

Bibliography

[1] Andrews, J. F. (1969) Dynamic Model of the Anaerobic Digestion Process, *J. San. Eng. Div.*, ASCE, 95(1):95-116.

[2] Andrews, J. F., G. Olsson, and R. D. Hill (1981) ACSL, An Activated Sludge Process Control Language, Technical Report, Weyerhaeuser Company, Tacoma, Washington.

[3] Argylan, J., A. P. Mathews, R. Nassar, and S. M. Lee (1991) A Non-linear Stochastic Model for Adsorption in Batch Reactors,*Water Sci. Tech.*, 24(6):49-56.

[4] Ayesa, E., J. Florez, J. L. García-Heras, and L. Larrea (1991) State and Coefficients Estimation for the Activated Sludge Process Using a Modified Kalman Filter Algorithm, *Water Sci. Tech.*, 24(6):235-247.

[5] Batchelor, B. (1982) Kinetic Analysis of Alternative Configurations for Single-sludge Nitrification-Denitrification, *J. Wat. Pollut. Control Fed.*, 54:1493-1504.

[6] Beck, M. B. (1986) Identification, Estimation and Control of Biological Wastewater Treatment Processes, *Proc. Inst. Elect. Eng.*, 133(D):254-264.

[7] Beck, M. B. (1989) System Identification and Control, In: *Dynamic Modeling and Expert Systems in Wastewater Engineering* (Eds. G. G. Patry and D. Chapman), 261-323, Lewis Publishers, Inc., Chelsea, Michigan.

[8] Beck, M. B. (1991) Principles of Modelling, *Water Sci. Tech.*, 24(6):1-8.

[9] Benefield, L. and F. Molz (1984) A Model for the Activated

Sludge Process which considers Wastewater Characteristics, Floc Behavior, and Microbial Population, *Biotech. & Bioeng.*, 26:352-361.

[10] Benefield, L. and R. B. Reed (1985) An Activated Sludge Model which considers Toxicant Concentration: Simulation and Sensitivity Analysis, *Appl. Math. Modelling*, 9:454-465.

[11] Boyle, W. C. and P. M. Berthouex (1974) Biological Wastewater Treatment Model Building Fits and Misfits, *Biotech. & Bioeng.*, 16:1139-1159.

[12] Cech J. S., J. Chudoba, and P. Grau (1985) Determination of Kinetic Constants of Activated Sludge Microorganisms, *Water Sci. Tech.*, 17:259-272.

[13] Derco, J., M. Kralik, M. Hutnan, and I. Bodik (1990a) Dynamic Modelling of Activated Sludge Process. (1) Nonlinear Kinetics Models, *Chemical Papers - Chemicke Zvesti*, 44(3):409-420.

[14] Derco, J., M. Kralik, P. Gubova, and I. Bodik (1990b) Dynamic Modelling of Activated Sludge Processes. (2) Linearized Kinetic Models, *Chemical Papers - Chemicke Zvesti*, 44(5):659-668.

[15] Dochain, D. and G. Bastin (1984) Adaptive Identification and Control Algorithms for Nonlinear Bacterial Growth Systems, *Automatica*, 20:621-634.

[16] Dold, P. L., G. A. Ekama, and G. v. R. Marais (1980) A General Model for the Activated Sludge Process, *Prog. Water Technology*, 12:47-77.

[17] Dold, P. L. and G. v. R. Marais (1986) Evaluation of the General Activated Sludge Model Proposed by the IAWPRC Task Group, *Water Sci. Tech.*, 18:63-89.

[18] Dold, P. L., M. C. Wentzel, A. E. Billing, G. A. Ekama and G.v.R. Marais (1991) *Activated Sludge System Simulation Programs*, Creda Press, Cape Town.

[19] Dold, P. L. (1992) Activated Sludge System Model incorporating Biological Nutrient (N & P) Removal, Dept. of Civil Eng. & Eng. Mech., McMaster University, Hamilton, Ontario, Canada.

[20] Dupont, R. and M. Henze (1988) Mathematical Modelling of Nitrogen and Phosphorus Removal, *Trans. Seminar on Nutrient Removal from Municipal Waste Water*, Sep. 4-6, 1989, Tampere, Finland.

[21] Eckenfelder, W. W. (1966) *Industrial Water Pollution Control*, McGraw-Hill, New York.

[22] Ekama, G. A. and G. v. R. Marais (1979) Dynamic Behaviour of the Activated Sludge Process, *J. Water Pollut. Control Fed.*, 51:534-556.

[23] Ekama, G. A., P. L. Dold, and G. v. R. Marais (1986) Procedures for determining Influent COD Fractions and the Maximum Specific Growth Rate of Heterotrophs in Activated Sludge Systems, *Water Sci. Tech.*, 18:91-114.

[24] Fletcher, R. (1987) *Practical Methods of Optimization*, John Wiley & Sons, Chichester, UK.

[25] Fujie, K., T. Tsubone, H. Moriya, and H. Kubota (1988) A Simplified Kinetic Model to Simulate Soluble Organic Substances Removal in an Activated Sludge Aeration Tank, *Water Research*, 22(1):29-36.

[26] Godfrey, K. R. and J. J. DiStefano (1985) Identifiability of Model Parameters, *Proc. IFAC Identification and System Parameter Estimation*, York, UK.

[27] Grady, Jr., C. P. L. and H. C. Lim (1980) *Biological Wastewater Treatment*, Marcel Dekker, New York.

[28] Grady, Jr., C. P. L., J. S. Dang, D. M. Harvey, A. Jobbagy, and X. L. Wang (1989) Determination of Biodegradation Kinetics through Use of Electrolytic Respirometry, *Water Sci. Tech.*, 21:957-968.

[29] Gujer, W. and M. Henze (1991) Activated Sludge Modelling

and Simulation, *Water Sci. Tech.*, 23:1011-1023.

[30] Hendricks, E. (1992) *Identification and Estimation of Nonlinear Systems using Physical Modelling*, Ph.D. Dissertation no. 61, Danish Inst. of Technology, Lyngby, Denmark.

[31] Henze, M., C. P. L. Grady Jr., W. Gujer, G. v. R. Marais and T. Matsuo (1987a) Activated Sludge Model No. 1, IAWPRC Scientific and Technical Report No. 1, IAWPRC, London.

