

Modeling, estimation, and control of biological wastewater treatment plants

Qian Chai
Faculty of Technology
Telemark University College
Porsgrunn, Norway

April 5, 2008

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree of *philosophiae doctor* (PhD) at the Norwegian University of Science and Technology (NTNU) and Telemark University College (TUC).

First of all, I am indebted to my main supervisor professor Bernt Lie for offering me the opportunity to study for a PhD degree, for the excellent guidance during this work, for a good number of valuable discussions, for reading and correcting diverse reports, articles, and the drafts of this thesis, and for help in improving my English. Your broad knowledge and excellent capability of disseminating knowledge within modeling and control area have been very useful for me. I am also thankful to you for your friendship and for so many enjoyable trips, café-visits, and parties.

My sincere gratitude also goes to my co-supervisor professor Rune Bakke for always being helpful and interested in my work, for the valuable discussions, and for the comments and tips on practical issues of wastewater treatment. I am thankful for your contribution to organize and guide the nitrogen removal project at TUC to put theories into practice.

I would like to thank my other co-supervisors professor Tor Lønnestad and professor Steinar Sælid for the participation and support in this work. I really appreciate that you took the time to participate in the meetings and to give useful comments and suggestions to this work.

I want to thank professor Peter Fritzson for being my kind host during my short visit to Linköping University, Sweden, in June 2005. I am thankful to Dr. Peter Bunus for many valuable discussions and for tremendous help in the Modelica project. I am also grateful to the PELAB (Programming Environments Laboratory) staff at Linköping University for aiding me with administrative issues and for enjoyable coffee breaks.

During the case studies of this work, I have obtained a good deal of help and support of many people. I would like to thank Sverre Amrani from BioTek AS, Norway, for providing measurement data of the Duvbacken plant, for sharing process knowledge, and for collaboration on two papers. Thanks to Cristina Sarmiento Ferrero for her contribution through her M.Sc. thesis. I have much to thank the group members of the nitrogen removal project at TUC. Thanks to lecturer Morten Pedersen at TUC for help with building the control box, visiting researchers Dr. Zhiqiang Chen and professor Tao He from China for their assistance on process design issues, PhD fellow Deshai Botheju at TUC for his support on process construction, master students Yuan Song, Shadi Lotfi, and Lei Zhao for their contribution on laboratory experiments and LabVIEW program.

Research is always depending on a well functioning working environment, fun during coffee breaks as well as serious discussions. In this respect I would like to thank my colleagues,

with whom I have shared many moments. Thanks to assistant professor Finn Haugen for the scientific help, for enjoyable conversations, and for invitations to the NI-days seminars. Thanks also to Beathe Furenes and Juan Videla for sharing an office with me, for so many fruitful discussions, and for a lot of support during good and hard times, and I really appreciate that Juan offered me to quote from his unpublished report on the Dymola-Matlab interface. Thanks to Marta Dueñas Díez for the interesting discussion on parameter estimation and thank you for the times you have invited me to your parties. Thank you all the nice people contributing to have such a positive working environment and so enjoyable lunch breaks, especially to Nils-Olav Skeie, Kjell Joar Alme, Inger Hedvik Matveyev, Urmila Datta, and Xiaoli Liang.

Finally, I would like to thank my dear parents, Jie and Jianxiang, for a lifetime of unlimited love and support. Although we live in two countries with the distance of some thousands of km, your support and encouragement has always inspired me to overcome the difficulties in my PhD work. Very special warm thanks to my husband Jon Are for his patience and loving support during all these years. Thanks also to my dear parents-in-law, Astri and Alf, and my little sister, Lu, for their love and support.

Stavanger, February 2008

Qian Chai

Summary

The activated sludge process (ASP), as the most generally applied biological wastewater purification technique, attracts a great deal of attention from the research community in a highly populated, industrialized world. The main objective of this thesis is to investigate methods for efficient development of activated sludge plant models and for possible improvements in process operation and performance through applications of advanced modeling and control techniques. Different topics in advanced control such as parameter estimation, state estimation, and model predictive control, are addressed, and the methods investigated in this thesis are illustrated with two activated sludge processes, a laboratory-scale nitrogen (N) removal plant and a full-scale phosphorus (P) removal plant: the Duvbacken wastewater treatment plant (WWTP).

As model representation is important in control oriented modeling, the object-oriented modeling approach providing open component-based architectures and reuse of components, is suggested in this thesis to achieve a more flexible model representation. The models of the N-removal process and the Duvbacken WWTP are developed based on the Activated Sludge Models (ASMs) which are generally accepted as the standard models of activated sludge processes (ASPs). The dynamic models are implemented in the object-oriented modeling language Modelica using the Dymola simulation environment, based on a modification of the Modelica library `WasteWater`. In order to build a good software development environment, special consideration is given to two Dymola interfaces: the Dymola-Matlab interface provided by Dymola and the Dymola-Python interface developed by the author. The application of the modeling approach in this thesis shows that the combination of Modelica/Dymola for efficient object-oriented modeling and Matlab for further analysis is a powerful tool for control system design.

The performances of some widely used state estimation techniques, such as the standard Kalman filter (KF), the extended Kalman filter (EKF), the unscented Kalman filter (UKF), the square-root unscented Kalman filter (SRUKF), and the moving horizon estimation (MHE), are evaluated with application to the N-removal process. The comparison of these filters for the N-removal process shows that the SRUKF improves numerical properties of the UKF and provides the best estimation performance, and the discrete-time EKF gives the fastest performance with a satisfactory estimation result. In addition, since a good characterization of the main compositions in the influent is very important to the performance of the ASM models, a method for simultaneously estimating both model state and influent composition (model disturbance) is developed using the discrete-time EKF and illustrated with the Duvbacken WWTP. The filter has a good convergence property for the observable process based on responses from the simulation model. However, when the filter is evalu-

ated based on the experimental data, the performance of the filter is relatively poor due to unobservability of the system. Additional measurements are therefore needed to render the Duvbacken system locally observable.

Parameter estimation of activated sludge plant models is a challenging task, since the models are in general large-scale, nonlinear, and overparameterized. A procedure for parameter identifiability analysis is discussed and tested with the Duvbacken model. This systematic method proves to be very useful, since it reduces a large number of parameters to a small subset of parameters that can be identified from the available experimental data. The fitting of the model to real data from the industrial plant is not great but satisfactory, especially for those outputs which are important to the plant operation (e.g. concentration of phosphate S_{PO_4} and concentration of total phosphorus TP).

Advanced control strategies for the ASP aeration system are suggested with the focus on the MPC algorithm, since MPC provides an optimization-based procedure with an elegant way of handling constraints and with capabilities of removing process disturbances. Both linear and nonlinear MPC techniques are evaluated with the N-removal process, and the simulation results show that the performances of the MPC controllers are satisfactory: the controllers guarantee that the strict effluent requirements are fulfilled. The nonlinear MPC strategy is also compared with some conventional rule-based controllers, and the MPC controller shows the most reduction of the energy consumption and offers the easiest way to handle the inequality constraints.

Finally, recommendations for future research directions are suggested to an N-removal project group at Telemark University College, Norway.

Contents

Preface	iii
Summary	v
Nomenclature	xi
I Overview	1
1 Introduction	3
1.1 Background	3
1.2 Previous work	4
1.3 Scope and outline of thesis	5
1.4 Main contributions	7
2 ASPs for Wastewater Treatment	9
2.1 Introduction	9
2.2 Nitrogen removal process	11
2.2.1 Biological nitrogen removal	11
2.2.2 The laboratory-scale nitrogen removal plant	12
2.3 Phosphorus removal process	15
2.3.1 Biological phosphorus removal	15
2.3.2 The Duvbacken WWTP	15
2.4 Object-oriented modeling	16
2.4.1 Modelica and Dymola	17
2.4.2 Model implementation	19
2.4.3 Dymola interfaces	19
2.A <code>dymosimPy</code> — A Python module for Dymola-Python interface	22
2.A.1 Python vs. Matlab	23
2.A.2 Python fundamentals	24
2.A.3 The <code>dymosimPy</code> module	25
2.A.4 Applications of the <code>dymosimPy</code> module	28
2.A.5 Conclusions	32

3	State Estimation	33
3.1	Introduction	33
3.2	State estimation techniques	35
3.2.1	Observability and detectability	35
3.2.2	The standard Kalman filter (KF)	37
3.2.3	The extended Kalman filter (EKF)	37
3.2.4	The unscented Kalman filter (UKF)	38
3.2.5	The square-root unscented Kalman filter (SRUKF)	39
3.2.6	Moving horizon estimator (MHE)	41
3.3	State estimation of the N-removal process	45
3.3.1	Performance indices	45
3.3.2	Results and discussions	46
3.A	Optimization-based estimation formulated as QP problem	51
3.A.1	The optimization-based estimation problem	51
3.A.2	Standard QP problem	51
4	Parameter Estimation	53
4.1	Introduction	53
4.2	Parameter identifiability analysis	54
4.2.1	Theoretical background	54
4.2.2	Procedure for selecting identifiable parameters	56
4.3	Estimation of kinetic parameters in the Duvbacken model	59
4.3.1	Parameter identifiability and estimation	59
4.3.2	Model validation	60
4.3.3	Confidence interval of parameter estimates	60
4.4	Disturbance estimation using the Kalman filter	63
4.4.1	State and disturbance estimation	63
4.4.2	Observability analysis	64
4.4.3	Application to the Duvbacken plant	67
4.A	Structural observability analysis of ASM2d model	75
5	Control Strategies	77
5.1	Introduction	77
5.2	Rule-based control strategies	78
5.3	Model predictive control (MPC)	80
5.3.1	Linear MPC for continuously aerated ASP	81
5.3.2	Nonlinear MPC for intermittently aerated ASP	82
6	Laboratory Case Study	85
6.1	Experimental set-up	85
6.2	Recommendations to the N-removal project	88
6.3	Implementation architecture — future perspective	89

7	Conclusions and Future Directions	93
7.1	Conclusions	93
7.2	Future directions	95
	Bibliography	106
II	Published and Submitted Papers	107
	Paper A: Systematic Analysis of Parameter Identifiability for Improved Fitting of a Biological Wastewater Model to Experimental Data	109
	Paper B: Parameter Identifiability Analysis and Model Fitting of a Biological Wastewater Model	131
	Paper C: Comparison of state estimation techniques, applied to a biological wastewater treatment process	139
	Paper D: Object-oriented Modeling and Optimal Control of a Biological Wastewater Treatment Process	147
	Paper E: Optimal control of a biological wastewater treatment process with intermittent aeration	155
	Paper F: Predictive Control of an Intermittently Aerated Activated Sludge Process	163

Nomenclature

List of symbols

a	Duration of the aeration cycle
B	Input matrix of a linear system
C	Disturbance matrix of a linear system
d	Aeration time
d_k	Disturbance vector at time step k
D_k	Noise covariance matrix of the disturbance model at time step k
e_t	Observation error at time t , $e_t \triangleq y_t - y_t^m$
F	Process matrix of a linear system
$f(\cdot)$	State transition function (possibly nonlinear)
$f_d(\cdot)$	State transition function of the disturbance model
f_k	Aeration fraction at the k th aeration cycle
G	Process noise matrix of a linear system
H	Output matrix of a linear system
H	Hessian of the objective function for parameter estimation
$h(\cdot)$	Measurement function (possibly nonlinear)
$h_d(\cdot)$	Measurement function of the disturbance model
J	Objective function
K_k	Kalman filter gain
K_{La}	Oxygen mass transfer coefficient
L	Measurement noise matrix of a linear system
M_P	MHE weighting matrix for initial state estimate
M_Q	MHE weighting matrix for process noise
M_R	MHE weighting matrix for measurement noise
N	Prediction horizon
$\mathcal{N}(\cdot)$	Normal probability distribution
N_c	The number of aeration cycles
n_d	Dimension of the disturbance vector d_k
n_u	Dimension of the input vector u_k
n_v	Dimension of the measurement noise vector v_k
n_w	Dimension of the process noise vector w_k

n_x	Dimension of the state vector x_k
n_y	Dimension of the output vector y_k
n_θ	Dimension of the model parameter vector θ
$P_{k k}$	<i>A posteriori</i> state covariance
$\tilde{P}_{k k}$	<i>A posteriori</i> augmented state covariance
$P_{k k-1}$	<i>A priori</i> state covariance
P_{xy}	Cross covariance between $\hat{x}_{k k-1}$ and $\hat{y}_{k k-1}$
P_y	Innovation covariance
Q	Weighting matrix in linear MPC
Q	Process noise covariance matrix of the continuous-time process
Q_{eff}	Effluent flowrate
Q_{in}	Influent flowrate
Q_k	Process noise covariance matrix at time step k
$Q_k^{\tilde{w}}$	Augmented process noise covariance matrix at time step k , $Q_k^{\tilde{w}} = \text{diag}(Q_k, D_k)$
Q_{rs}	Recycled sludge flowrate
Q_w	Wasted sludge flowrate
r	Vector of reaction rates
R	Weighting matrix in linear MPC
r_k	White noise in the disturbance model at time step k
R_k	Measurement noise covariance matrix at time step k
S	Weighting matrix in linear MPC
S	Sensitivity matrix
S_A	Fermentation product (acetate)
S_{ALK}	Alkalinity, bicarbonate
S_F	Readily biodegradable substrate for fermentation
S_I	Soluble inert organics
S_j	Dissolved components in the ASMs
$S_{k k}$	<i>A posteriori</i> Cholesky factor of state covariance
$S_{k k-1}$	<i>A priori</i> Cholesky factor of state covariance
$\tilde{S}_{k k}$	Cholesky factor of augmented state covariance
S_{N_2}	Dinitrogen, released by denitrification
S_{NH_4}	Ammonium
S_{NOX}	Nitrite plus nitrate
S_{O_2}	Dissolved oxygen
$S_{\text{O}_2}^{\text{sat}}$	Saturated dissolved oxygen concentration
S_{PO_4}	Phosphate
S_S	Readily biodegradable substrates
S_y	Cholesky factor of measurement error covariance
t	Time
T	Number of available observations
T_h	Control horizon
$t_{\text{on,min}}$	Minimum air-on period within an aeration cycle
$t_{\text{on,max}}$	Maximum air-on period within an aeration cycle
$t_{\text{off,min}}$	Minimum air-off period within an aeration cycle
$t_{\text{off,max}}$	Maximum air-off period within an aeration cycle
u_a	State of blower (on/off) for aeration system

u_k	Control input vector at time step k
u_t	Control input at time t , $t \in \{1, \dots, T\}$
V_a	Working volume of the aeration tank
v_k	Measurement noise vector at time step k
V_x	Estimator prediction accuracy for state sequence
V_y	Estimator prediction accuracy for output sequence
w_k	Process noise vector at time step k
\tilde{w}_k	Augmented process noise vector at time step k , $\tilde{w}_k \triangleq [w_k^T, r_k^T]^T$
W	Weighting matrix
x_a	Vector of concentrations in the aeration tank
X_A	Autotrophic, nitrifying biomass
\hat{x}_{con}	Constrained state estimate
X_H	Heterotrophic biomass
X_I	Inert particulate organics
x_{in}	Concentrations in the influent
X_j	Particulate components in the ASMs
x_k	State vector at time step k
\tilde{x}_k	Augmented state vector at time step k , $\tilde{x}_k \triangleq [x_k^T, \xi_k^T]^T$
$\hat{x}_{k k}$	<i>A posteriori</i> state estimate
$\hat{x}_{k k-1}$	<i>A priori</i> state estimate
$\tilde{x}_{k k}$	<i>A posteriori</i> augmented state vector of UKF, $\tilde{x}_{k k} \triangleq [(x_{k k})^T, (w_{k k})^T, (v_{k k})^T]^T$
X_{PAO}	Phosphorus-accumulating organisms, PAO
X_{PP}	Stored poly-phosphate of PAO
X_{PHA}	Organic storage products of PAO
x_{rs}	Concentrations in the recycled sludge
X_S	Slowly biodegradable substrates
X_{SS}	Total suspended solids
X_{STO}	Organics stored by heterotrophs
y_k	Controlled output vector at time step k
$\hat{y}_{k k-1}$	<i>A priori</i> output estimate
y_t	Output observed from the real plant at time t
y_t^m	Model output at time t
γ	Filter parameter of the UKF
γ_K	Collinearity index
δ^{msqr}	Sensitivity measure
θ	Model parameter vector
θ^0	Vector of initial parameter values
$\hat{\theta}$	Vector of estimated parameter values
θ^*	Vector of correct parameter values
$\kappa(\cdot)$	Condition number of a matrix
Λ_c	Sigma point covariance weight of the UKF
Λ_m	Sigma point mean weight of the UKF
ξ_k	State of the disturbance model at time step k
$\sigma_{\max}(\cdot)$	Largest singular value of a matrix
$\sigma_{\min}(\cdot)$	Smallest singular value of a matrix