[32] Henze, M., C. P .L. Grady Jr., W. Gujer, G. v. R. Marais and T. Matsuo (1987b) A General Model for Single-sludge Wastewater Treatment Systems, *Water Research*, 21(5): 505-515.

[33] Henze, M. (1988) Constants in the Mathematical Models, IAWPRC Technology Transfer Seminar on *Mathematical Modelling of Biological Wastewater Treatment Processes*, May 4-6, 1988, Rome, Italy.

[34] Henze, M., P. Harremoës, J. l. C. Jansen, and E. Arvin (1992) *Spildevandsrensning, Biologisk og Kemisk* (2nd edition, in Danish), Polyteknisk Forlag, Lyngby, Denmark.

[35] Holmberg, A. (1981) *A Systems Engineering Approach to Biotechnical Processes - Experiences of Modelling, Estimation and Control Methods*, Ph.D. Dissertation, Acta Polytechnica Scandinavica, Helsinki, Finland.

[36] Holmberg, A. and J. Ranta (1982) Procedures for Parameter and State Estimation of Microbial Growth Process Models, *Automatica*, 18:181-193.

[37] Holmberg, A. (1982) On the Practical Identifiability of Microbial Growth Models incorporating Michaelis-Menten Type Nonlinearities, *Math. Bioscience*, 62:23-43.

[38] Holmberg, U. (1991) *Relay Feedback of Simple Systems*, Ph.D. Dissertation, CODEN:LUTFD2/(TFRT-1034), Dept. of Automatic Control, Lund Inst. of Technology, Sweden.

[39] Jeppsson, U. and G. Olsson (1993) Reduced Order Models

for On-line Parameter Identification of the Activated Sludge Process. Paper accepted to the *6th IAWQ Workshop on Instrumentation, Control and Automation of Water and Wastewater Treatment and Transport Systems*, Banff-Hamilton, Ontario, Canada.

[40] Kappeler, J. and W. Gujer (1991) Estimation of Kinetic Parameters of Heterotrophic Biomass under Aerobic Conditions and Characterization of Wastewater for Activated Sludge Modelling, IAWPRC Seminar on *Interactions of Wastewater, Biomass and Reactor Configurations in Biological Treatment Plants*, Aug. 21-23, 1991, Copenhagen, Denmark.

[41] Kuo, B. C. (1991) *Automatic Control Systems*, Prentice Hall, Inc., Englewood Cliffs, NJ.

[42] Larrea, L., J. L. García-Heras, E. Ayesa, and J. Florez (1991) Designing Experiments to Determine the Coefficients of Activated Sludge Models by Identification Algorithms, IAWPRC Seminar on *Interactions of Wastewater, Biomass and Reactor Configurations in Biological Treatment Plants*, Aug. 21-23, 1991, Copenhagen, Denmark.

[43] Lawrence, A. W. and P. L. McCarty (1970) Unified Basis for Biological Treatment Design and Operation, *J. San. Eng. Div.*, ASCE, 96:757-778.

[44] Ljung, L. (1983) *Theory and Practice of Recursive Identification*, Massachusetts Institute of Technology.

[45] Ljung, L. (1987) *System Identification: Theory for the User*, Prentice-Hall, Englewood Cliffs, NJ.

[46] Little, J. N. and L. Shure (1988) *Signal Processing Toolbox*, The MathWorks, Inc., South Natick, MA.

[47] Marais, G. v. R. and G. A. Ekama (1976) The Activated Sludge Process: Part 1 - Steady State Behaviour, *Water SA*, 2:163-200.

[48] MathWorks (1992) *Simulink™, A Program for Simulating*

Dynamic Systems, The MathWorks, Inc., Natick, MA.

[49] McKinney, R. E. (1962) Mathematics of complete Mixing Activated Sludge, *J. San. Eng. Div.*, ASCE, 88(3):87-113.

[50] McKinney, R. E. and R. J. Ooten (1969) Concepts of complete Mixing Activated Sludge, *Trans. 19th San. Eng. Conf.*, University of Kansas, 32-59.

[51] Monod, J. (1942) The Growth of Bacterial Cultures, *Ann. Rev. Microb.*, 3:371-394.

[52] Ogata, K. (1987) *Discrete-time Control Systems*, Prentice Hall, Inc., Englewood Cliffs, NJ.

[53] Olsson, G. and J. F. Andrews (1978) The Dissolved Oxygen Profile – a Valuable Tool for Control of the Activated Sludge Process, *Water Research*, 12:985-1004.

[54] Olsson, G. (1987) Dynamic Modelling, Estimation and Control of Activated Sludge Systems, In: *Systems Analysis in Water Quality Management* (Ed. M. B. Beck), 349-356, Pergamon Press, Oxford.

[55] Olsson, G. (1989) Practical Experiences of Identification and Modeling from Experiments, In: *Dynamic Modeling and Expert Systems in Wastewater Engineering* (Eds. G. G. Patry and D. Chapman), 325-344, Lewis Publishers, Inc., Chelsea, Michigan.

[56] Olsson, G. (1991) Operations and Control in Wastewater Treatment - some Swedish Experiences, *Wat. Sci. Tech.*, 24(6):193-200.

[57] Olsson, G. and G. Piani (1992) Computer Systems for Automation and Control, Prentice Hall, Inc., UK.

[58] Ossenbruggen, P. J., H. Spanjers, H. Aspegren, and A. Klapwijk (1991) Designing Experiments for Model Identification of the Nitrification Process, *Water Sci. Tech.*, 24(6):9-16.

[59] Padukone, N. and G. F. Andrews (1989) A Simple,

Conceptual Mathematical Model for the Activated Sludge Process and its Variants, *Water Research*, 23(12):1535-1543.

[60] Pohjanpalo, H. (1978) System Identifiability based on the Power Series Expansion of the Solution, *Math. Biosci.*, 41:21-33.

[61] Seborg, D. E., T. F. Edgar, and D. A. Mellichamp (1989) *Process Dynamics and Control*, John Wiley & Sons, New York.

[62] Sheffer M. S., M. Hiraoka, and K. Tsumura (1984) Flexible Modelling of the Activated Sludge System - Theoretical and Practical Aspects, *Water Sci. Tech.*, 17, 247-258.

[63] Spanjers, H. and G. Olsson (1992) Modelling of the Dissolved Oxygen Probe Response in the Improvement of the Performance of a Continuous Respiration Meter, *Water Research*, 26(7):945-954.