List of abbreviations

ASM	Activated Sludge Model
ASP	Activated Sludge Process
COD	Chemical Oxygen Demand
CSTR	Continuous Stirred Tank Reactor
DAE	Differential Algebraic Equation
DDE	Dynamic Data Exchange
DO	Dissolved Oxygen
Dymosim	Dynamic model simulator
EKF	Extended Kalman Filter
HRT	Hydraulic Retention Time
INCOOP	INtegrated process COntrol and plantwide OPTimization
IWA	International Water Association
KF	(standard) Kalman Filter
MHE	Moving Horizon Estimation
MPC	Model Predictive Control
ODE	Ordinary Differential Equation
OPC	OLE for Process Control
PDE	Partial Differential Equation
QP	Quadratic Programming
RMSE	Root Mean Squared Error
RSE	Root Squared Error
SNdN	Simultaneous Nitrification-denitrification
SRT	Sludge Retention Time
SRUKF	Square-root Unscented Kalman Filter
TN	Total Nitrogen
TP	Total Phosphorus
TUC	Telemark University College
UKF	Unscented Kalman Filter
WWTP	Wastewater Treatment Plant

Part I

Overview

Chapter 1

Introduction

1.1 Background

Water is one of our most precious resources, and wastewater treatment is becoming more and more important in a highly populated, industrialized world. Sewage treatment is accomplished most economically by processes with biological oxidation, and among them, the activated sludge process (ASP) is the most generally applied biological wastewater purification technique. This thesis deals with mathematical modeling and process control of ASPs for domestic wastewater treatment. The ASP is typically characterized by:

- requirement of a proper dynamic model to describe the complex system that incorporate a large number of biological and biochemical processes,
- uncertain model parameters due to the varying operation, influent, and sludge characteristics of the process,
- relatively few measurements for the key state variables due to the lack of cheap and reliable sensors and techniques,
- requirement of advanced control strategies because of the following incentives (see Olsson & Newell (1999) and Olsson et al. (2005)):
 - the increased public awareness as reflected in more stringent regulations is an efficient driving force,
 - control and optimization can help to minimize the operating costs and get more out of the existing plants,
 - rejection of large process disturbances, e.g. large variations in the load to the plant requires monitoring and process control,
 - the current manual control procedures are difficult to use for controlling more complex process alternatives.

1.2 Previous work

As for any industrial process, models in biological wastewater treatment have always played an important role in the process design and in the determination of control and operating strategies. The Activated Sludge Models (ASMs) presented by the International Water Association (IWA) (Henze et al. 2000) are generally accepted as the mathematical description of the biochemical processes in the ASP reactors.

Model representation is important in control oriented modeling. Traditionally, vendors of control systems have had proprietary/closed systems for model representation. However, a relatively recent trend is software developments supporting open component-based architectures and reuse of components based on object-oriented methods (see Heck et al. (2003)). One object-oriented, general-purpose modeling language is Modelica (Fritzson 2004), with commercial simulation environments such as Dymola from Dynasim (Dynasim 2004). Modelica/Dymola provides important advantages such as suitability for multi-domain modeling, the usage of general equations of physical phenomena, the reusability of model components, and the existence of a `WasteWater` library containing three ASMs (Reichl 2003).

In the early 1970s, Andrews (1974) depicted many potential benefits from the development and use of dynamic models for wastewater treatment processes and the incorporation of modern control systems into wastewater treatment plant (WWTP) design. In the last two decades, the application of advanced control techniques has received more and more interest due to the characteristics of ASPs. Advanced control is not the solution to all the problems of ASPs, but it can indeed help to reduce the requirement for long residence time, improve the effluent water quality, and/or decrease the use of chemicals. In general, the operational objective for a WWTP is to achieve the required effluent standards with the least operational costs (Olsson et al. 2005). Typically, advanced/modern control theory may include the following concepts (Ray 1981), (Balchen & Mummé 1988):

- *System identification*, which determines the model structure and estimates the model parameters (Söderström & Stoica 1989), (Walter & Pronzato 1997), (Ljung 1999). When the model is to be used for process performance evaluation and optimization, it is necessary to have an accurate description of the actual process. Although the ASMs are the state-of-the-art models for ASPs, the model parameters are often required to be adjusted to fit a certain set of data obtained from the real plant. However, a major problem encountered in calibration of ASP models is indeed the lack of identifiability of the model parameters (Vanrolleghem et al. 1995), (Petersen 2000), (Brun et al. 2002). Parameter identifiability analysis and estimation is thus an interesting area of research.
- *State estimation*. In processes where digital sensors are not available or are expensive, state estimation may be of significant practical importance (Gelb 1992), (Soroush 1998), (Crassidis & Junkins 2004), (Simon 2006). A major bottleneck in the application of advanced monitoring and control strategies for WWTPs lies in the difficulty of measuring reactant and product concentrations in a reliable and cost effective manner (Lindberg 1997), (Dochain 2003). Therefore, the application of state estimation techniques in ASPs is of great interest.
- *Optimal control theory*, which allows the design of control schemes that are optimal

in the sense that the controller performance minimizes some specified cost functional (Bubnicki 2005), (Bequette 2003). The optimal control algorithms belong to the class of model-based control strategies (Brosilow & Joseph 2002). Thus, the development of ASP models motivates the study of applying optimal control theory to the ASP control. However, the optimal control strategy in itself is unsuitable for practical implementation since it is an open-loop solution.

- *Model predictive control (MPC)*, which is the approach to implement the theory of optimal control on-line, introducing feedback from the process (Rossiter 2003), (Camacho & Bordons 2004), (Seborg et al. 2004). MPC is a proven advanced control technique, suitable for ASPs, especially due to its abilities to systematically handle constraints and to reduce the effects of process disturbances (see Zhao et al. (1994a), Lukasse et al. (1998), Holenda et al. (2007), etc.)

1.3 Scope and outline of thesis

The goal of this thesis is to investigate methods for developing and implementing activated sludge plant models in an efficient way to provide a good prediction of process dynamics, and to study possible improvements in process operation and performance through applications of advanced modeling and control techniques. In a broad perspective, the thesis will cover the modeling methodology of ASPs and all the aforementioned topics in advanced control.

The modeling and control techniques described in this thesis are in principle applicable to any ASP; as case studies, a laboratory-scale nitrogen (N) removal process and a full-scale phosphorus (P) removal plant: the Duvbacken WWTP, are introduced. The studies focus on the modeling and control of the biochemical processes occurring in the biological reactors, based on the complete ASM models. For a complex system like the ASP, the ASM model is probably the simplest description of the processes in consideration of the predictive purpose of the model, and model reduction of the ASM model is not considered in the thesis. With respect to the control of ASPs, the implementation and testing of controllers is not covered in this work. The developed control strategies are evaluated through simulation studies, since simulations make it possible to efficiently compare the different control techniques.

This thesis is composed of two parts. Part I basically gives an overview of the results obtained in the papers provided in Part II. However, a few results in Part I are not presented in any paper, either because the topics do not fit with the scope of the papers or because the results were not ready at the time of submission or publishing of papers. Chapter 2 gives an introduction to ASPs and to the object-oriented modeling of ASPs. Chapter 3 provides an overview of the state estimation techniques and compares the performances of these techniques applied to a nonlinear ASP. Parameter estimation of the ASPs is described in chapter 4, where both estimation of kinetic parameters and model disturbances are investigated. Chapter 5 presents the process control strategies used for controlling the aeration system of the ASPs. Chapter 6 gives a short introduction to a laboratory case study. Finally, chapter 7 contains conclusions and suggestions for future work.

Part II provides six papers published or submitted in conjunction with this thesis:

Abstract of Paper A A general description of wastewater treatment based on activated

sludge is given, with emphasis on the ASM2d model. Next, a systematic approach for experimental parameter identifiability analysis is described. After a presentation of the wastewater treatment plant at Duvbacken in Gävle, Sweden, the experimental identifiability of the dynamic model is analyzed. Out of the 45 original parameters in the model, the analysis indicates that with the given experimental conditions, 12 parameters can be identified.

Abstract of Paper B This paper carries out the parameter identifiability analysis with the approach described in Paper A, and fits the model to experimental data from the Duvbacken plant. The model is implemented in an object-oriented, general-purpose modeling language Modelica, and is executed in Matlab for further analysis. A standard least squares criterion is used in combination with a direct optimization method to calculate the improved parameter estimates, and the adjusted model is validated against experimental data.

Abstract of Paper C The problem of optimal state estimation in a biological wastewater treatment process is considered. The standard Kalman filter (KF) and its extensions: the extended Kalman filter (EKF) and the unscented Kalman filter (UKF), are used to estimate the unmeasured state. The prediction of the state with the standard KF is poor due to the high nonlinearity of the process. Thus, the nonlinear estimation approaches are focused, with a comparison between the EKF and the UKF. The simulation results show that the UKF provides slightly better state estimate than the EKF for both an observable process and an unobservable process.

Abstract of Paper D This paper deals with optimal control of a biological wastewater treatment process designed to remove nitrogen with simultaneous nitrification/ denitrification in a single reactor. The objective of the control is to force the total nitrogen concentration to track given set points. The optimization problem is specified and formulated using a quadratic performance index. Dynamic optimization is used in model predictive control (MPC) based on a linearized model. The control strategy is successfully applied to the simulation model of the pilot plant.

Abstract of Paper E The optimal control of a nitrogen removal process with intermittent aeration is discussed in order to design the aeration strategy (air-on and air-off periods) which minimizes the energy dissipated by the aeration system, with adherence to the limits of the EU effluent requirements and the operating constraints. The optimization problem is formulated and used with a receding horizon in nonlinear MPC based on the complete ASM3 model. The simulation results show that significant improvements can be achieved with respect to the initial aeration strategy, with reductions in the energy consumption by 33% for the optimal control over one day and by 16% for MPC over 7 days.

Abstract of Paper F The model-based optimal control and predictive control of the intermittently aerated nitrogen removal process is presented. The optimization problem is developed based on the discussion in Paper E. The MPC aeration profile guarantees that the plant fulfills the effluent requirements at any time over long time periods, and

significant energy saving is also obtained when comparing MPC to three traditional rule-based control strategies.

1.4 Main contributions

The main contributions of this thesis are:

- Two activated sludge plant models are developed based on standard ASM models: one for the laboratory-scale N-removal plant and one for the Duvbacken P-removal plant. The model implementation in Modelica/Dymola is described, and the possible Dymola interfaces are introduced.
- Some widely used state estimation techniques, such as standard Kalman filter (KF), extended Kalman filter (EKF), unscented Kalman filter (UKF), square-root unscented Kalman filter (SRUKF), and moving horizon estimation (MHE), are discussed, and the performances of these filters are evaluated with application to the N-removal process.
- The systematic approach to parameter identifiability prior to parameter estimation (Brun et al. (2002), Dueñas Díez et al. (2006)) is described and applied to the Duvbacken-model.
- A methodology for simultaneously estimating state and disturbance in ASPs is developed using the extended Kalman filter, and the suggested method is illustrated on the Duvbacken WWTP, based on experimental data.
- Advanced control strategies for the ASP aeration system, based on linear and nonlinear MPC techniques, are suggested, with promising results compared with the traditional rule-based controllers.
- Some recommendations for future work are suggested for a laboratory N-removal project.

Chapter 2

Activated Sludge Processes for Wastewater Treatment

2.1 Introduction

Wastewater generally contains large quantities of organic matter that results in the depletion of natural oxygen resources in the water. Also, the nutrients in the wastewater, such as nitrogen and phosphorus, may lead to algae growth, which causes oxygen depletion, odor problems, excretion of toxins, etc. Therefore, for wastewater treatment, the main objective is to remove or reduce the content of organic matter, as well as nitrogen (N) and phosphorus (P). A general description of biological wastewater treatment, containing wastewater characterization, wastewater species, biological processes, and wastewater kinetics and dynamic models, is given in Paper A.

The strict European Union Guideline *Urban Wastewater Directive 91/271/EEC*¹ (a summary is shown in Table 2.1) came out in 1991, where the limits on effluent concentrations of total chemical oxygen demand (COD_{tot})², total suspended solids (SS), total phosphorus (TP)³, and total nitrogen (TN)⁴ are required. Due to legislation on wastewater treatment,

¹For more information, see http://ec.europa.eu/environment/water/water-urbanwaste/index_en.html.

² COD_{tot} indicates the total organic carbon in wastewater. The COD_{tot} related to the components in ASM3 is defined as (Petersen 2000)

$$\text{COD}_{\text{tot}} \triangleq S_{\text{I}} + S_{\text{S}} + X_{\text{I}} + X_{\text{H}} + X_{\text{STO}} + X_{\text{A}}.$$

³Total phosphorus (TP) includes both soluble phosphorus and particulate phosphorus. The TP related to the components in ASM2d is defined as

$$\text{TP} \triangleq S_{\text{PO}_4} + S_{\text{F}} \cdot i_{\text{P},S_{\text{F}}} + S_{\text{I}} \cdot i_{\text{P},S_{\text{I}}} + X_{\text{PP}} + X_{\text{S}} \cdot i_{\text{P},X_{\text{S}}} + X_{\text{I}} \cdot i_{\text{P},X_{\text{I}}} + (X_{\text{H}} + X_{\text{PAO}} + X_{\text{AUT}}) \cdot i_{\text{P},\text{BM}}.$$

⁴Total nitrogen (TN) is comprised of organic nitrogen, ammonia, nitrite, and nitrate (Metcalf & Eddy 2003). The TN related to the components in ASM3 is defined as

$$\text{TN} \triangleq S_{\text{NH}_4} + S_{\text{NOX}} + S_{\text{I}} \cdot i_{\text{N},S_{\text{I}}} + S_{\text{S}} \cdot i_{\text{N},S_{\text{S}}} + X_{\text{I}} \cdot i_{\text{N},X_{\text{I}}} + X_{\text{S}} \cdot i_{\text{N},X_{\text{S}}} + (X_{\text{H},\text{eff}} + X_{\text{A},\text{eff}}) \cdot i_{\text{N},\text{BM}}.$$

Table 2.1: Effluent requirements in EU directive 91/271/EEC.

Parameters	Component ^{a)}	Efficiency ^{b)} (% of reduction)
Chemical oxygen demand (COD _{tot})	125 g m ⁻³	75%
Total suspended solids (SS)	35 g m ⁻³	90%
Total phosphorus (TP)	2 g m ⁻³ (10, 000–100, 000 p.e.)	80%
	1 g m ⁻³ (> 100, 000 p.e.)	80%
Total nitrogen (TN)	15 g m ⁻³ (10, 000–100, 000 p.e.)	70 – 80%
	10 g m ⁻³ (> 100, 000 p.e.)	70 – 80%

a) These values of concentration are annual average.

b) Either concentration or the percentage of reduction shall apply.

One or both parameters are to be applied depending on local conditions.

activated sludge plants has been introduced on a large scale. In an activated sludge process (ASP), wastewater is mixed with a bacterial biomass suspension (called an activated sludge) which degrades the pollutants. Depending on the plant design and the specific application, an activated sludge plant can achieve biological nitrogen removal and biological phosphorus removal, besides removal of organic matter.

The modeling of ASPs is a very active research area, see e.g. Dochain & Vanrolleghem (2001) and Hu et al. (2003) for reviews on the subject. It should be mentioned that there is an essential difference between a WWTP model and an activated sludge model, as discussed in Gernaey et al. (2004). The WWTP model indicates the ensemble of activated sludge model, hydraulic model, oxygen transfer model, and settler model needed to describe an actual plant. The activated sludge model is used to represent the biological (and chemical) reactions taking place in one activated sludge reactor.

To date, the most frequently used activated sludge models are the family of Activated Sludge Models (ASMs) developed by the International Water Association (IWA). The Activated Sludge Model No. 1 (ASM1) (Henze et al. 1987) is usually considered as the reference model, which was primarily developed for municipal ASPs to describe the removal of organic carbon substances and nitrogen. With basically the same goals as ASM1, the ASM3 model (Gujer et al. 1999) was developed for nitrogen removal ASPs. The ASM3 model was intended to become the new standard model, correcting a number of defects that have been noticed during the usage of the ASM1 model. In addition, the models including biological phosphorus are the ASM2 model (Gujer et al. 1995) and its extension ASM2d (Henze et al. 1999), which extend the capabilities of ASM1 to represent the dynamics of phosphorus removal. Henze et al. (2000) also mentioned that modeling of phosphorus removal (as mentioned in ASM2 and ASM2d) can be integrated with ASM3 by including additional processes and states.

This chapter gives an introduction to the ASP and to the ASM model implementation using an object-oriented modeling method. Two activated sludge processes are studied: one at laboratory scale for N-removal is described in section 2.2, and one at the plant scale for P-removal is described in section 2.3. The object-oriented modeling method using Modelica/Dymola is presented in section 2.4. In addition, appendix 2.A describes a Python

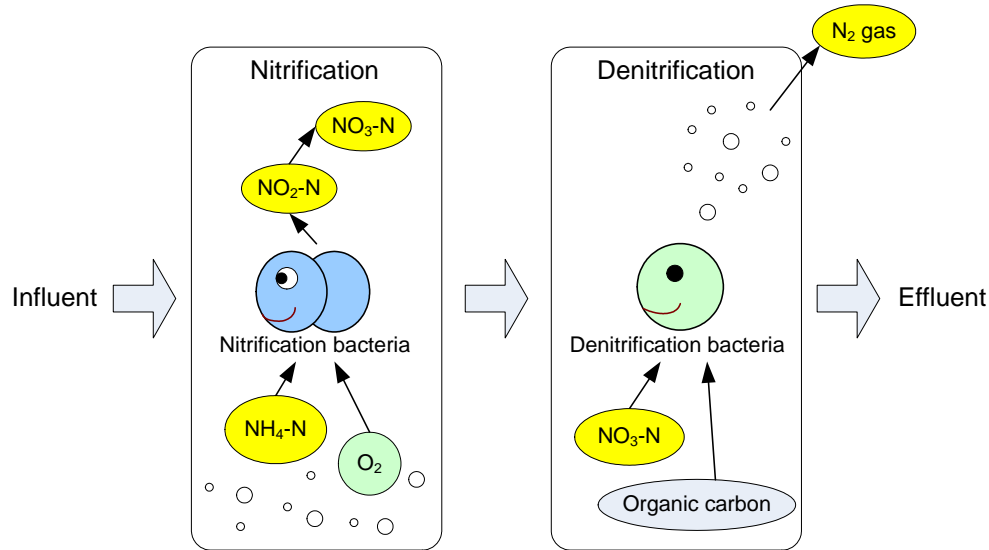


Figure 2.1: Nitrogen transformations in biological treatment processes, based on description in Metcalf & Eddy (2003).

tool for a Dymola-Python interface.

2.2 Nitrogen removal process

2.2.1 Biological nitrogen removal

Nitrogen is removed in these two steps: ammonium ($\text{NH}_4\text{-N}$) is first oxidized to nitrate ($\text{NO}_3\text{-N}$) under nitrification; the produced nitrate is then transformed into nitrogen gas (N_2) under denitrification. The nitrogen transformations are schematically shown in Figure 2.1. Hence, all of the biological nitrogen removal processes include an aerobic zone in which nitrification occurs. Anoxic condition must also be included to provide denitrification to complete the objective of total nitrogen removal by both $\text{NH}_4\text{-N}$ oxidation and $\text{NO}_3\text{-N}$ and $\text{NO}_2\text{-N}$ reduction to nitrogen gas (Metcalf & Eddy 2003). Thus, the approaches for designing anoxic conditions can be categorized as

1. separate anoxic zones which are either spatial distributions in the reactors or separate reactors from aeration zones (Droste 1997),
2. simultaneous nitrification-denitrification (SNdN) processes where both nitrification and denitrification occur in the same tank without alternating (in time or in space) aerobic and anoxic conditions (Collivignarelli & Bertanza 1999), or
3. intermittently aerated processes where continuous alternating aerobic-anoxic conditions are created by switching the air on and off in the same tank (Ip et al. 1987), (Nakajima & Kaneko 1991), (Hao & Huang 1996), (Kimochi et al. 1998).

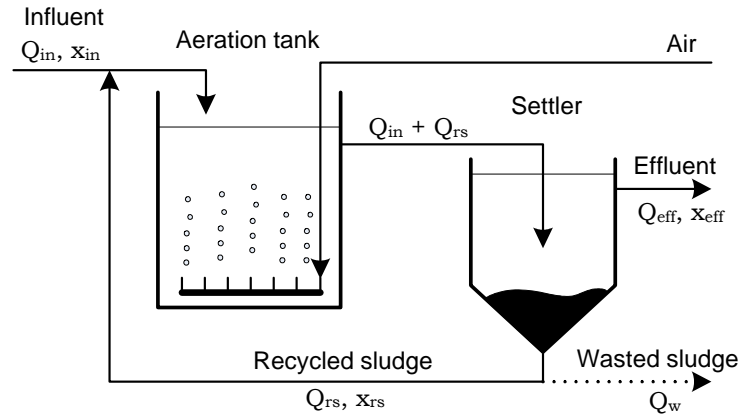


Figure 2.2: Schematic diagram of a typical activated sludge process. (Adapted from Chachuat et al. (2005).)

In the last two decades, intermittently aerated ASPs have been developed and widely used for small-sized WWTPs. Compared to conventional ASPs, intermittently aerated ASPs have the following distinctive features:

- The process offers significant energy savings due to no continuous transport of oxygen to the anoxic phase and high oxygen transfer efficiency during the initial period of aeration (Ip et al. 1987), (Kim et al. 2000).
- The degree of operational flexibility is high because reaction times of nitrifying/denitrifying can easily be adjusted by changing the aeration scheduling, which makes the process better accommodate variable influent loading conditions and suitable for optimization of operational costs (Hao & Huang 1996), (Isaacs 1997) .
- The single tank configuration provides space saving, low investments of instruments, and easy plant retrofitting.

In this work, mainly the intermittently aerated process is studied. In addition, control of the SNdN process is also considered in subsection 5.3.1.

2.2.2 The laboratory-scale nitrogen removal plant

A laboratory-scale N-removal plant is considered in Papers D, E and F, for modeling and control studies of an ASP. The plant was originally built to study the removal of nitrogen from the wastewater of an oil refinery by Baksaas (1996); later the reactor system was modified and used for municipal wastewater treatment in student projects at Telemark University College. This nitrogen removal process consists of a unique aeration tank with a working volume $V_a = 401$ and a cylindrical settler located in the aeration tank; the process is schematically shown in Figure 2.2. More details are available in Paper E. Hence, the plant model consists of an aeration tank model and a settler model.

Aeration tank model

The derivation of the reactor model is based on mass balances of the wastewater species, which are generally expressed as follows:

$$\text{Accumulation} = \text{Inflow} - \text{Outflow} + \text{Reaction}.$$

The ASM3 model is used to describe the biological processes involved in the aeration tank. The model contains 13 state variables described in Table 2.2. Assuming perfect mixing in the reactor, the mass balance of each component is expressed by a nonlinear ordinary differential equation (ODE), thus the mass balance in the aeration tank results in:

$$\frac{dx_a}{dt} = \frac{Q_{in}x_{in} + Q_{rs}x_{rs} - (Q_{in} + Q_{rs})x_a}{V_a} + r + e_1\mathcal{A}_{O_2},$$

where $x_{in}, x_{rs}, x_a \in \mathbb{R}^{13}$ contain the concentrations in the influent, in the recycled sludge, and in the reactor, respectively; their components are

$$x_l = [S_{O_2,l} \ S_{I,l} \ S_{S,l} \ S_{NH_4,l} \ S_{N_2,l} \ S_{NOX,l} \ S_{ALK,l} \ X_{I,l} \ X_{S,l} \ X_{H,l} \ X_{STO,l} \ X_{A,l} \ X_{SS,l}]^T$$

$l \in \{in, rs, a\}$; $r \in \mathbb{R}^{13}$ is the vector formed by the reaction rates of each component (defined in ASM3); e_1 is the standard basis for the first coordinate in \mathbb{R}^{13} . \mathcal{A}_{O_2} describes the oxygen transfer:

$$\mathcal{A}_{O_2} = u_a \cdot K_L a (S_{O_2}^{sat} - S_{O_2,a}),$$

where $K_L a$ is the oxygen transfer coefficient which is a function of the supplied air flow rate, and $S_{O_2}^{sat}$ is the saturated dissolved oxygen concentration ($S_{O_2}^{sat} \approx 10 \text{ g m}^{-3}$). According to the aeration modes of ASPs, the control input of the process may be described as:

- For the continuously aerated ASP, variable u_a is chosen 1, and $K_L a$ is often considered as control input.
- For the intermittently aerated ASP, the value of $K_L a$ is chosen constant, instead, variable u_a is a binary sequence switching between 0 and 1 that represents the state of the blower (on/off) aerating the plant. Hence, the control input is the sequence of switching times, i.e., times when the blower switches on/off.

Settler model

The purpose of the secondary settler is twofold:

- to efficiently separate the solids and liquid — the clarification;
- to rapidly thicken the sludge at the bottom of the settler — the thickening process.

Olsson & Newell (1999) points out that dynamic modeling of settlers has received much less attention compared to modeling of the aeration tank, and an elegant solution still evades the activated sludge modelers. In the literature, different settler models have been introduced mainly based on mass balances and on settling flux theories, see e.g. Jeppsson (1996), Dochain & Vanrolleghem (2001), and Takács et al. (1991). However, for simplicity, the settler is often considered as an ideal settler. We thus assume that

Table 2.2: State variables of the ASM3 model.

State	Description	Unit
<i>Dissolved components (S_j)</i>		
S_{O_2}	Dissolved oxygen	g O ₂ m ⁻³
S_I	Soluble inert organics	g COD m ⁻³
S_S	Readily biodegradable substrates	g COD m ⁻³
S_{NH_4}	Ammonium	g N m ⁻³
S_{N_2}	Dinitrogen, released by denitrification	g N m ⁻³
S_{NOX}	Nitrite plus nitrate	g N m ⁻³
S_{ALK}	Alkalinity, bicarbonate	mol HCO ₃ ⁻ m ⁻³
<i>Particulate components (X_j)</i>		
X_I	Inert particulate organics	g COD m ⁻³
X_S	Slowly biodegradable substrates	g COD m ⁻³
X_H	Heterotrophic biomass	g COD m ⁻³
X_{STO}	Organics stored by heterotrophs	g COD m ⁻³
X_A	Autotrophic, nitrifying biomass	g COD m ⁻³
X_{SS}	Total suspended solids	g SS m ⁻³

1. the separation of liquid and solid/particulate phases is perfect and no reaction occurs in the settler, and
2. the sum of all settler outflows equals the settler influent flowrate.

Based on these assumptions, the settler model can be expressed as follows:

Effluent concentration

$$S_{j,\text{eff}} = S_{j,a}, \quad X_{j,\text{eff}} = 0$$

Recycled sludge concentration

$$S_{j,rs} = S_{j,a}, \quad X_{j,rs} = \frac{Q_{in} + Q_{rs}}{Q_{rs} + Q_w} X_{j,a}.$$

General model

In summary, the general model of the dynamic system can be simply described as a set of ODEs:

$$\frac{dx_a}{dt} = \mathcal{F}(x_a, u, t),$$

in which u is the control input described in modeling of the aeration tank.

This simplified expression will be used in the definition of the optimization problem in chapter 5.

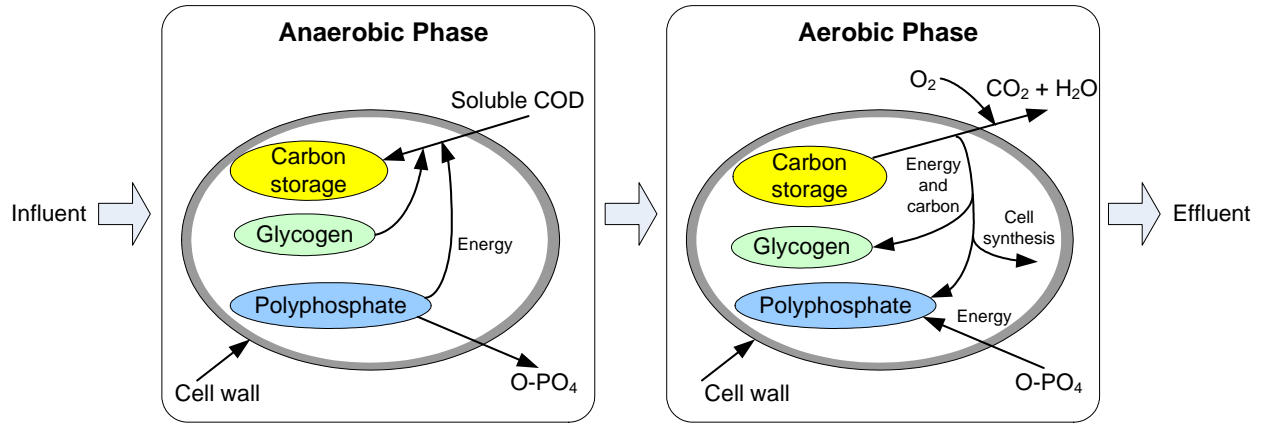


Figure 2.3: Principle of biological phosphorus removal. (Adapted from (Jenkins 2007).)

2.3 Phosphorus removal process

2.3.1 Biological phosphorus removal

In biological removal of phosphorus, the phosphorus is released under anaerobic conditions and taken up under aerobic/anoxic conditions, see Figure 2.3.