[64] Spanjers, H., G. Olsson, and A. Klapwijk (1993) Estimation of Short-term Biochemical Oxygen Demand and Respiration Rate in the Aeration Tank, Paper accepted to the *6th IAWQ Workshop on Instrumentation, Control and Automation of Water and Wastewater Treatment and Transport Systems*, Banff-Hamilton, Ontario, Canada.

[65] SSPA Systems (1991) *Simnon™, User's Guide for UNIX Systems, SSPA Systems*, Gothenburg, Sweden.

[66] Stewart, I. (1990) *Does God play Dize - the new Mathematics of Chaos*, Penguin Books Ltd., Harmondsworth, Middlesex, UK.

[67] Söderström, T. and P. Stoica (1989) *System Identification*, Prentice Hall, Inc., Hertfordshire, UK.

[68] Tang, C., E. D. Brill Jr., and J. T. Pfeffer (1987) Comprehensive Model of Activated Sludge Wastewater Treatment System, *J. Env. Eng.*, ASCE, 113(5):952-969.

[69] Thomann, R. V. (1963) Mathematical Model for Dissolved

Oxygen, *J. San. Eng. Div.*, ASCE, 89(5):1-30.

[70] Thomann, R. V. (1982) Verification of Water Quality Models, *J. Env. Eng. Div., Proc. Am. Soc. Civ. Eng.*, 108(EE5):923-940.

[71] Vaccari, D. A. and C. Christodoulatos (1990) Estimation of the Monod Model Coefficients for Dynamic Systems using actual Activated Sludge Plant Data, *Water Sci. Tech.*, 10:449-454.

[72] Van Haandel, A. C., G. A. Ekama, and G. v. R. Marais (1981) The Activated Sludge Process: Part 3 - Single Sludge Denitrification, *Water Research*, 15:1135-1152.

[73] Vanrolleghem, P. A. and W. Verstraete (1993) Simultaneous Biokinetic Characterization of Heterotrophic and Nitrifying Populations of Activated Sludge with an On-line Respirographic Biosensor, Paper accepted to the *6th IAWQ Workshop on Instrumentation, Control and Automation of Water and Wastewater Treatment and Transport Systems*, Banff-Hamilton, Ontario, Canada.

[74] Vialas, C., A. Cheruy, and S. Gentil (1985) An Experimental Approach to Improve the Monod Model Identification, In: *Modelling and Control of Biotechnological Processes, 1st IFAC Symposium*, 175-179.

[75] Wentzel, M. C., L. H. Lotter, R. E. Loewenthal and G. v. R. Marais (1986) Metabolic Behaviour of Acinetobacter Spp. in Enhanced Biological Phosphorus Removal - A Biochemical Model, *Water SA*, 12(4):209-224.

[76] Åström, K. J. and B. Wittenmark (1990) *Computer Controlled Systems, Theory and Design* (2nd edition), Prentice Hall, Inc., Englewood Cliffs, NJ.

Appendix

A

List of Symbols

State variables

unspecified biodegradable single substrate	S
soluble inert organic matter	S_{I}, S_{us}
readily biodegradable substrate	S_{S}, S_{bs}
adsorbed slowly biodegradable substrate	S_{ads}
enmeshed slowly biodegradable substrate	S_{enm}
particulate inert organic matter	X_{I}
particulate products arising from biomass decay	X_{P}
slowly biodegradable substrate	X_{S}
biodegradable substrate	X_{COD}
inert mass	Z_{I}
endogenous mass	Z_{E}
active unspecified biomass (ub)	X
active heterotrophic biomass (hb)	$X_{\mathrm{B,H}}$, Z_{BH}
active autotrophic biomass (ab)	$X_{\mathrm{B,A}}$, Z_{BA}
nitrate and nitrite nitrogen	S_{NO}, N_{o3}
ammonia nitrogen	S_{NH}, N_{a}
soluble biodegradable organic nitrogen	S_{ND}, N_{obs}
particulate biodegradable organic nitrogen	X_{ND}, N_{obp}
oxygen	S_{O}, O
alkalinity	S_{ALK}, Alk

Parameters

maximum specific growth rate for *ub*	μ
decay rate coefficient for *ub*	b
yield factor for *ub*	Y
reaction rate factor for *ub*	r
growth saturation concentration for *ub*	S_{sat}
maximum specific growth rate for *hb*	μ_H
decay rate coefficient for *hb*	b_H
yield factor for *hb*	Y_H
reaction rate factor for *hb*	r_H
maximum specific growth rate for *ab*	μ_A
decay rate coefficient for *ab*	b_A
yield factor for *ab*	Y_A
reaction rate factor for *ab*	r_A
organic substrate half-saturation coefficient	K_S
nitrate half-saturation coefficient	K_{NO}
ammonia half-saturation coefficient	K_{NH}
oxygen half-saturation coefficient	$K_{O,H}$
half-saturation coefficient for hydrolysis of slowly biodegradable substrate	K_X
maximum specific hydrolysis rate	k_h
ammonification rate	k_a
correction factor for anoxic growth of *hb*	η_g
correction factor for anoxic hydrolysis	η_h
fraction of biomass yielding particulate products	f_P
mass N/mass COD in biomass	i_{XB}
mass N/mass COD in products from biomass	i_{XP}