1. In the anaerobic phase, the phosphorus accumulating organisms (PAO) take up carbon sources and produce intracellular polyhydroxylalkanoate (PHA) storage products. The energy needed for this process is derived by the degradation of intracellular polyphosphates, resulting in a subsequent release of orthophosphate (O-PO_4). Some glycogen contained in the cell is also used.
2. In the subsequent aerobic phase, the stored PHA is metabolized, providing energy from oxidation and carbon for new cell growth. Some glycogen is also produced. Using the released energy, soluble O-PO_4 is removed from solution and incorporated into polyphosphates within the bacterial cell. The new biomass with high polyphosphate storage accounts for phosphorus removal.

Since biological phosphorus removal is an extremely complex procedure, the mechanisms of phosphorus removal are still under investigation. The mechanisms as currently understood are summarized by the ASM2 and ASM2d models (Henze et al. 2000).

2.3.2 The Duvbacken WWTP

Duvbacken is a wastewater treatment plant which is designed to treat the municipal wastewater from 100,000 people equivalents (p.e.) in the city of Gävle, Sweden, see Figure 2.4. The average feed flowrate to the plant is approximately $1,600 \text{ m}^3/\text{h}$. The first construction phase was completed in 1967, and consisted of facilities for mechanical and biological treatment according to the activated sludge approach. This plant has been recently modified to mainly remove phosphorus, together with nitrogen and organic matter. The treatment



Figure 2.4: Aerial photo of the Duvbacken wastewater treatment plant. (Taken from Duvbacken plant leaflet.)

process of the plant consists of grit removal, primary treatment, chemical treatment, biological treatment, and sludge treatment. In this work, we only consider biological treatment of wastewater, including processes for biological reactions and clarification. The process layout of the biological treatment stage is described in Paper A, where the process consists of 1 anaerobe step, 3 aerobe steps, and 1 secondary sedimentation step.

A dynamic model of the plant is developed, based on the ASM2d model. A minor simplification of the presentation here compared to the full ASM2d, is that we neglect the possibility of phosphorus to bind to metal-hydroxides and the associated chemical processes. In the plant model, each of the 4 reaction steps (anaerobe, aerobe 1–3) is modeled as a perfectly stirred reactor. This leads to 17 states (listed in Table 2.3) in each of the reactors, i.e., in total 68 states for the plant model. The model has 42 parameters, i.e., kinetic parameters of the simplified ASM2d model. For the settler, it is assumed that 99.8% of the sludge in the settler influent is deposited, and the remaining 0.2% of the sludge flows out in the effluent of the plant. The concentrations of dissolved components (S_j) in the inflow of the settler are equal to the S_j contained in the outflows of the settler.

2.4 Object-oriented modeling

During the previous decade of research in modeling and simulation, the concept of object-oriented modeling has achieved a prominent position. Many studies have demonstrated that object-oriented concepts can be successfully employed to support hierarchical structuring, reuse, and evolution of large and complex models, independently from the application domain and specialized graphical formalism (Maffezzoni et al. 1999), (Bortolin et al. 2004), (Bastogne 2004).

The object-oriented approach to modeling is a credible methodology to deal with complex and heterogeneous systems, while computer simulation is the means to use such methods in

Table 2.3: State variables of the AMS2d model.

State	Description	Unit
<i>Dissolved components (S_j)</i>		
S_{O_2}	Dissolved oxygen	g O ₂ m ⁻³
S_F	Readily biodegradable substrate for fermentation	g COD m ⁻³
S_A	Fermentation product (acetate)	g COD m ⁻³
S_{NH_4}	Ammonium	g N m ⁻³
S_{NO_3}	Nitrate (plus nitrite)	g N m ⁻³
S_{PO_4}	Phosphate	g P m ⁻³
S_I	Inert, non-biodegradable organics	g COD m ⁻³
S_{ALK}	Bicarbonate alkalinity	mol HCO ₃ ⁻ m ⁻³
S_{N_2}	Dinitrogen	g N m ⁻³
<i>Particulate components (X_j)</i>		
X_I	Inert, non-biodegradable organics	g COD m ⁻³
X_S	Slowly biodegradable substrate	g COD m ⁻³
X_H	Heterotrophic biomass	g COD m ⁻³
X_{PAO}	Phosphorus-accumulating organisms, PAO	g COD m ⁻³
X_{PP}	Stored poly-phosphate of PAO	g P m ⁻³
X_{PHA}	Organic storage products of PAO	g COD m ⁻³
X_{AUT}	Autotrophic, nitrifying biomass	g COD m ⁻³
X_{TSS}	Particulate material as model component	g TSS m ⁻³

practice for assessing the global system performance. For wastewater treatment systems, a very good survey of modeling and simulation tools is provided in Olsson & Newell (1999) and Germaey et al. (2004).

2.4.1 Modelica and Dymola

In 1996 an international effort was initiated for the purpose of bringing together expertise in object-oriented physical modeling and defining a modern uniform modeling language; this language was named Modelica. Modelica is an object-oriented, general-purpose modeling language specially designed for representing complex and heterogeneous physical systems (Tiller 2001), (Fritzson 2004). Today, there are at least four commercial simulation environments based on the Modelica language: Dymola, MathModelica System Designer, MOSILAB, and SimulationX, as well as a number of useful tools such as EXITE, TISC, and StateViewer. In addition, the Open Source Modelica (named OpenModelica) is developed by the Programming Environment Laboratory (PELAB) at Linköping University, Sweden. For more information, we refer the reader to the Modelica home page: <http://www.modelica.org>.

In this work, Modelica is used with one of its commercial simulation environments, Dymola from Dynasim (<http://www.dynasim.se>), as the model representation tool for activated sludge systems. Compared with most other modeling and simulation tools available

today, such as gPROMS⁵ and Matlab & Simulink, Modelica/Dymola offers several important advantages:

For modeling and simulation in general

- Object-oriented architecture allows the user to easily modify and reuse of model components.
- Equation-based modeling supports many different model descriptions, such as ODEs, differential algebraic equations (DAEs), bond graphs, Petri nets, etc.⁶
- Acausal modeling, meaning that the model's terminals do not necessarily have to be assigned an input or an output role, enables both simpler models and more efficient simulation.
- Hybrid DAE systems can present both continuous-time behavior and discrete-time behavior of the model.
- Class parameters make it possible to parameterize models in a high-level fashion by exchanging submodels or even whole subsystems for a compatible type replacement.
- The Dymola model editor offers a good intuitive user interface, and the graphical representation of the model is embedded in the Modelica code.

For modeling and simulation of activated sludge processes

- A free Modelica library `WasteWater` (Reichl 2003) is available, which contains three activated sludge models (ASM1, ASM2d, and ASM3) and several secondary clarifier models. The `WasteWater` library includes relatively complete components needed for modeling and simulation using ASMs. In addition, with the flexibility of the Modelica language, the user can easily extend the specific components.
- Modelica provides flexibility to build large plant models. A full wastewater plant usually contains many components. The hierarchical modeling support and available basic models are vital in helping the user to deal with the complexity.
- It is possible to include measurement data as input or data tables (using the component `CombiTimeTable`) to the simulator. Since measurement data are not always represented with an equidistant time interval, Dymola provides possibilities to present time tags to each measurement point.

⁵gPROMS is an equation-based modeling system designed primarily for chemical process engineering. Currently, Modelica has less support for process models than gPEOMS, but the modeling scope is more general. More detailed comparison of language features between gPROMS and Modelica can be found in Tummescheit (2002).

⁶gPROMS includes a notation for partial differential equations (PAEs) in the language, while integration of language constructs for PAEs is still under development in Modelica.

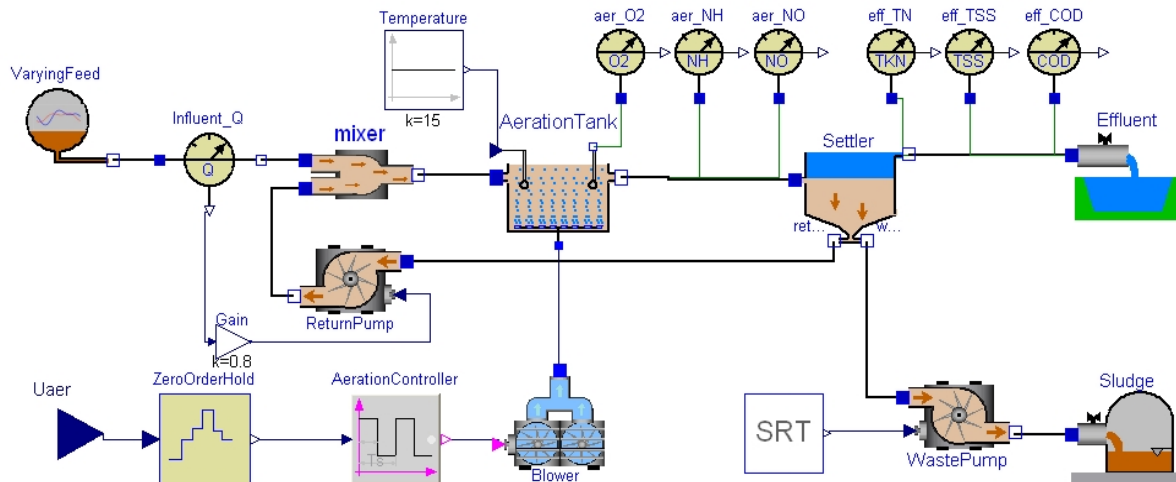


Figure 2.5: Model diagram of the laboratory-scale N-removal plant.

2.4.2 Model implementation

Although the WWTP is often a very large and complex system, within the Dymola simulation environment, the process model can be easily constructed by dragging, dropping, and connecting components chosen from the `WasteWater` library. The model diagrams of the laboratory-scale N-removal plant and the Duvbacken P-removal plant are shown in Figures 2.5 and 2.6. More details of the models implementation for these two presented ASPs can be found in Papers E and B, respectively.

2.4.3 Dymola interfaces

Dymola offers an open interface to other programs. A well-developed interface exists for Matlab/Simulink (MathWorks 2007), (Dynasim 2004). Recently, a prototype of an interface to LabVIEW version 8.20 is available in Dymola version 6.1 (Dynasim 2007). In addition, the author developed a Dymola-Python interface to execute Modelica models in the Python environment, to test out the possibility.

Dymosim – Dynamic model simulator

Dymosim – **dynamic model simulator** – is a stand-alone program that contains the machine code necessary for model simulation with event handling support (Dynasim 2004). When translating a model in Dymola, an executable Dymosim (`dymosim.exe`) is generated. Dymosim can be compiled with no graphical user interface or as a Windows application with Dynamic Data Exchange (DDE) server capabilities. With the different operating modes, the `dymosim` executable can be used for simulation, linearization, and initial value calculation. Moreover, the `dymosim` executable uses files to define its inputs and outputs. Figure 2.7 schematically presents the relationship between Dymosim and these files. The `dymosim` executable can be independently called from environments other than Dymola. Next, two

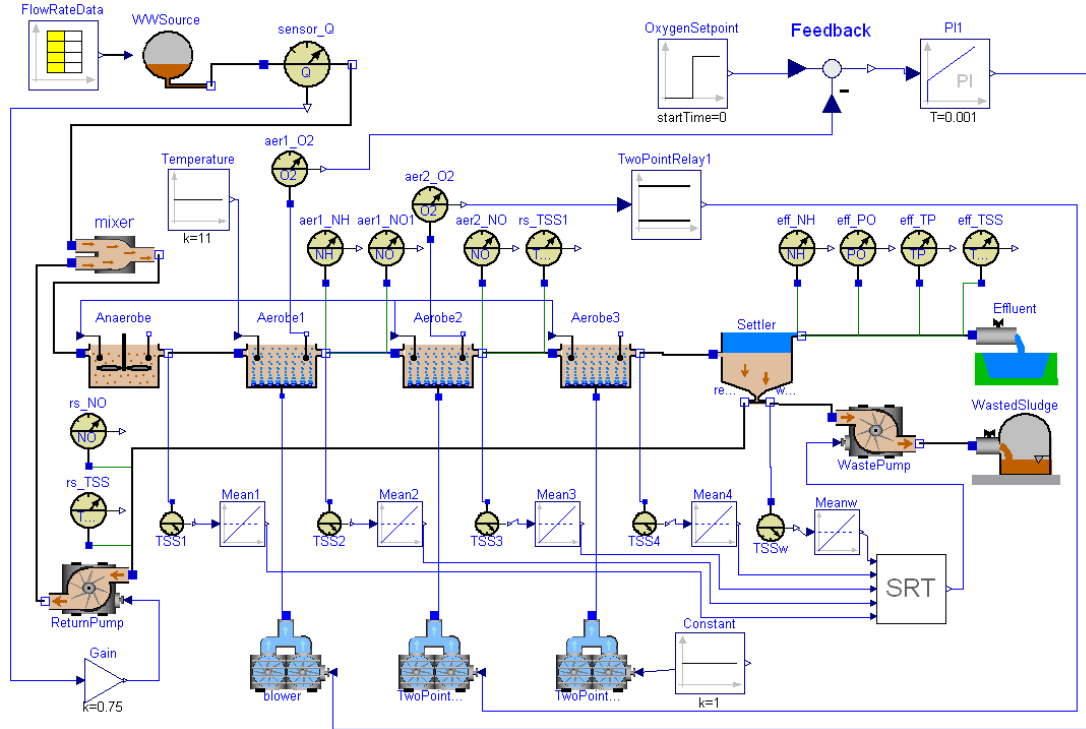


Figure 2.6: Model diagram of the Duvbacken P-removal plant.

interfaces to Dymola will be discussed.

Dymola-Matlab interface

Dymola provides convenient interfaces to Matlab and the popular block-diagram simulator Simulink. This enables the powerful physical modeling capabilities of Modelica to be combined with efficient control system design using Matlab/Simulink. The dymosim executable can be called from Matlab using the function `dymosim.m`. Execution of the Modelica model in Matlab is described in Paper B. In addition, the model compiled into a dymosim executable can be linearized using the `'-l'` modifier of the dymosim command⁷. Matlab allows us to execute the dymosim executable modifier using a Matlab script function:

```
eval(['! dymosim ', '-l ', 'dslin.mat']);
```

Then, the linearized model saved in `dslin.mat` can be loaded into Matlab using the Dymola add-on m-function `tloadlin` which loads matrices `A,B,C,D` and string vectors `uname`, `yname`, and `xname`.

For the approaches on how to transform Modelica models into Simulink see Dynasim (2004) and Dempsey (2003).

⁷As a stand-alone program, Dymosim accepts command line arguments defined in Dynasim (2004).

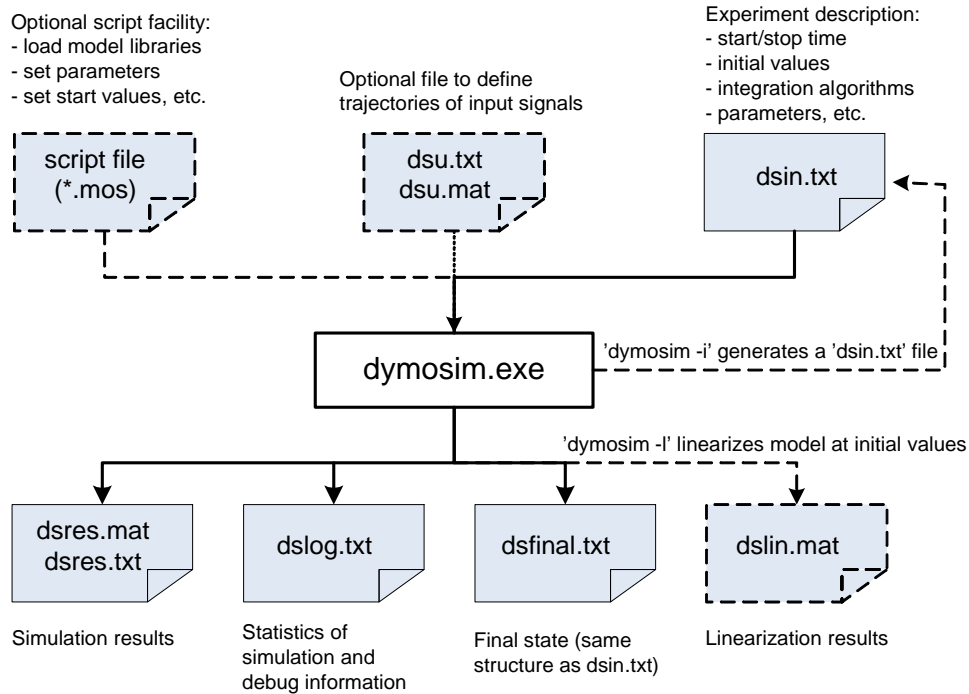


Figure 2.7: Connection of dymosim executable file and its input and output files, based on (Videla 2007). Dashed lines mean that the connection is optional.

Dymola-Python interface

As a dynamic object-oriented programming language, Python may be an alternative tool for prototyping algorithms in advanced control techniques. Compared to Matlab, Python has some clear advantages (Langtangen 2006), such as:

- the Python environment is completely open and made for integration with external tools,
- a complete toolbox/module with lots of functions and classes can be contained in a single file (in contrast to a number of M-files),
- object-oriented programming is better developed,
- the source code is free and runs on more platforms.

Unfortunately, Dymola does not have the functionality to support the communication with Python. Thus, a Dymola-Python interface module⁸, called `dymosimPy`, is developed in this work in order to execute a Modelica model and load simulation results in Python. Table 2.4 summarizes the functions in the `dymosimPy` module. These functions are developed using similar principles as for the Matlab functions offered by Dymola. For instance, the dymosim executable is called by the `dymosim` function in Python with the command line

⁸A module is a file containing Python definitions and statements.

Table 2.4: Functions in the `dymosimPy` module.

Function	Description
<code>dymosim</code>	Perform simulation of a Modelica model, save and load results
<code>parload</code>	Load parameter values and names from the Dymola input file
<code>initload</code>	Load initial state values and names from the Dymola input file
<code>dymload</code>	Load simulation results from a <code>.mat</code> file
<code>dymget</code>	Get simulation result of a specific single variable
<code>tnlist</code>	List a string vector with preceding row numbers
<code>loadName</code>	Load string variables from the Dymola input file
<code>findIndex</code>	Find indices of a value in a list
<code>setInitialValue</code>	Set a list (initial values) in the <code>initialValue</code> array

```
d = dymosimPy.dymosim(exp,x0,p,quiet,file)
```

By running this command, the `dymosim` executable reads the experiment description from the input file specified in the `file` variable, performs simulation, stores the simulation result in `dsres.mat`, and loads the result into a dictionary⁹ data structure in Python.

A detailed introduction to the `dymosimPy` module is given in appendix 2.A.

2.A `dymosimPy` — A Python module for Dymola-Python interface

Python is a dynamic object-oriented programming language that can be used for many kinds of software development. Python offers powerful features such as classes, modules, exceptions, dynamic typing and very high level collection data types in a concise, regular, and very readable syntax. Both the Python language and software packages are free, and a huge collection of library components written in or interfaced with Python exists. Hence, Python may be a powerful integration tool for advanced control. It is of interest to develop a Dymola-Python interface to call the `dymosim` executable from the Python environment. Here, a Dymola-Python interface module, called `dymosimPy`, is developed to execute a Modelica model and load simulation results in Python.

Actually, using Python for development of Modelica tools has been the subject of several studies. German National Research Center for Information Technology (GMD) has developed a Modelica translator in the Python-based development environment TRAP (Translator RAPid Prototyping) for integration in the Smile¹⁰ dynamic simulation environment. An open source research prototype compiler¹¹ for μ Modelica (a subset of Modelica) is implemented in

⁹A dictionary, also called hash or associative array in other computer languages, is a kind of list where the index, referred to as key, can be an arbitrary text. For example, a dictionary can be initialized by defining `dict = {}`; `dict['list'] = [1,2,3]`.

¹⁰Smile (<http://www.first.fraunhofer.de/en/simulation-technology>) is an object-oriented dynamic simulation environment developed by Technische Universität Berlin and GMD.

¹¹More information and the source code can be found on the website <http://msdl.cs.mcgill.ca/people/wxu/research.html>.

Python with the focus on acausal modeling. The GENSIM project is developing the generic simulation tool MOSILAB (Schmitz 2005) for the analysis of mixed time-continuous/time-discrete (hybrid) models of heterogeneous technical systems. MOSILAB uses the Python language and interpreter as its central mechanism for experiment control.

For more information on the Python language, see the Python home page <http://python.org> and the reference books Lutz & Ascher (2003), Gauld (2000), and Langtangen (2006). However, since Python is nonprofit, the technical support and documentation of Python are not as good as for other commercial software.

2.A.1 Python vs. Matlab

In computational science, Matlab is the *de factor* standard, and offers a high-level language and an interactive environment for algorithm development, data visualization, data analysis, and numeric computation. As a powerful scripting programming language, Python also offers a number of advantages for scientific computing. A comparison of Matlab and Python is described in Langtangen (2006). Matlab and Python have indeed many things in common, including no declaration of variables, simple and convenient syntax, easy creation of GUIs, and gluing of simulation and visualization. Nevertheless, Python has some clear advantageous over Matlab and similar environments:

- the Python programming language is more powerful,
- the Python environment is completely open and made for integration with external tools,
- a complete toolbox/module with lots of functions and classes can be contained in a single file (in contrast to a number of M-files),
- transferring functions as arguments to functions is simpler,
- nested, heterogeneous data structures are simple to construct and use,
- object-oriented programming is better developed,
- interfacing C, C++, and Fortran code is better supported and therefore simpler,
- scalar functions work with array arguments to a larger extent (without modifications of arithmetic operators),
- the source is free and runs on more platforms.

However, Matlab has significantly more comprehensive numerical functionality than Python (linear algebra, ODE solvers, optimization, time series analysis, image analysis, etc.). The graphical capabilities of Matlab are also more convenient than Python, since Python graphics relies on external packages that must be installed separately. Some Python programmers have developed interfaces, e.g. `pymat`, `Mlabwrap`, that allows Python programs to use Matlab as a computational and graphics engine. Also, the `matplotlib` library offers high-quality curve plotting with Matlab-like syntax. A good comparison of how to use Python for those with Matlab knowledge is given in the books Kiusalaas (2005*b*) and Kiusalaas (2005*a*)..

2.A.2 Python fundamentals

Installing Python

Simply visit <http://www.python.org/download> to get the most recent version of Python. If you are running Windows XP, the most well-known Python distribution is ActivePython. ActivePython is a free, all-in-one distribution from ActiveState (<http://www.activestate.com>). More Python distributions and integrated development environments (IDEs) are introduced in Lutz & Ascher (2003).

Some useful Python libraries

A large number of extension modules have been developed for Python. A complete list of available modules can be obtained from the PyPI page (<http://www.python.org/pypi>). Here, two packages are used in the `dymosimPy` module:

- Scientific Tools for Python (`SciPy`) is open-source software for mathematics, science, and engineering. The core library is `NumPy` which provides convenient and fast N-dimensional array manipulation. In this work, the functions `loadmat` and `savemat` of the `SciPy` library are used to read and write data of a Matlab `.mat` file.
- `matplotlib` is a python 2D plotting library which produces high-quality curve plotting with Matlab-like syntax.

Defining PATH and PYTHONPATH variables

Sometimes, the user needs to add two important environment variables: `PATH` and `PYTHONPATH`. The operating system searches in the directories contained in the `PATH` variable to find executable files. Similarly, Python searches modules to be imported in the directories contained in the `PYTHONPATH` variable. The standard PythonWin has included the default Python path which can be browsed or edited in the menu `Tools/Edit Python Path`. To use the `dymosimPy` module, you need to add the location of `dymosimPy` into the `PYTHONPATH` variable.

When the user has added the `dymosimPy` module in `PYTHONPATH`, all the functions in the module can be imported using the command:

```
>>> import dymosimPy
```

Debugging in Python

The PythonWin IDE (available with the ActivePython distribution) has a built-in debugger. The debugger supports setting breakpoints and single stepping at the source line level, examining the current context, post-mortem diagnostics, etc.. An introduction to debugging in Python is given on http://www.ferg.org/papers/debugging_in_python.html.

Handling files and text

The `dymosimPy` module directly calls the `dymosim` executable that uses files to define its inputs and outputs. In Python, the basic steps of handling files are as follows:

1. Open the input or output files

A file is opened by the `open` function, taking the filename as first argument and a read/write indication as second argument:

```
infile = open(infile, 'r')    # open file for reading
ofile = open(outfile, 'w')    # open file for writing
```

The second parameter is optional:

```
'r'    read data from the file
'w'    write data to the file (an existing file with the same name will be replaced)
'a'    write data to the file, appending to the end
'r+'   read from and write to (erasing any previous data) the file
'r+a'  read from and write to (appending) the file
'b'    read or write binary data
```

The `open` function creates a Python file object that you use for reading from or writing to a file.

2. Read from the input file and write to the output file

Here are some basic Python methods for file reading and writing:

```
infile.read()           read a file's whole content
infile.readline()      read a single line from the file
infile.readlines()     return a list containing all the lines of data in the file
ofile.write(string)    write the contents of string to the file
ofile.writelines(list) write a list of lines to the file
```

3. Close the files

The files are closed at the end with the `close()` method:

```
infile.close(); ofile.close()
```

When working with binary files, strings, numbers etc. must be converted into sequences of bytes. To encode/decode binary data, Python provides a module called `struct`. Text searching and manipulation can often be greatly simplified by utilizing regular expressions which is available through the `re` module. A basic reference for regular expression syntax in Python is the “Regular Expression HOWTO” (<http://www.amk.ca/python/howto/regex>).

2.A.3 The `dymosimPy` module

Principle of Dymola-Python interface

The Dymola-Python interface is built based on the structure of the Dymola-Matlab interface. More details of the Dymola-Matlab interface are given in Dynasim (2004).

The functionality of the `dymosimPy` module is twofold:

Firstly, Dymosim is called as a stand-alone application from Python environment: the experiment description is read from an input file, one simulation run is performed, the result is stored in an output file, and the statistics information is saved in another file.

Secondly, the simulation result is loaded into Python, and the result for a specific simple variable can be obtained.

Dymosim is supposed to be executed in Python by a similar command as in Dymola-Matlab interface with the following Matlab syntax:

```
d = dymosim(exp,x0,p)
```

where the three input arguments correspond to the three matrices specified in a Dymosim input file `dsin.txt`. `exp` is the experiment setup vector, `x0` is the initial state vector, and `p` is the parameter vector. `d` is a structure variable containing the simulation result.

The input file as well as the result file of Dymosim are given in ASCII (human readable) or Matlab binary format. With the executable program `alist` provided by Dymola, the ASCII and binary form can be transformed into each other, e.g.

```
alist -b dsin.txt temp.mat
```

transforms the ASCII format file `dsin.txt` into the Matlab data file `temp.mat`. The option `-b` is used to transform ASCII format to Matlab binary format.

According to the functionality, the Dymola-Python interface is built with the following steps (see Figure 2.8):

1. Load the variables of the experimental description from the Dymosim input file (the default file is `dsin.txt`). `dsin.txt` is a self-explanatory file with general data structure depicted in Table 2.5. The necessary variables required to run `dymosim.exe` are `Aclass`, `experiment`, `initialName`, and `initalValue`.
2. Set the user defined input arguments: define `exp` as the variable `experiment`, and put `x0` and `p` into the variable `initialValue`.
3. Save the modified variables (`Aclass`, `experiment`, `initialName`, `initialValue`) into a new input file, e.g. `dsin_new.txt`.
4. Run `dymosim.exe` with the command line:

```
dymosim.exe -w dslog.txt dsin_new.txt dsres.mat
```

By running this command, Dymosim reads the experiment description from the file `dsin_new.txt`, performs simulation, and stores the simulation results in `dsres.mat` and the statistics information in `dslog.txt`.

5. Load the simulation result from `dsres.mat` and save the result in a dictionary data structure to make it easy to get the simulation result for a specific variable.

Table 2.5: Data structures of Dymosim input file.

Variable name	Size	Description
<code>Aclass</code>	<code>char(3, 24)</code>	General description
<code>experiment</code>	<code>double(7, 1)</code>	Experimental parameters
<code>method</code>	<code>double(27, 1)</code>	Method tuning parameters
<code>settings</code>	<code>int(13, 1)</code>	Output parameters
<code>initialName</code>	<code>char(n, 8)</code>	Name of initial variables
<code>initialValue</code>	<code>double(n, 6)</code>	Matrix of initial value calculation
<code>initialDescription</code>	<code>char(n, 53)</code>	Description of initial variables

n is the number of initial variables.

Functions in the `dymosimPy` module

The functions of the `dymosimPy` module are summarized in Table 2.4. All the functions are developed by the author. Some descriptions of the functions in the `dymosimPy` module are given as follows:

- `dymosim` is the main function of `dymosimPy`. `dymosim` is called by the command

```
d = dymosimPy.dymosim(exp,x0,p,quiet,file)
```

There are two optional input arguments: `quiet` indicating whether simulation messages are displayed and `file` indicating the name of the Dymosim input file. The function follows the Dymola-Python communication procedure described in Figure 2.8.

- `dymload` is used to load simulation results from a Matlab data file generated by Dymosim. The output `dymstr` is a dictionary variable containing the following fields:

<code>fname</code>	name of the Matlab data file
<code>nname</code>	the number of variable names
<code>ndatamat</code>	the number of data matrices
<code>name</code>	names of the variables
<code>description</code>	description of the variables
<code>dataInfo</code>	<code>dataInfo</code> array
<code>data_#</code>	data matrices of the variables, <code>#</code> has range from 1 to <code>ndatamat</code>

- After running `dymload`, the simulation result of a specific variable can be obtained by the `dymget` function. For a single variable, a column vector containing the data is returned with the command

```
data = dymosimPy.dymget(dymstr,'name')
```

where `dymstr` is a dictionary variable introduced above and `name` is the full variable name.

- `parload` and `initload` are used to load names and values of model parameters and states from the Dymosim input file.

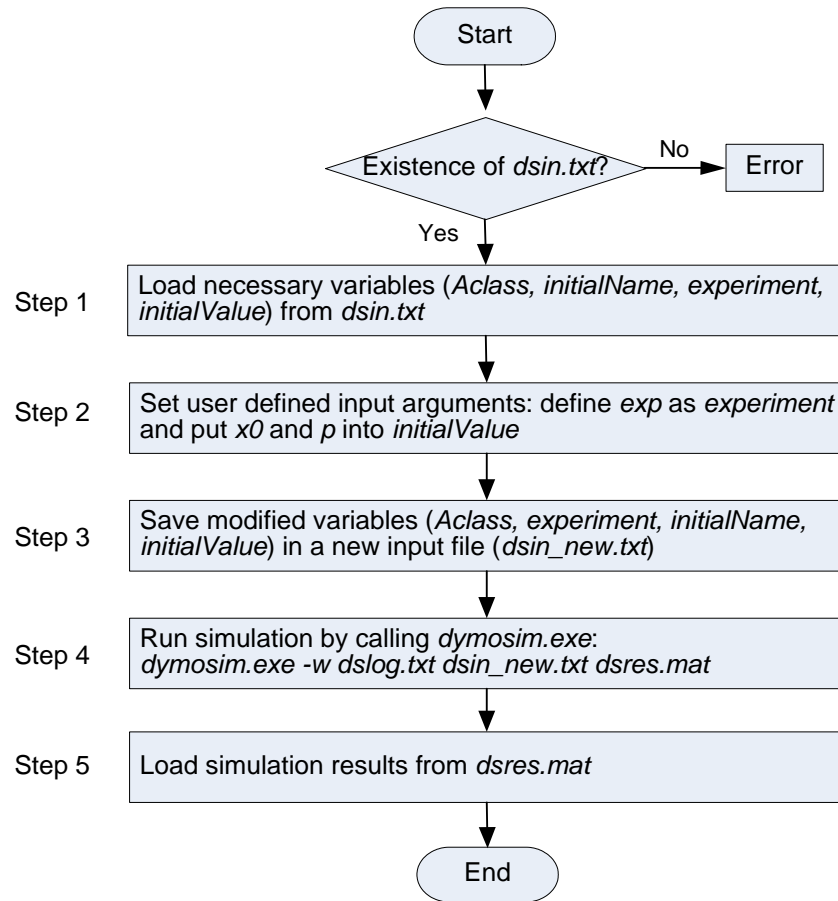


Figure 2.8: Structure of the Dymola-Python interface.