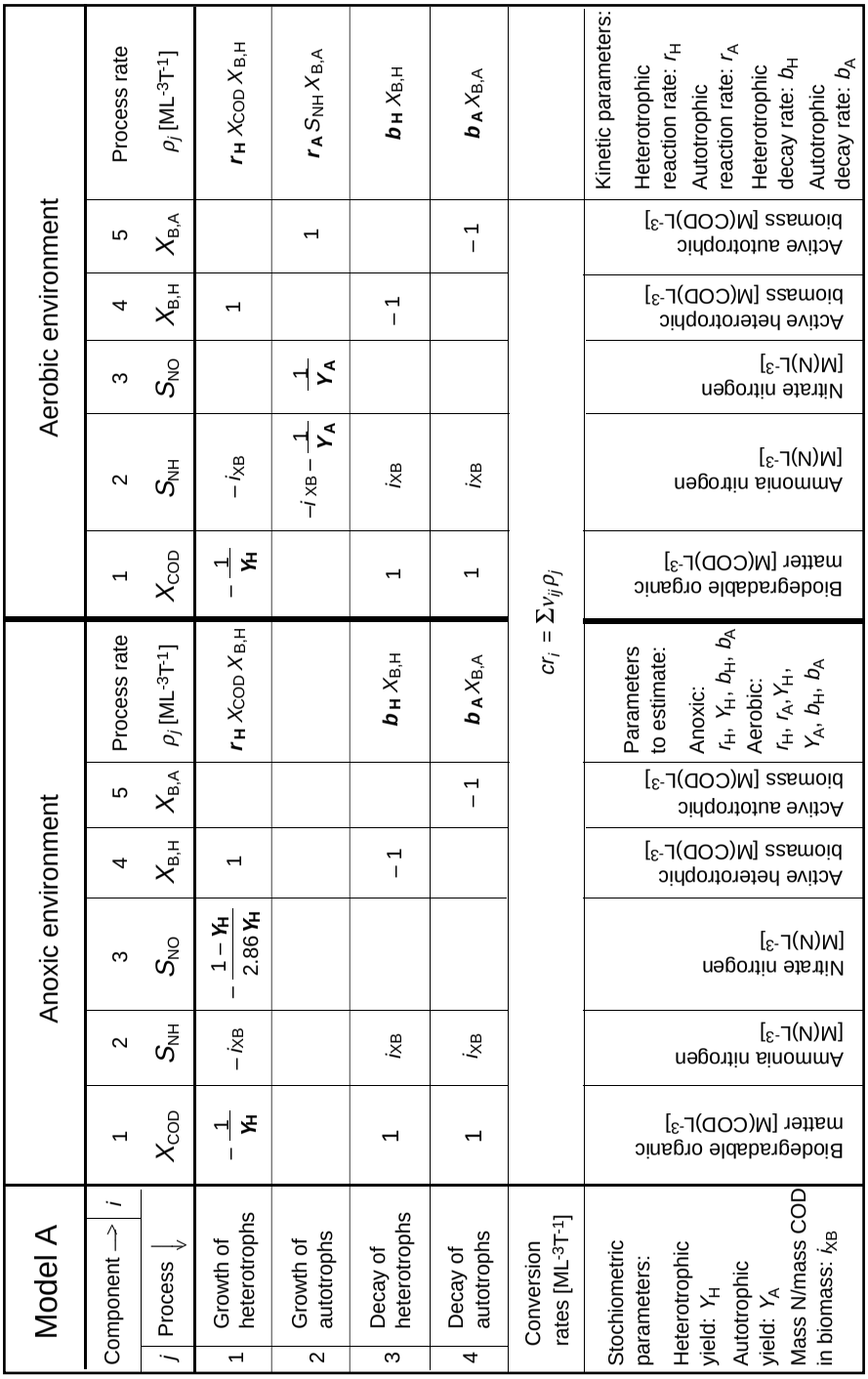
B

The IAWPRC Model

Figure B.1 Matrix formulation of the IAWPRC model [Henze et al., 1987a].

C

The Reduced Order Models



Model A		Anoxic environment						Aerobic environment					
Component → i		1	2	3	4	5		1	2	3	4	5	
j	Process ↓ v	X_{COD}	S_{NH}	S_{NO}	$X_{B,H}$	$X_{B,A}$	Process rate ρ_j [ML^{-3}T^{-1}]	X_{COD}	S_{NH}	S_{NO}	$X_{B,H}$	$X_{B,A}$	Process rate ρ_j [ML^{-3}T^{-1}]
1	Growth of heterotrophs	$-\frac{1}{\boldsymbol{Y_H}}$	$-i_{XB}$	$-\frac{1-\boldsymbol{Y_H}}{2.86\,\boldsymbol{Y_H}}$	1		$\boldsymbol{r_H}\,X_{COD}\,X_{B,H}$	$-\frac{1}{\boldsymbol{Y_H}}$	$-i_{XB}$		1		$\boldsymbol{r_H}\,X_{COD}\,X_{B,H}$
2	Growth of autotrophs								$-i_{XB}-\frac{1}{\boldsymbol{Y_A}}$	$\frac{1}{\boldsymbol{Y_A}}$		1	$\boldsymbol{r_A}\,S_{NH}\,X_{B,A}$
3	Decay of heterotrophs	1	i_{XB}		-1		$\boldsymbol{b_H}\,X_{B,H}$	1	i_{XB}		-1		$\boldsymbol{b_H}\,X_{B,H}$
4	Decay of autotrophs	1	i_{XB}			-1	$\boldsymbol{b_A}\,X_{B,A}$	1	i_{XB}			-1	$\boldsymbol{b_A}\,X_{B,A}$
Conversion rates [ML^{-3}T^{-1}]		$cr_i = \Sigma v_{ij}\rho_j$											
Stochiometric parameters: Heterotrophic yield: Y_H Autotrophic yield: Y_A Mass N/mass COD in biomass: i_{XB}		Biodegradable organic matter [M(COD)L^{-3}]	Ammonia nitrogen [M(N)L^{-3}]	Nitrate nitrogen [M(N)L^{-3}]	Active heterotrophic biomass [M(COD)L^{-3}]	Active autotrophic biomass [M(COD)L^{-3}]	Parameters to estimate: Anoxic: r_H, Y_H, b_H, b_A Aerobic: r_H, r_A, Y_H, Y_A, b_H, b_A	Biodegradable organic matter [M(COD)L^{-3}]	Ammonia nitrogen [M(N)L^{-3}]	Nitrate nitrogen [M(N)L^{-3}]	Active heterotrophic biomass [M(COD)L^{-3}]	Active autotrophic biomass [M(COD)L^{-3}]	Kinetic parameters: Heterotrophic reaction rate: r_H Autotrophic reaction rate: r_A Heterotrophic decay rate: b_H Autotrophic decay rate: b_A

Figure C.1 Matrix formulation of the reduced order model A. The parameters in **bold** are assumed to be variable and should be identified (separately for the anoxic and aerobic zone).

Model B		Anoxic environment						Aerobic environment					
Component →	*i*	1	2	3	4	5	Process rate	1	2	3	4	5	Process rate
j	Process ↓	X_{COD}	S_{NH}	S_{NO}	$X_{B,H}$	$X_{B,A}$	ρ_j [ML^{-3}T^{-1}]	X_{COD}	S_{NH}	S_{NO}	$X_{B,H}$	$X_{B,A}$	ρ_j [ML^{-3}T^{-1}]
1	Growth of heterotrophs	$-\frac{1}{\boldsymbol{Y_H}}$	$-i_{XB}$	$-\frac{1-\boldsymbol{Y_H}}{2.86\,\boldsymbol{Y_H}}$	1		$\boldsymbol{r_H}\,X_{COD}\,X_{B,H}$	$-\frac{1}{\boldsymbol{Y_H}}$	$-i_{XB}$		1		$\boldsymbol{r_H}\,X_{COD}\,X_{B,H}$
2	Growth of autotrophs								$-i_{XB}-\frac{1}{\boldsymbol{Y_A}}$	$\frac{1}{\boldsymbol{Y_A}}$		1	$\boldsymbol{r_A}\,S_{NH}\,X_{B,A}$
3	Decay of heterotrophs	1	i_{XB}		-1		$\boldsymbol{b}\,X_{B,H}$	1	i_{XB}		-1		$\boldsymbol{b}\,X_{B,H}$
4	Decay of autotrophs	1	i_{XB}			-1	$\boldsymbol{b}\,X_{B,A}$	1	i_{XB}			-1	$\boldsymbol{b}\,X_{B,A}$
Conversion rates [ML^{-3}T^{-1}]		$cr_i = \Sigma v_{ij}\rho_j$											
Stochiometric parameters: Heterotrophic yield: Y_H Autotrophic yield: Y_A Mass N/mass COD in biomass: i_{XB}		Biodegradable organic matter [M(COD)L^{-3}]	Ammonia nitrogen [M(N)L^{-3}]	Nitrate nitrogen [M(N)L^{-3}]	Active heterotrophic biomass [M(COD)L^{-3}]	Active autotrophic biomass [M(COD)L^{-3}]	Parameters to estimate: Anoxic: r_H, Y_H Aerobic: r_H, r_A,Y_H, Y_A Common: *b*	Biodegradable organic matter [M(COD)L^{-3}]	Ammonia nitrogen [M(N)L^{-3}]	Nitrate nitrogen [M(N)L^{-3}]	Active heterotrophic biomass [M(COD)L^{-3}]	Active autotrophic biomass [M(COD)L^{-3}]	Kinetic parameters: Heterotrophic reaction rate: r_H Autotrophic reaction rate: r_A Heterotrophic and autotrophic decay rate: *b*

Figure C.2 Matrix formulation of the reduced order model B. The parameters in **bold** are assumed to be variable and should be identified (*b* common for both anoxic and aerobic zone).

D

The Simplex Optimization Algorithm

Many methods for optimization of algebraic functions are based on rough ideas without much theoretical background (*ad hoc* methods). One such possible method implies generating a number of points at random within a certain region and selecting the one which gives the best function value over a large number of trials. Unfortunately, this type of methods suffer from the 'curse of dimensionality' since the amount of effort required to solve actual problems goes up rapidly (typically as 2^n) as the number of degrees of freedom increases.

The most successful of the methods which merely compare function values is the *simplex method*, still widely used [Fletcher, 1987]. A regular simplex is a set of $n+1$ equidistant points in $\mathbb{R}^n$, such as the triangle for $n=2$ and tetrahedron for $n=3$. The current information kept in the method is the coordinates of the $n+1$ points and their corresponding function values.

On the first iteration of the simplex method the vertex at which the function value is largest is determined. The vertex is then reflected in the centroid of the other n vertices, thus forming a new simplex. The function value at this new vertex is evaluated and the process repeated. On iterations after the first it might appear that the newest vertex still has the largest function value in the new simplex, and to reflect this vertex would cause oscillation. Hence the largest function value other than that at the newest vertex is subsequently used to decide which vertex to reflect. Ultimately this iteration will fail to make further process, so an additional rule has to be introduced. When a certain vertex i has been in the current simplex for more than a fixed number M iterations, then the simplex should be contracted by replacing the other vertices by new ones half way along the edge to the vertex i. The value of M is normally determined by the dimension of the problem.

The typical progress of the iteration is illustrated in Figure D.1

using a two dimensional example. Vertices 1, 2, and 3 form the initial simplex and increasing numbers indicate the new vertices added at each iteration. Note that vertex 7 has the largest function value for the simplex (4,6,7) but is not reflected immediately since it is the newest vertex in that simplex. When simplex (6,9,10) is reached, vertex 6 has been in the current simplex for four iterations and if M is assumed to equal 3.5, the simplex is contracted at this stage to the new simplex (6,11,12) and the iteration continues from this simplex.

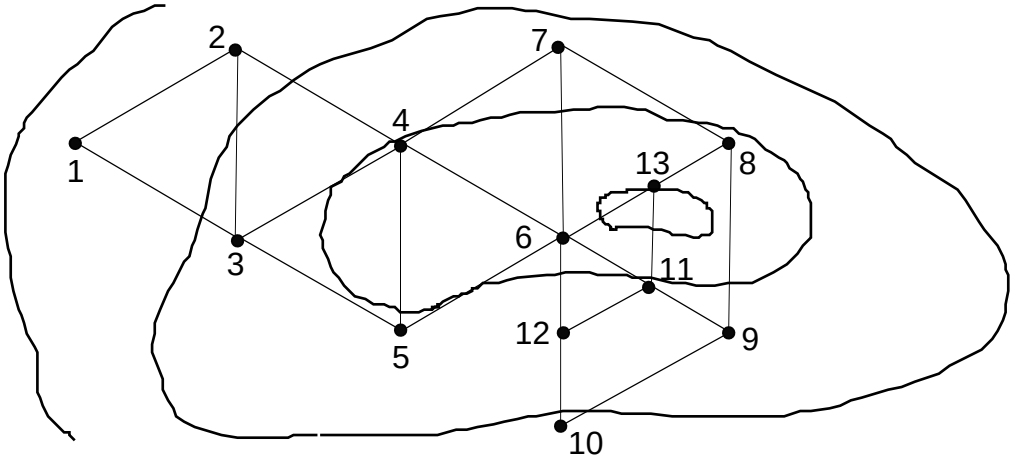


Figure D.1 The simplex method in two variables.

The Nelder-Mead algorithm applied in this work is a slightly modified simplex method which allows irregular simplexes, and distorsions of the simplex are performed automatically in an attempt to take into account the local geometry of the function.