- `tnlist` is used to list a string vector, for instance, names of parameters or states, with preceding row numbers.
- Other functions in `dymosimPy`, e.g. `loadName`, `findIndex`, `setInitialValue`, are useful for loading and changing variables in the Dymosim input file.

Usually, you can get detailed help for a function by the `help` command provided in the `pydoc` module, e.g.:

```
>>> import pydoc
>>> help(dymosimPy.dymosim)
```

2.A.4 Applications of the `dymosimPy` module

Two examples are given below to demonstrate applications of the `dymosimPy` module.

Example 1 (Simulation of a CSTR system model) Consider the classical continuous stirred tank reactor (CSTR) for an exothermic, irreversible reaction, $A \rightarrow B$. In Henson &

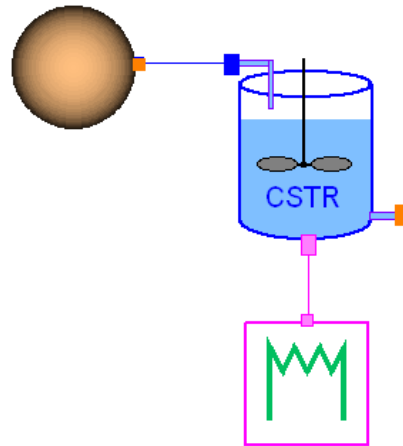


Figure 2.9: Model diagram of a CSTR system implemented in Modelica/Dymola.

Seborg (1997), the following dynamic model is derived based on the component balance for reactant A and the energy balance :

$$\begin{aligned}\frac{dC_A}{dt} &= \frac{q}{V} (C_{Af} - C_A) - k_0 \exp\left(-\frac{E}{RT}\right) C_A \\ \frac{dT}{dt} &= \frac{q}{V} (T_f - T) + \frac{(-\Delta H)}{\rho C_p} k_0 \exp\left(-\frac{E}{RT}\right) C_A + \frac{UA}{\rho V C_p} (T_c - T)\end{aligned}$$

where C_A is the concentration of A in the reactor, T is the reactor temperature, and T_c is the temperature of the coolant stream. The CSTR system is implemented in Modelica/Dymola, see Figure 2.9. The Modelica model is then simulated in the Python environment using the `dymosimPy` module. The evolution of the reactor temperature T is shown in Figure 2.10 by manipulating the temperature of the coolant stream T_c . The Python code appears as follows.

```
# Version 1.0, 2006-07-18
# Written by Qian Chai, Telemark University College

import dymosimPy
from pylab import *      # import matplotlib library

# Define path and name of the Dymosim input file
path = "D:\\Python_Dymola\\testCSTRmodel\\" # \\ denotes \ in Python
file = path + "dsincstr.txt"
# Define experiment parameters
exp = [0, 10, 0, 500, 1e-4, 0, 8]
# Define model parameters
p, pname = dymosimPy.parload(file)
##dymosimPy.tnlist(pname)
```

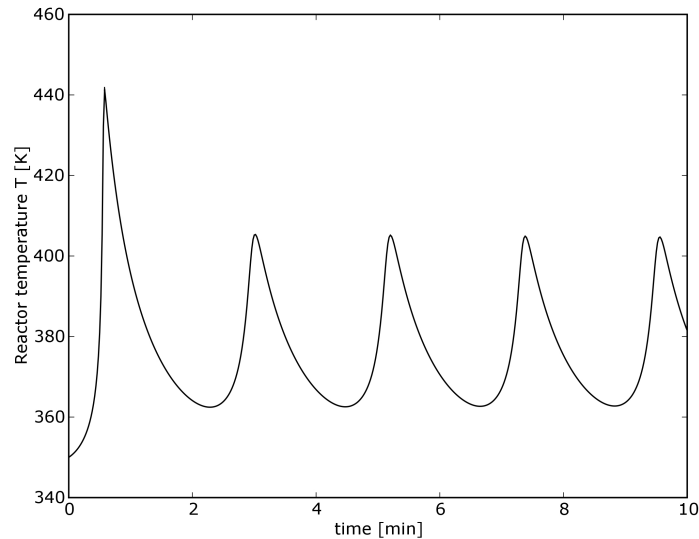


Figure 2.10: Simulation result of the CSTR system, plotted with the `matplotlib` library in Python.

```
p[-5] = 305 # the 5th item from the end of p is the manipulated input
# Define initial state
x0, xname = dyMosimPy.initload(file)
# Call the dyMosim function
quiet = 1
d = dyMosimPy.dyMosim(exp, x0, p, quiet, file)
# Get simulation results for specific variables
time = dyMosimPy.dymget(d, 'Time')
T = dyMosimPy.dymget(d, 'CSTR1.RRate.T')
# Plot results using the matplotlib library
plot(time, T, 'k')
xlabel('time [min]')
ylabel('Reactor temperature T [K]')
show()
```

Example 2 (Simulation of a nitrogen removal process) *Now, we use the `dyMosimPy` module to simulate the N-removal process presented in Paper E. Figure 2.11 shows the simulation result of the concentration of total nitrogen TN when the optimal aeration strategy is applied, i.e., 16 aeration cycles each day. This result is as same as the one we get in the Matlab simulation environment (see Figure 6 in Paper E). The Python program is given as follows:*

```
# Version 1.0, 2006-07-18
# Written by Qian Chai, Telemark University College
```

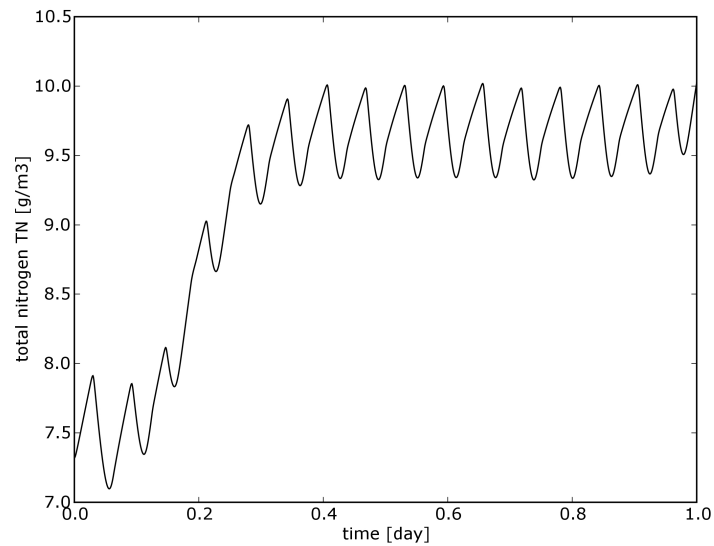


Figure 2.11: Time evolution of the concentration of TN with 16 aeration cycles each day.

```

import dyMosimPy
from pylab import *
import pydoc

help(dyMosimPy.dyMosim) # general help
# Define path and the name of DyMosim input file
path = "D:\\Conference\\Tools\\Python_Dymola\\testSIMS06model\\"
file = path + "dsin.txt"
# Give user-defined input arguments
step = 1.0/(60.0*24.0)
exp = [0,1,step,0,1E-4,0,8]
p, pname = dyMosimPy.parload(file)
dyMosimPy.tnlist(pname)
p[1] = 1.0/16.0 # define length of aeration cycle
x0, xname = dyMosimPy.initload(file)
# Call the dyMosim function
d = dyMosimPy.dyMosim(exp, x0, p)
# Get simulation results
time = dyMosimPy.dymget(d,'Time')
TNeff = dyMosimPy.dymget(d,'eff_TN.TN')
# Plot
plot(time, TNeff, 'k')
xlabel('time [day]')
ylabel('total nitrogen TN [g/m3]')

```

show()

2.A.5 Conclusions

A Python module, named `dymosimPy`, is developed by the author to build up a convenient communication between Modelica/Dymola and Python. `dymosimPy` allows the user to simulate Modelica models and load simulation results in the Python environment. However, the functions included in the module are quite basic and limited. In order to carry out more advanced analysis (e.g. linearization, parameter estimation), additional useful functions should be supplemented to the `dymosimPy` module.

Acknowledgement 3 *The author would like to acknowledge the valuable suggestions from Dr. V. Siepmann of StatoilHydro, Porsgrunn, Norway on handling files in Python.*

Chapter 3

State Estimation

3.1 Introduction

This chapter provides an overview of some commonly applied techniques for state estimation of linear, nonlinear, and constrained systems. The following techniques are discussed: *standard Kalman filter (KF)*, *extended Kalman filter (EKF)*, *unscented Kalman filter (UKF)*, *square-root unscented Kalman filter (SRUKF)*, and *moving horizon estimator (MHE)*. Because of the wide scope of the state estimation problem, the attention will be focused on systematic and exact presentation of algorithms in order to make implementation of the techniques more efficient. The N-removal process described in subsection 2.2.2 is used as case study.

To formulate state estimation as a mathematical problem, we consider a dynamic system of the general form

$$x_k = f(x_{k-1}, u_{k-1}, w_{k-1}) \quad (3.1)$$

$$y_k = h(x_k, v_k), \quad (3.2)$$

where $x_k \in \mathbb{R}^{n_x}$ is the state, $u_k \in \mathbb{R}^{n_u}$ is the control input, and $y_k \in \mathbb{R}^{n_y}$ is the controlled output. $w_k \in \mathbb{R}^{n_w}$ and $v_k \in \mathbb{R}^{n_v}$ represent the process noise and measurement noise, respectively. They are assumed to be zero-mean Gaussian noise processes

$$\begin{aligned} w_k &\sim \mathcal{N}(0, Q_k) \text{ and} \\ v_k &\sim \mathcal{N}(0, R_k). \end{aligned}$$

$f(x_{k-1}, u_{k-1}, w_{k-1})$ is the (possibly nonlinear) state transition function and $h(x_k, v_k)$ is the (possibly nonlinear) measurement function. The state estimation problem is to determine an estimate of the state \hat{x}_k given the chosen model structure and a sequence of noisy observations (measurements) $\{y_0, \dots, y_k\}$.

The standard KF is the algorithm of choice in state estimation for linear systems. Nonlinear filtering is more difficult and complex and there are still many research activities for advances and improvement in nonlinear estimation techniques. However, some nonlinear estimation methods have become widespread. In Simon (2006) an overview of the nonlinear estimation techniques, which include nonlinear extensions of the standard KF, unscented

filtering, and particle filtering, has been given. The EKF is undoubtedly the most widely used nonlinear estimation technique that has been applied in the past few decades. Recently, new variants of the Kalman filter which approximate the nonlinearity to a higher order have been proposed. Among them, the UKF and SRUKF generates a population of so-called *sigma points* based on the current mean and covariance of the state, and permits the direct propagation of the mean and covariance through the actual nonlinear system (Julier et al. 1995), (van der Merwe & Wan 2001). These variants of the Kalman filter are described in a recursion similar in structure to the standard KF. All members of this family of filters will comply with a structured sequence of 6 steps per iteration, as outlined in Plett (2006a):

1. *State estimate time update.* The first step is to compute a prediction $\hat{x}_{k|k-1}$ of the current x_k based on *a priori* information $\{y_0, \dots, y_{k-1}\}$ and the system model.
2. *State error covariance time update.* The second step is to calculate the predicted state estimate error covariance $P_{k|k-1}$ based on *a priori* information and the system model.
3. *Estimate system output $\hat{y}_{k|k-1}$.* The third step is to estimate the system output using the *a priori* state estimate and Equation 3.2.
4. *Estimator gain matrix K_k .* The fourth step is to determine the estimator gain matrix by evaluating $K_k = P_{xy,k}P_{y,k}^{-1}$, where P_{xy} represents the cross correlation matrix between $\hat{x}_{k|k-1}$ and $\hat{y}_{k|k-1}$, and P_y represents the innovation covariance.
5. *State estimation measurement update.* The fifth step is to compute the *a posteriori* state estimate by updating the *a priori* estimate using the estimator gain and the output prediction error $y_k - \hat{y}_{k|k-1}$, i.e., $\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k (y_k - \hat{y}_{k|k-1})$.
6. *State error covariance measurement update.* The final step computes the *a posteriori* state error covariance $P_{k|k}$.

Figure 3.1 illustrates the concept of a state estimator using the Kalman filter-based techniques.

Although the aforementioned Kalman filter-based techniques are widely used for estimating states of unconstrained systems subject to normally distributed state and measurement noises, many physical systems have states subject to hard constraints, such as non-negative concentrations or pressures. Hence, Kalman filtering is no longer directly applicable. One solution for determining an optimal estimate is to formulate the estimation problem as the minimization of a weighted quadratic function of the unknown variables subject to the model equations. This formulation allows for the natural addition of inequality constraints. The basic form of the estimation problem is non-recursive, which is easily understandable, but the burden of computing the optimal estimates will become prohibitively large when more and more measurements become available. It is thus necessary to introduce some kind of approximation or simplification in the estimation procedure. Based on the idea of repetitive optimal control as used in predictive control (Sunan et al. 2002), (Qin & Badgwell 2003), (Lie et al. 2005), moving horizon estimation is suggested as a practical approach to implement optimization-based estimation.

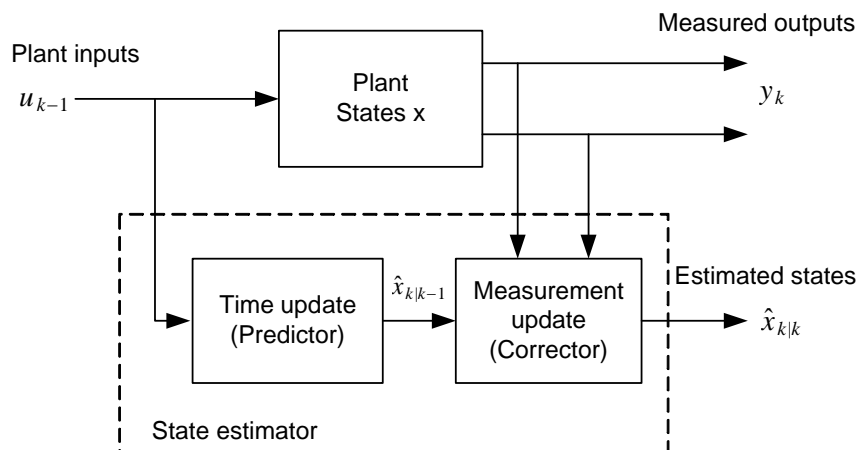


Figure 3.1: Principle of state estimation.