Due to the problems with the computational effort there is a practical limit to the size of systems which the method can be applied to. The convergence rate is slow but the algorithm is very robust and quite insensitive to noise. Often the method can be used in combination with more sophisticated ones. The simplex method is then applied in an early stage of the optimization in order to get the convergence going in the right direction and thereby producing suitable initial values for methods which converge faster. Such algorithms are usually less robust and likely to diverge if the initial estimates are far from the true ones (for example the Gauss-Newton algorithm in the example in section 4.1).

E

The Extended Kalman Filter

The technique of Kalman filters is a general filtering technique which can be applied to such problems as optimal estimation, prediction, noise filtering, and stochastic control. Adaptive gain tuning capability is the characteristic of the Kalman filter. The method can also be applied to both stationary and non-stationary processes.

The following time discrete linear system is assumed:

$$\begin{cases} \mathbf{x}(t_{k+1}) = \Phi\, \mathbf{x}(t_k) + \Gamma\, \mathbf{u}(t_k) + \nu(t_k) \\ \mathbf{y}(t_k) = \mathbf{C}\, \mathbf{x}(t_k) + \varepsilon(t_k) \end{cases} \tag{E.1}$$

where ν and ε are Gaussian white noise processes with zero mean and the covariance matrices are given as:

$$\begin{cases} \mathrm{E}\left[\nu(t_k)\, \nu^{\mathrm{T}}(t_k)\right] = \mathbf{R}_1 \\ \mathrm{E}\left[\nu(t_k)\, \varepsilon^{\mathrm{T}}(t_k)\right] = \mathbf{R}_{12} \\ \mathrm{E}\left[\varepsilon(t_k)\, \varepsilon^{\mathrm{T}}(t_k)\right] = \mathbf{R}_2 \end{cases} \tag{E.2}$$

Let the estimator have the form:

$$\hat{\mathbf{x}}(t_{k+1} \mid t_k) = \Phi\, \hat{\mathbf{x}}(t_k \mid t_{k-1}) + \Gamma\, \mathbf{u}(t_k) + \mathbf{K}(t_k)\left[\mathbf{y}(t_k) - \mathbf{C}\, \hat{\mathbf{x}}(t_k \mid t_{k-1})\right] \tag{E.3}$$

The reconstruction error $\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}$ is governed by:

$$\begin{aligned} \tilde{\mathbf{x}}(t_{k+1}) &= \Phi\, \tilde{\mathbf{x}}(t_k) + \nu(t_k) - \mathbf{K}(t_k)\left[\mathbf{y}(t_k) - \mathbf{C}\, \hat{\mathbf{x}}(t_k \mid t_{k-1})\right] \\ &= \left(\Phi - \mathbf{K}(t_k)\, \mathbf{C}\right) \tilde{\mathbf{x}}(t_k) + \nu(t_k) - \mathbf{K}(t_k)\, \varepsilon(t_k) \end{aligned} \tag{E.4}$$

The property of the noise is taken into account and the criterion is to minimize the variance of the estimation error, $\mathbf{P}(t_k)$, by determining the best gain matrix, $\mathbf{K}(t_k)$. $\mathbf{P}(t_k)$ is defined as:

$$\mathbf{P}(t_k) = \mathrm{E}\left[\left(\tilde{\mathbf{x}}(t_k) - \mathrm{E}\left[\tilde{\mathbf{x}}(t_k)\right]\right)\left(\tilde{\mathbf{x}}(t_k) - \mathrm{E}\left[\tilde{\mathbf{x}}(t_k)\right]\right)^{\mathrm{T}}\right] \qquad (E.5)$$

The mean value of **x** is obtained from (E.4) as:

$$\mathrm{E}\left[\tilde{\mathbf{x}}(t_{k+1})\right] = \left(\Phi - \mathbf{K}(t_k)\,\mathbf{C}\right)\mathrm{E}\left[\tilde{\mathbf{x}}(t_k)\right] \qquad (E.6)$$

If $\mathrm{E}[\mathbf{x}(0)]=m_0$ then the mean value of the reconstruction error is zero for times $t_k \geq 0$ independent of **K** if $\mathrm{E}[\mathbf{x}(0)]=m_0$. This is assumed to be true and equation (E.4) gives:

$$\begin{aligned}\mathbf{P}(t_{k+1}) &= \mathrm{E}\left[\tilde{\mathbf{x}}(t_{k+1})\,\tilde{\mathbf{x}}^{\mathrm{T}}(t_{k+1})\right] \\ &= \left(\Phi - \mathbf{K}(t_k)\,\mathbf{C}\right)\mathbf{P}(t_k)\left(\Phi - \mathbf{K}(t_k)\,\mathbf{C}\right)^{\mathrm{T}} + \mathbf{R}_1 + \mathbf{K}(t_k)\,\mathbf{R}_2\,\mathbf{K}^{\mathrm{T}}(t_k) - 2\,\mathbf{K}(t_k)\,\mathbf{R}_{12}\end{aligned} \qquad (E.7)$$

The criterion is to minimize the scalar $\alpha^{\mathrm{T}}\,\mathbf{P}(t_{k+1})\,\alpha$ where α is an arbitrary vector, by choosing the best possible $\mathbf{K}(t_k)$. If the criterion is developed using (E.7), two terms occur. The first term is independent of $\mathbf{K}(t_k)$ whereas the second term is determined by $\mathbf{K}(t_k)$. If $\mathbf{K}(t_k)$ is chosen such that the second part is zero, a minimum is obtained. The following two equations result:

$$\mathbf{K}(t_k) = \left(\Phi\,\mathbf{P}(t_k)\,\mathbf{C}^{\mathrm{T}} + \mathbf{R}_{12}\right)\left(\mathbf{C}\,\mathbf{P}(t_k)\,\mathbf{C}^{\mathrm{T}} + \mathbf{R}_2\right)^{-1} \qquad (E.8)$$

$$\mathbf{P}(t_{k+1}) = \Phi\,\mathbf{P}(t_k)\,\Phi^{\mathrm{T}} + \mathbf{R}_1 - \mathbf{K}(t_k)\left(\mathbf{C}\,\mathbf{P}(t_k)\,\mathbf{C}^{\mathrm{T}} + \mathbf{R}_2\right)\mathbf{K}^{\mathrm{T}}(t_k) \qquad (E.9)$$

The reconstruction defined by (E.3), (E.8), and (E.9) is called the *Kalman filter*. The main difficulty is usually to determine the proper covariance matrices of (E.2) and to initially select a suitable variance matrix $\mathbf{P}(0)$. Note that $\mathbf{P}(t_k)$ does not depend on the observations. Thus, the gain can be precomputed in forward time and stored in the computer.