The design and application of state observers, also called *software sensors*¹, in WWTPs has been an active area over the past decades. The approaches that are relevant to the field of WWTPs can be categorized into three groups:

- Classical observers (e.g. Kalman filter) are based on the perfect knowledge of the system model and parameters. The EKF has been widely applied in wastewater engineering, e.g. Lindberg (1997), Boaventura et al. (2001), Keesman (2002), Lukasse, Keesman & van Straten (1999b).
- Asymptotic observers are used to make the model independent of the process kinetics (Perrier et al. 2000), (Dochain 2003). The asymptotic observer requires that the number of measurements is larger than or equal to the number of process reactions.
- Interval observers can be used for predicting intervals in which the unmeasured variables are guaranteed to belong, if bounds are known for the uncertain inputs and/or parameters (Hadj-Sadok & Gouzé 2001), (Alcaraz-González et al. 2002).

3.2 State estimation techniques

3.2.1 Observability and detectability

Observability is a critical property of a system for observer design. An important consequence of the state observability is the ability to reconstruct the time evolution of the state variables from measured variables in an arbitrary finite time from any initial conditions. In the following, it is assumed that the model is identical to the true system.

¹In the scientific literature, the expression of *software sensors* has largely been used for state and parameter estimation. In fact, a software sensor can be defined as an algorithm using state estimation techniques to reconstruct time evolution of the unmeasured states or unknown parameters on the basis of the knowledge of the process dynamics (the dynamic model) and on the (few) available on-line measurements (Dochain 2003).

For simplicity, we shall consider the notion of observability for the linear time invariant system which is described as

$$\begin{aligned}x_k &= Fx_{k-1} \\ y_k &= Hx_k.\end{aligned}$$

Next we address concepts related to observability, with reference to Chen (1999), Rugh (1996), and Simon (2006).

Definition 4 (Observability) *The discrete-time system or the pair (F, H) is said to be observable if for any unknown initial state x_0 , there exists a finite integer $k > 0$ such that knowledge of the input u_i and output y_i for all $i \in [0, k]$ suffices to determine uniquely the initial state x_0 .*

Definition 5 (Detectability) *The discrete-time system or the pair (F, H) is said to be detectable if there exists a matrix M such that*

$$x_k = (F - MH)x_{k-1}$$

is exponentially stable.

The pair (F, H) is detectable if the unobservable part in the observable canonical decomposition is stable, i.e., has eigenvalues either in the left half of the complex plane or inside the unit disk, depending on whether we are dealing with a continuous- or a discrete-time system. If the system is observable, then an observer gain can be found so that the estimated \hat{x} converges to the true state x arbitrarily fast. If the system is detectable, then the convergence of the observer can be guaranteed but the dynamics of the observer can not be assigned arbitrarily.

It is well known that the linear system is observable if and only if the *observability matrix*

$$\mathcal{O} = \begin{pmatrix} H \\ HF \\ \vdots \\ HF^{n_x-1} \end{pmatrix}$$

has full column rank, i.e., $\text{rank}(\mathcal{O}) = n_x$. Another equivalent test for observability is the *PBH test*.

Theorem 6 (PBH observability test) *The pair (F, H) is observable if and only if*

$$\text{rank} \begin{pmatrix} H \\ \lambda I - F \end{pmatrix}$$

has full rank at every eigenvalue λ of F .

For nonlinear systems, the observability analysis is rather complex, and some approaches are introduced in e.g. Sontag (1998), Kurtz & Henson (1998), Dochain & Chen (1992), Bogaerts & Wouwer (2004).

3.2.2 The standard Kalman filter (KF)

We consider a system that can be represented with a linear time invariant model

$$\begin{aligned}x_k &= Fx_{k-1} + Bu_{k-1} + Gw_{k-1} \\y_k &= Hx_k + Lv_k.\end{aligned}$$

Then, the algorithm of the standard KF for state estimation is presented in Table 3.1.

Table 3.1: Algorithm of the standard Kalman filter.

Linear state-space model	
$x_k = Fx_{k-1} + Bu_{k-1} + Gw_{k-1}$	
$y_k = Hx_k + Lv_k$, where $w_k \sim \mathcal{N}(0, Q_k)$ and $v_k \sim \mathcal{N}(0, R_k)$.	
Initialization: for $k = 0$, set $x_0 \sim \mathcal{N}(\hat{x}_{0 0}, P_{0 0})$	
Computation: for $k = 1, 2, \dots$, compute	
State estimate time update	$\hat{x}_{k k-1} = F\hat{x}_{k-1 k-1} + Bu_{k-1}$
Error covariance time update	$P_{k k-1} = FP_{k-1 k-1}F^T + GQ_{k-1}G^T$
Output estimate	$\hat{y}_{k k-1} = H\hat{x}_{k k-1}$
Kalman gain	$K_k = P_{k k-1}H^T (HP_{k k-1}H^T + LR_kL^T)^{-1}$
State estimate meas. update	$\hat{x}_{k k} = \hat{x}_{k k-1} + K_k(y_k - \hat{y}_{k k-1})$
Error covariance meas. update	$P_{k k} = (I - K_kH)P_{k k-1}(I - K_kH)^T + K_kLR_kL^TK_k^T$

3.2.3 The extended Kalman filter (EKF)

The EKF linearizes the nonlinear system around the Kalman filter estimate, and the Kalman filter estimate is based on the linearized system. The nonlinear model is used for computation of state estimate time update $\hat{x}_{k|k-1}$ and output estimate $\hat{y}_{k|k-1}$. To reduce the linearization error in the EKF for highly nonlinear systems, some refined linearization techniques such as the iterated EKF and the second-order EKF, have been described in Simon (2006) and Crassidis & Junkins (2004).

The discrete-time EKF, which considers systems with discrete-time dynamics and discrete-time measurements, is given in Table 3.2.

However, in practice, many engineering systems are governed by continuous-time dynamics whereas the measurements are obtained at discrete instants of time. We therefore consider a hybrid system as follows

$$\begin{aligned}\frac{dx}{dt} &= \check{f}(x, u, w) \\y_k &= h(x_k, v_k) \\w(t) &\sim \mathcal{N}(0, Q) \\v_k &\sim \mathcal{N}(0, R_k).\end{aligned}$$

In steps 1 and 2 of the algorithm of the hybrid EKF, the state estimate and its covariance from time $k - 1$ to time k are integrated as

$$\begin{aligned}\frac{d\hat{x}}{dt} &= \check{f}(\hat{x}, u, 0) \\ \frac{dP}{dt} &= FP + PF^T + GQG^T,\end{aligned}$$

where $F = \left. \frac{\partial \check{f}}{\partial x} \right|_{\hat{x}, u, 0}$ and $G = \left. \frac{\partial \check{f}}{\partial w} \right|_{\hat{x}, u, 0}$. We begin the integration process with $\hat{x} = \hat{x}_{k-1|k-1}$ and $P = P_{k-1|k-1}$. At the end of the integration we have $\hat{x} = \hat{x}_{k|k-1}$ and $P = P_{k|k-1}$. It should be noticed that the covariance matrix Q of the continuous-time process noise is not identical but related to the discrete-time process noise covariance Q_k . An approximation of the relationship can be described in the form of: $Q_k = (t_{k+1} - t_k) Q$, see Grewal & Andrews (2001) and Crassidis & Junkins (2004). At each measurement time, we update the state estimate and the covariance as derived in the discrete-time EKF (steps 3–6).

Table 3.2: Algorithm of the discrete-time extended Kalman filter.

Nonlinear state-space model	
$x_k = f(x_{k-1}, u_{k-1}, w_{k-1})$	
$y_k = h(x_k, v_k)$, where $w_k \sim \mathcal{N}(0, Q_k)$ and $v_k \sim \mathcal{N}(0, R_k)$.	
Definition: $F_k = \left. \frac{\partial f}{\partial x} \right _{\hat{x}_{k k}, u_k, 0}$, $G_k = \left. \frac{\partial f}{\partial w} \right _{\hat{x}_{k k}, u_k, 0}$, $H_k = \left. \frac{\partial h}{\partial x} \right _{\hat{x}_{k k-1}, u_k, 0}$, $L_k = \left. \frac{\partial h}{\partial v} \right _{\hat{x}_{k k-1}, u_k, 0}$	
Initialization: for $k = 0$, set $x_0 \sim \mathcal{N}(\hat{x}_{0 0}, P_{0 0})$	
Computation: for $k = 1, 2, \dots$, compute	
State estimate time update	$\hat{x}_{k k-1} = f(\hat{x}_{k-1 k-1}, u_{k-1}, 0)$
Error covariance time update	$P_{k k-1} = F_{k-1}P_{k-1 k-1}F_{k-1}^T + G_{k-1}Q_{k-1}G_{k-1}^T$
Output estimate	$\hat{y}_{k k-1} = h(\hat{x}_{k k-1}, 0)$
Estimator gain	$K_k = P_{k k-1}H_k^T (H_kP_{k k-1}H_k^T + L_kR_kL_k^T)^{-1}$
State estimate meas. update	$\hat{x}_{k k} = \hat{x}_{k k-1} + K_k(y_k - \hat{y}_{k k-1})$
Error covariance meas. update	$P_{k k} = (I - K_kH_k)P_{k k-1}(I - K_kH_k)^T + K_kL_kR_kL_k^TK_k^T$

3.2.4 The unscented Kalman filter (UKF)

The UKF is an extension of the traditional Kalman filter for the estimation of non-linear systems (Julier et al. 1995). Different from the EKF, the UKF approximates probability distributions by use of the unscented transform. The UKF uses the so-called *sigma points* and propagates all of them through the system model to numerically estimate the state and the output as well as their covariances. The weights of the sigma points for calculating mean and covariances are defined in Table 3.3. The formulations of the UKF is given in Table 3.4, where $\text{chol}(\cdot)$ denotes Cholesky factorization of a matrix, $\text{diag}(\cdot)$ denotes a block diagonal matrix constructed from input arguments, and A_i denotes column i of matrix A . See Julier & Uhlmann (2004), Wan & van der Merwe (2000), and VanDyke et al. (2004) for more details on the algorithm. A recent comparative study of the EKF and the UKF shows

Table 3.3: Weighting constants of the UKF.

γ	Λ_m^0	Λ_m^i	Λ_c^0	Λ_c^i
$\sqrt{n_{\tilde{x}} + \lambda}$	$\frac{\lambda}{n_{\tilde{x}} + \lambda}$	$\frac{\lambda}{2(n_{\tilde{x}} + \lambda)}$	$\frac{\lambda}{n_{\tilde{x}} + \lambda} + (1 - \alpha^2 + \beta)$	$\frac{\lambda}{2(n_{\tilde{x}} + \lambda)}$

$n_{\tilde{x}}$ is the size of the augmented state $\tilde{x}_{k|k}$ (defined in Table 3.4), i.e., $n_{\tilde{x}} = n_x + n_w + n_v$.

$\lambda = \alpha^2 (n_{\tilde{x}} + \kappa) - n_{\tilde{x}}$ is a scaling parameter (typically $\alpha \in [10^{-4}, 1]$ and $\kappa = 0$).

β is used to incorporate prior knowledge of the distribution (with Gaussian distribution $\beta = 2$).

that the estimation performance of the UKF is better than the EKF in terms of robustness and speed of convergence, due to the increased time-update accuracy and the improved covariance accuracy of the UKF, see Kandepe et al. (2008).

3.2.5 The square-root unscented Kalman filter (SRUKF)

The unscented Kalman filter may be used directly for state estimation, but a major drawback of the UKF is its vulnerability for numerical instability. In particular, the state estimation error covariance may lose its positive definiteness in the update step of the UKF. To overcome this flaw, van der Merwe & Wan (2001) introduced the square-root unscented Kalman filter (SRUKF). This approach has mainly two advantages:

1. The state covariances are guaranteed to be positive semi-definite because of the implementation.
2. Compared with the UKF, the SRUKF has an equal computational complexity for state estimation and a lower computational complexity for parameter estimation.

Three important linear algebra techniques are used to implement the SRUKF (van der Merwe & Wan 2001), (Plett 2006b):

- *QR decomposition.* The QR decomposition algorithm computes two factors $Q \in \mathbb{R}^{N \times N}$ and $R \in \mathbb{R}^{L \times N}$ for a matrix $A \in \mathbb{R}^{L \times N}$ such that $A = QR$ where Q is orthogonal, R is upper-triangular, and $L \leq N$. An important property of the QR factorization is that R is related to the Cholesky factor we wish to find. Specifically, if $\tilde{R} \in \mathbb{R}^{L \times L}$ is the upper triangular portion of R , then \tilde{R}^T is the Cholesky factor of $P = A^T A$. That is, if $\tilde{R} = \text{qr}(A^T)^T$, where $\text{qr}(\cdot)$ performs the QR decomposition and returns the upper-triangular portion of R , then $\tilde{R} = \text{chol}(AA^T)$. Thus, the time update of state covariance can be calculated using a QR decomposition of the compound matrix $\sqrt{\Lambda_c^i} \left(x_{k|k-1}^{0:2n_{\tilde{x}}} - \hat{x}_{k|k-1} \right)^T$ for $i = 0, 1, \dots, 2n_{\tilde{x}}$, where $x_{k|k-1}^{0:2n_{\tilde{x}}}$ denotes the propagated sigma points with index from 0 to $2n_{\tilde{x}}$.
- *Cholesky downdating.* To use the previous method, the Cholesky downdate, which is denoted as $\text{cholupdate}(\cdot)$, is necessary since the weight Λ_c^0 may be negative. In addition, the Cholesky downdate is used in the final step of the SRUKF. If $K_k S_y$ is a matrix not a vector, downdating is consecutively done using each column of $K_k S_y$.

Table 3.4: Algorithm of the unscented Kalman filter.