Extended Kalman filters (EKF) are a logical generalization of linear Kalman filters for the case where the system dynamics vary with operating and control points in nonlinear systems. The first step in the generalization is to exchange the linear process model (E.1) for a nonlinear one:

$$\begin{cases} \mathbf{x}(t_{k+1}) = \mathbf{f}\left[\mathbf{x}(t_k), \mathbf{u}(t_k)\right] + \nu(t_k) \\ \mathbf{y}(t_k) = \mathbf{h}\left[\mathbf{x}(t_k)\right] + \varepsilon(t_k) \end{cases} \quad \text{(E.10)}$$

where **f** and **h** represent general nonlinear vector functions.

The second step is to use a linearization of the plant dynamics in order to minimize the effect of process and measurement noise. This linearization is performed around the current state estimates, $\mathbf{x}(t_k)$, on-line. The main element of an extended Kalman filter are thus a description of the process dynamics (and a linearized version of it) and a noise model.

There are three principal types of extended Kalman filters; discrete EKF, continuous EKF, and continuous-discrete EKF. The continuous-discrete EKF uses a continuous time update of the nonlinear observer while it employs a discrete measurement update. Such a filter is often a well suited approach because the model can be kept in the traditional continuous form while the measurements are most conveniently digitized using a zero order hold network. A continuous-discrete EKF was applied for this work.

As was discussed in section 4.3, the calculations are often divided in a *prediction* and a *correction* phase. If $\mathbf{R}_{12}$ for simplicity is assumed to equal zero, the EKF can be described in a straightforward way. The predictor phase includes the following calculations:

$$\hat{\mathbf{x}}(t_{k+1} \mid t_k) = \mathbf{f}\left[\hat{\mathbf{x}}(t_k \mid t_k), \mathbf{u}(t_k)\right] \quad \text{(E.11)}$$

$$\mathbf{P}(t_{k+1} \mid t_k) = \mathbf{F}(t_k)\,\mathbf{P}(t_k \mid t_k)\,\mathbf{F}^{\mathrm{T}}(t_k) + \mathbf{R}_1 \quad \text{(E.12)}$$

and the corrector phase includes:

$$\hat{\mathbf{x}}(t_{k+1} \mid t_{k+1}) = \hat{\mathbf{x}}(t_{k+1} \mid t_k) + \mathbf{K}(t_{k+1})\left[\mathbf{y}(t_{k+1}) - \mathbf{h}\left[\hat{\mathbf{x}}(t_{k+1} \mid t_k)\right]\right] \quad \text{(E.13)}$$

$$\mathbf{P}(t_{k+1} \mid t_{k+1}) = \mathbf{P}(t_{k+1} \mid t_k) - \mathbf{K}(t_{k+1})\,\mathbf{H}(t_{k+1})\,\mathbf{P}(t_{k+1} \mid t_k) \quad \text{(E.14)}$$

$$\mathbf{K}(t_{k+1}) = \mathbf{P}(t_{k+1} \mid t_k)\,\mathbf{H}^T(t_{k+1})\left[\mathbf{H}(t_{k+1})\,\mathbf{P}(t_{k+1} \mid t_k)\,\mathbf{H}^T(t_{k+1}) + \mathbf{R}_2\right]^{-1} \quad \text{(E.15)}$$

where $\mathbf{F}(t_k)$ and $\mathbf{H}(t_k)$ correspond to the Jacobian matrices of **f** [.] and **h**[.] respectively. The Jacobians are defined as:

$$\mathbf{F}(t_k) = \left[\frac{\delta\,\mathbf{f}(\mathbf{x}(t_k), \mathbf{u}(t_k))}{\delta\,\mathbf{x}(t_k)}\right]_{\mathbf{x}(t_k) = \hat{\mathbf{x}}(t_k)} \quad \text{(E.16)}$$

$$\mathbf{H}(t_k) = \left[\frac{\delta\,\mathbf{h}(\mathbf{x}(t_k))}{\delta\,\mathbf{x}(t_k)}\right]_{\mathbf{x}(t_k) = \hat{\mathbf{x}}(t_k)} \quad \text{(E.17)}$$

The significant real time computational burden imposed by the use of extended Kalman filters have motivated the search for more simple estimators which can retain some of the robustness characteristics of the full EKF. The constant gain EKF is one such simplification. It is quite easily designed by setting $\mathbf{P}(t_{k+1})$ and $\mathbf{P}(t_k)$ equal and solving the resulting time independent equation for the variance of the estimation error. By using the resulting **P** to find **K**, a *constant* gain matrix is achieved for a selected operating point $\mathbf{x}_0$ of the system. Such a filter maintains its robust behaviour even when exposed to significantly varying signals [Hendricks, 1992]. For practical reasons this approach was applied for the on-line estimations performed in section 5.3.

The methods of this section not only hold for state estimation but also for simultaneous state and parameter estimation. The expressions are still valid although **x** becomes a generalized state vector which includes both the unknown state variables and the uncertain model parameters. The difficulty of determining the proper covariance matrices (E.2) is, however, more emphasized.

A more detailed and theoretical derivation of the Kalman filter and its variants is for example given in [Ogata, 1987 and Ljung, 1983].

F

Simulation Environment

When working with modelling, identification and verification it is important to have access to a good simulation environment - both hardware and software. All the computations included in this work have been performed on a Sun™ Sparcstation IPC using the simulation programs Simnon™ and Simulink™.

Simnon is designed for solving ordinary differential and difference equations and for simulating dynamical systems. Numerical integration routines are used to simulate differential equations and difference equations are solved by iteration. No symbolic analysis of the systems are possible. The systems may be described as an interconnection of subsystems (promoting a hierarchical system description) which may be both in continuous and discrete time. The user interacts with the program by typing commands (a graphical interface is available as an add-on product). Parameters, initial conditions, and system descriptions can be modifyed interactively and the results are graphically or numerically displayed on the screen. A built-in macro facility allows the user to create his own set of commands. As an example, the straightforward text file for simulating the simple model (4.1) is illustrated in Figure F.1.

Simulink is an interactive system for simulating dynamical systems. It is a graphical, mouse-driven program that allows the user to model a system by drawing a block diagram on the screen and manipulating it dynamically. It can handle linear, nonlinear, continuous-time, discrete-time, multivariable, and multirate systems. A large number of predefined building blocks is included in the program and it is easy for the user to extend this library with blocks of his own. Hierarchical models are recommended since blocks may include other blocks and allows for graphical 'information zooming'. Results are numerically and graphically available in numerous ways. The comprehensible block diagram for describing the small model (4.1) is shown in Figure F.2.

```
CONTINUOUS SYSTEM  model4_1
"Model of the Monod growth equation in a single-
"substrate (S)/single-organism (X) batch reactor
"with no other growth limitations.

STATE  S X
DER    dS dX

TIME   t

"Values for the model parameters
mumax :  6         "maximum specific growth rate
Ks    :  10        "half saturation coefficient
b     : 0.48       "decay rate factor
Y     : 0.66       "yield factor

"Initial values for the state variables
S : 100            "substrate concentration
X :  2             "organism concentration

"Dynamic equations
dS = -mu*X/Y
dX = (mu - b)*X
mu = mumax*S/(Ks + S)  "Monod growth rate

END
```

Figure F.1 Simnon text file describing the simple model (4.1).

The major strongpoint of Simulink is the fact that it is an integral part (toolbox) of the complete Matlab™ (mat rix laboratory) computing environment. Matlab is an interactive system whose basic data element is a matrix that does not require dimensioning. It includes a huge library of predefined functions and a simple way for the user to define functions of his own expressed as they are written mathematically - without traditional programming.

A whole family of application-specific toolboxes that extend the Matlab environment in order to solve particular classes of problems is also available. These toolboxes include signal processing, control system design, system identification, optimization, neural networks, etc. Altogether this means that the Simulink user not only has the possibility to perform simulations but an enormous capability to manipulate and further investigate the results. All this power is available at the user's fingertips in one complete environment.

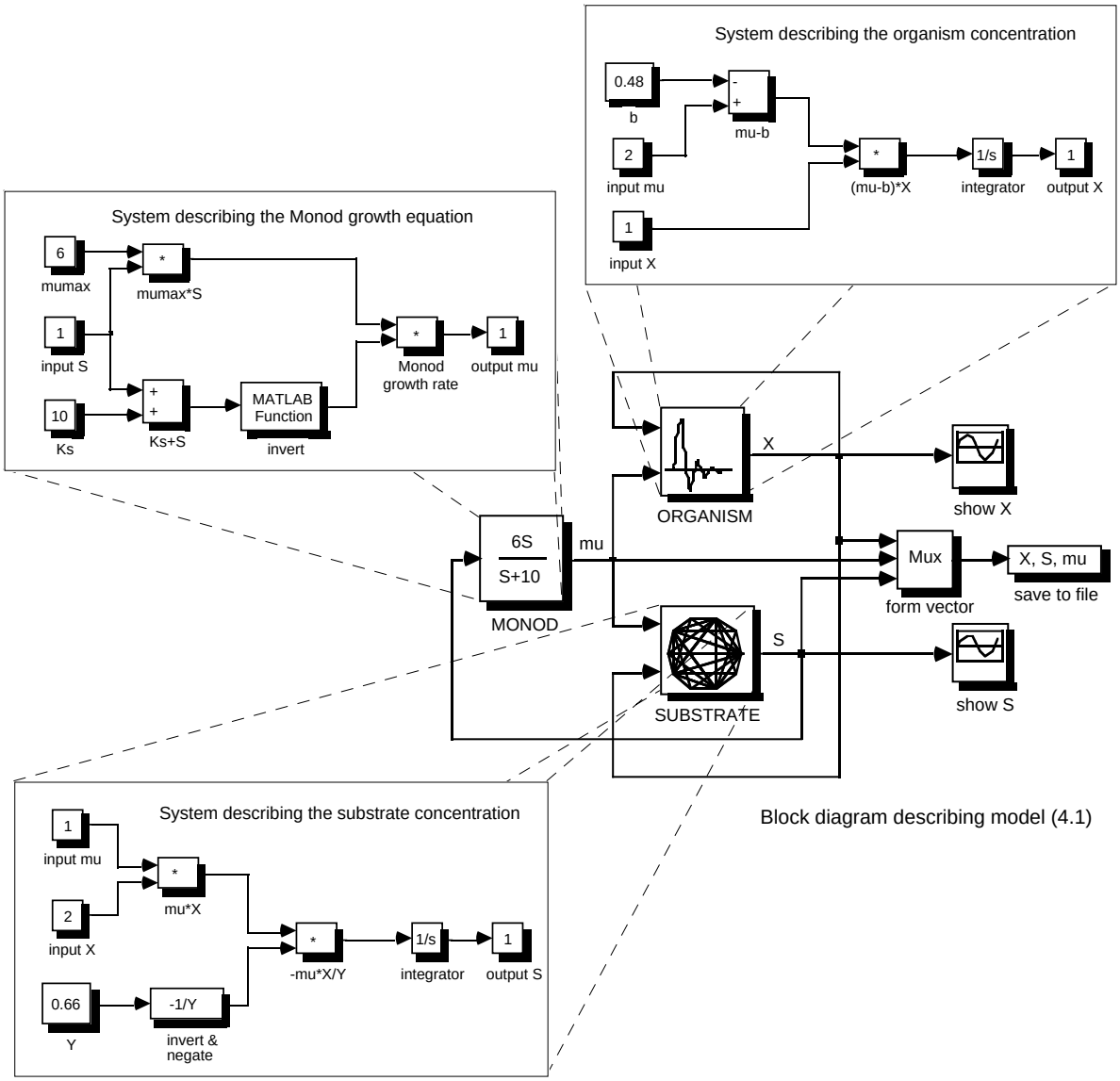


Figure F.2 Simulink block diagram of the simple model (4.1).

The compatibility between Matlab and Simnon is also quite good. It is easy so exchange data files since both systems accept simple ASCII files in tabular form. Linear, time-invariant systems can be directly transferred from Matlab to Simnon by a special translation script. Therefore the user can combine the two programs and take advantage of the respective strongpoints and use both of them as a complete model building, simulation, data analysis, and data manipulation software environment.