Nonlinear state-space model	
$x_k = f(x_{k-1}, u_{k-1}, w_{k-1})$	
$y_k = h(x_k, v_k)$, where $w_k \sim \mathcal{N}(0, Q_k)$ and $v_k \sim \mathcal{N}(0, R_k)$.	
Definition: $\tilde{x}_{k k}^i = \left[\left(x_{k k}^i \right)^T, \left(w_{k k}^i \right)^T, \left(v_{k k}^i \right)^T \right]^T$	
Initialization: for $k = 0$, set	
$x_0 \sim \mathcal{N}(\hat{x}_{0 0}, P_{0 0})$,	
$\tilde{x}_{0 0} = [\hat{x}_{0 0}^T, 0, 0]^T$, $\tilde{P}_{0 0} = \text{diag}(P_{0 0}, Q_0, R_0)$	
Computation: for $k = 1, 2, \dots$, compute	
Sigma points calculation	$\tilde{x}_{k-1 k-1} = \left[\hat{x}_{k-1 k-1}^T, 0_{1 \times n_w}, 0_{1 \times n_v} \right]^T$ $\tilde{P}_{k-1 k-1} = \text{diag}(P_{k-1 k-1}, Q_{k-1}, R_{k-1})$ $\tilde{x}_{k-1 k-1}^0 = \tilde{x}_{k-1 k-1}$ $\tilde{x}_{k-1 k-1}^i = \tilde{x}_{k-1 k-1} + \left(\gamma \cdot \text{chol} \left(\tilde{P}_{k-1 k-1} \right) \right)_i \quad i = 1, 2, \dots, n_{\tilde{x}}$ $\tilde{x}_{k-1 k-1}^i = \tilde{x}_{k-1 k-1} - \left(\gamma \cdot \text{chol} \left(\tilde{P}_{k-1 k-1} \right) \right)_{i-n_{\tilde{x}}} \quad i = n_{\tilde{x}} + 1, \dots, 2n_{\tilde{x}}$
State time update	$x_{k k-1}^i = f \left(x_{k-1 k-1}^i, u_{k-1}, w_{k-1 k-1}^i \right), \quad i = 0, 1, \dots, 2n_{\tilde{x}}$
Output estimate	$y_{k k-1}^i = h \left(x_{k k-1}^i, v_{k-1 k-1}^i \right), \quad i = 0, 1, \dots, 2n_{\tilde{x}}$ $\hat{x}_{k k-1} = \sum_{i=0}^{2n_{\tilde{x}}} \Lambda_m^i x_{k k-1}^i$ $\hat{y}_{k k-1} = \sum_{i=0}^{2n_{\tilde{x}}} \Lambda_m^i y_{k k-1}^i$
Covariance time update	$P_{k k-1} = \sum_{i=0}^{2n_{\tilde{x}}} \Lambda_c^i \left[x_{k k-1}^i - \hat{x}_{k k-1} \right] \left[x_{k k-1}^i - \hat{x}_{k k-1} \right]^T$
Estimator gain	$P_{y,k k-1} = \sum_{i=0}^{2n_{\tilde{x}}} \Lambda_c^i \left[y_{k k-1}^i - \hat{y}_{k k-1} \right] \left[y_{k k-1}^i - \hat{y}_{k k-1} \right]^T$ $P_{xy,k k-1} = \sum_{i=0}^{2n_{\tilde{x}}} \Lambda_c^i \left[x_{k k-1}^i - \hat{x}_{k k-1} \right] \left[y_{k k-1}^i - \hat{y}_{k k-1} \right]^T$ $K_k = P_{xy,k k-1} P_{y,k k-1}^{-1}$
State meas. update	$\hat{x}_{k k} = \hat{x}_{k k-1} + K_k \left(y_k - \hat{y}_{k k-1} \right)$
Covariance meas. update	$P_{k k} = P_{k k-1} - K_k P_{y,k k-1} K_k^T$

- *Back-substitution.* The estimator gain is calculated more efficiently via back-substitution which can be implemented with Matlab's '/' operator.

In detail, the SRUKF algorithm for state estimation is described in Table 3.5.

3.2.6 Moving horizon estimator (MHE)

Moving horizon estimation that is based on an optimization approach, is an efficient method for state estimation of constrained systems. Rawlings & Bakshi (2006) provide a good review of the recent research on moving horizon estimation, with a focus on advantages and disadvantages of different estimation techniques. To address the computational efficiency issues posed by MHE, Darby & Nikolaou (2007) introduces a parametric programming approach for the real-time implementation.

Compared with other estimation techniques, MHE may provide the following advantages:

- The formulation of the optimization-based approach allows for the natural addition of inequality constraints on states, noise, and other variables.
- For nonlinear estimation, the optimization formulation is quite general and has the advantage of using full nonlinear models.
- MHE has the flexibility of incorporating *a priori* information including the state and noise as well as non-Gaussian variable distributions.

Problem statement

x is assumed to be initially normally distributed with expectation $\hat{x}_{1|0}$ and covariance $P_{1|0}$, i.e., $x_1 \sim \mathcal{N}(\hat{x}_{1|0}, P_{1|0})$. $\hat{x}_{1|0}$, $P_{1|0}$, Q_k , and R_k are assumed to be known².

It is assumed that we know the following information \mathcal{I}_k of the system at the time step k :

$$\mathcal{I}_k = \{\hat{x}_{1|0}, y_1, \dots, y_k, u_1, \dots, u_k\},$$

where k is the current time (initial value of k is 1).

The state estimation problem can be stated as:

Given the “guess” of the initial estimate $x_1 \sim \mathcal{N}(\hat{x}_{1|0}, P_{1|0})$ and all of the information up to the current time k contained in the measurement sequence $\{y_0, \dots, y_k\}$, obtain the best estimate of the initial state $\hat{x}_{1|k}$ and the entire state trajectory $\hat{x}_{1|k}, \dots, \hat{x}_{k|k}$ subject to the process model.

²This initial state is different from the previous filters. $\hat{x}_{1|0}$ and $P_{1|0}$ can be obtained from $\hat{x}_{0|0}$ and $P_{0|0}$ using the standard EK (or EKF) time update steps.

Table 3.5: Algorithm of the square-root unscented Kalman filter.

Nonlinear state-space model	
$x_k = f(x_{k-1}, u_{k-1}, w_{k-1})$	
$y_k = h(x_k, v_k)$, where $w_k \sim \mathcal{N}(0, Q_k)$ and $v_k \sim \mathcal{N}(0, R_k)$.	
Definition: $\tilde{x}_{k k}^i = \left[\left(x_{k k}^i \right)^T, \left(w_{k k}^i \right)^T, \left(v_{k k}^i \right)^T \right]^T$	
Initialization: for $k = 0$, set	
$x_0 \sim \mathcal{N}(\hat{x}_{0 0}, P_{0 0})$	
$\tilde{x}_{0 0} = [\hat{x}_{0 0}^T, 0, 0]^T$, $\tilde{S}_{0 0} = \text{diag}(\text{chol}(P_{0 0}), \text{chol}(Q_0), \text{chol}(R_0))$	
Computation: for $k = 1, 2, \dots$, compute	
Sigma points calculation	$\tilde{x}_{k-1 k-1} = \left[\hat{x}_{k-1 k-1}^T, 0_{1 \times n_w}, 0_{1 \times n_v} \right]^T$ $\tilde{S}_{k-1 k-1} = \text{diag}(S_{k-1 k-1}, \text{chol}(Q_{k-1}), \text{chol}(R_{k-1}))$ $\tilde{x}_{k-1 k-1}^0 = \tilde{x}_{k-1 k-1}$ $\tilde{x}_{k-1 k-1}^i = \tilde{x}_{k-1 k-1} + \left(\gamma \tilde{S}_{k-1 k-1} \right)_i, \quad i = 1, 2, \dots, n_{\tilde{x}}$ $\tilde{x}_{k-1 k-1}^i = \tilde{x}_{k-1 k-1} - \left(\gamma \tilde{S}_{k-1 k-1} \right)_{i-n_{\tilde{x}}}, \quad i = n_{\tilde{x}} + 1, n_{\tilde{x}} + 2, \dots, 2n_{\tilde{x}}$
State time update	$x_{k k-1}^i = f\left(x_{k-1 k-1}^i, u_{k-1}, w_{k-1 k-1}^i\right), \quad i = 0, 1, \dots, 2n_{\tilde{x}}$
Output estimate	$y_{k k-1}^i = h\left(x_{k k-1}^i, v_{k-1 k-1}^i\right), \quad i = 0, 1, \dots, 2n_{\tilde{x}}$ $\hat{x}_{k k-1} = \sum_{i=0}^{2n_{\tilde{x}}} \Lambda_m^i x_{k k-1}^i$ $\hat{y}_{k k-1} = \sum_{i=0}^{2n_{\tilde{x}}} \Lambda_m^i y_{k k-1}^i$
Covariance time update	$\check{S}_{k k-1} = \text{qr}\left(\sqrt{\Lambda_c^1} \left(x_{k k-1}^{1:2n_{\tilde{x}}} - \hat{x}_{k k-1}\right)^T\right)^T$ $S_{k k-1} = \text{cholupdate}\left(\check{S}_{k k-1}, \sqrt{ \Lambda_c^0 } \left(x_{k k-1}^0 - \hat{x}_{k k-1}\right), \text{sign}(\Lambda_c^0)\right)$
Estimator gain	$\check{S}_{y,k k-1} = \text{qr}\left(\sqrt{\Lambda_c^1} \left(y_{k k-1}^{1:2n_{\tilde{x}}} - \hat{y}_{k k-1}\right)^T\right)^T$ $S_{y,k k-1} = \text{cholupdate}\left(\check{S}_{y,k k-1}, \sqrt{ \Lambda_c^0 } \left(y_{k k-1}^0 - \hat{y}_{k k-1}\right), \text{sign}(\Lambda_c^0)\right)$ $P_{xy,k k-1} = \sum_{i=0}^{2L} \Lambda_c^i \left[x_{k k-1}^i - \hat{x}_{k k-1}\right] \left[y_{k k-1}^i - \hat{y}_{k k-1}\right]^T$ $K_k = \left(P_{xy,k k-1} / S_{y,k k-1}^T\right) / S_{y,k k-1}$
State meas. update	$\hat{x}_{k k} = \hat{x}_{k k-1} + K_k (y_k - \hat{y}_{k k-1})$
Covariance meas. update	$S_{k k} = \text{cholupdate}(S_{k k-1}, K_k S_{y,k k-1}, -1)$

Full information estimation

In the full information estimation, an estimate of the states can be found by selecting the set $\{\hat{x}_1, \dots, \hat{x}_k\}$ that maximizes the probability density function

$$p(x_1, \dots, x_k | y_1, \dots, y_k)$$

i.e., with mathematical notation

$$\{\hat{x}_1, \dots, \hat{x}_k\} = \arg \max_{x_1, \dots, x_k} p(x_1, \dots, x_k | y_1, \dots, y_k).$$

According to Bayes' rule, the *maximum likelihood optimization* is solved, and the solution can be posed as the following *least squares problem*

$$\min_{x_1, \dots, x_k} \frac{1}{2} (x_1 - \hat{x}_{1|0})^T M_P^1 (x_1 - \hat{x}_{1|0}) + \frac{1}{2} \sum_{i=1}^k v_i^T M_R^i v_i + \frac{1}{2} \sum_{i=1}^{k-1} w_i^T M_Q^i w_i \quad (3.3)$$

subject to

$$\begin{aligned} x_k &= f(x_{k-1}, u_{k-1}, w_{k-1}) \\ y_k &= h(x_k, v_k) \end{aligned}$$

where the weighting matrices are usually chosen as $M_P^1 = P_{1|0}^{-1}$, $M_R^i = R_i^{-1}$, $M_Q^i = Q_i^{-1}$ and $P_{1|0}$, R_i , and Q_i are the covariance matrices which reflect the confidence in the initial state estimate, the measurement noise, and the process noise, respectively. Weighting factors M_R^i and M_Q^i might change with time. Here, we assume they are constants M_R and M_Q . We refer to this problem as the *full information estimator*, because we consider all of the available measurements.

Moving horizon estimation

As the number of measurements grows, the size of the unknown variables increases without bound. In practice, the objective function of the least squares problem at time k can be modified to

$$\min_{x_{k-N}, \dots, x_k} \frac{1}{2} (x_{k-N} - \hat{x}_{k-N})^T M_P^{k-N} (x_{k-N} - \hat{x}_{k-N}) + \frac{1}{2} \sum_{i=k-N}^k v_i^T M_R^i v_i + \frac{1}{2} \sum_{i=k-N}^{k-1} w_i^T M_Q^i w_i,$$

subject to

$$\begin{aligned} x_k &= f(x_{k-1}, u_{k-1}, w_{k-1}) \\ y_k &= h(x_k, v_k), \end{aligned}$$

where N is a constant interval. This optimization problem is therefore called a *moving horizon optimal estimation*. When the number of available measurements is less than the optimization horizon N , one may use an increasing window length until N measurements

are available. Alternatively, when the problem is unconstrained, e.g., an extended Kalman filter may be used to give optimal estimates in the initial phase.

For MHE, one interesting problem is then how to update \hat{x}_{k-N} and M_P^{k-N} . Here, we consider unconstrained systems. In the linear case, the least squares estimation (regardless of horizon size) is equivalent to the standard Kalman filter under certain choices of the weighting matrices. In the nonlinear case, updating of the weighting matrix M_P^{k-N} is more complicated than the linear case; and the following methods to update M_P^{k-N} have been discussed in the literature (Robertson & Lee 1995):

1. Assume that x_{k-N} is uniformly distributed over some interval. Then, $M_P^{k-N} = 0$ and constraints are used to bound the range of x_{k-N} .
2. Assume that the error distribution of the estimates remains approximately constant and M_P^{k-N} is a fixed weighting matrix.
3. Another possibility is based on the probabilistic interpretation of $\hat{x}_{k-N|k-N-1}$ and $(M_P^{k-N})^{-1}$ as the conditional mean and covariance of x_{k-N} . Any of the approximate nonlinear filtering techniques which are based on the propagation of the conditional mean and covariance (e.g. EKF) can be used. In this case, the moving estimator with a horizon size of one ($N = 1$) will be equivalent to the EKF. However, when $N > 1$, Haseltine & Rawlings (2005) show that MHE is more robust to poor guesses of the initial state in comparison to the EKF, because it fits the data in the horizon and uses the full nonlinear model.

Constrained state estimation

In practice, we may need to constrain the range of variables representing physical quantities. This gives the least squares estimates that are within the specified bounds, which may not be the case for Kalman filter-based methods. We assume that the state, process noise, and measurement noise satisfy the following constraints

$$\begin{aligned} x^\ell &\leq x_i \leq x^u \text{ for } k-N \leq i \leq k \\ v^\ell &\leq v_i \leq v^u \text{ for } k-N \leq i \leq k \\ w^\ell &\leq w_i \leq w^u \text{ for } k-N \leq i \leq k-1 \end{aligned}$$

where the superscript ℓ and u stand for the lower bound and the upper bound, respectively. Of course, more complicated constraint formulations are possible.

MHE for nonlinear estimation

The MHE problem is usually formulated as the following optimization problem

$$\min_{x_{k-N}, \dots, x_k} \frac{1}{2} (x_{k-N} - \hat{x}_{k-N})^T M_P^{k-N} (x_{k-N} - \hat{x}_{k-N}) + \frac{1}{2} \sum_{i=k-N}^k v_i^T M_R^i v_i + \frac{1}{2} \sum_{i=k-N}^{k-1} w_i^T M_Q^i w_i \quad (3.4)$$

subject to

$$\begin{aligned} x_k &= f(x_{k-1}, u_{k-1}, w_{k-1}) \\ y_k &= h(x_k, v_k) \\ x^\ell &\leq x_i \leq x^u \\ v^\ell &\leq v_i \leq v^u \\ w^\ell &\leq w_i \leq w^u. \end{aligned}$$

Since the nonlinear state estimation technique is dual to the nonlinear MPC techniques, various discretization strategies and solution procedures used in nonlinear MPC are directly applicable to nonlinear estimation as well (Robertson & Lee 1995). In some cases, solving the described full nonlinear optimization problem may be time-consuming and unable to find the global minimum of the criterion. Therefore, a simple suboptimal approximation to the optimal solution of the nonlinear estimation problem described in Equation 3.4 is often applied, see Russell et al. (2000). The approach is to approximate the sequential solution of Equation 3.4 by a succession of linear least squares problems that can be easily solved by the standard quadratic programming (QP) solvers. Appendix 3.A presents the QP formulation for solving linear estimation problems.

3.3 State estimation of the N-removal process

This section presents state estimation of the N-removal process by simulation studies. The objective is to compare performances of the EKF, UKF, SRUKF, and MHE for the nonlinear biological WWTP, where the exact models and the statistical information of the stochastic noises, are available to the filters. In Paper C, the standard KF, EKF, and UKF are presented for state estimation of this process. Since the simulation results show that the prediction of the state with the standard KF is poor due to the high nonlinearity of the process, only the nonlinear estimation techniques are focused on in this section.

3.3.1 Performance indices

To analyze the performance of the filters, the prediction accuracies V_y and V_x are defined by the following RMSE (root mean square error) performance criteria

$$\begin{aligned} V_y &= \sqrt{\frac{1}{N \cdot n_y} (y^N - \hat{y}^N)^T \cdot (y^N - \hat{y}^N)} \\ V_x &= \sqrt{\frac{1}{N \cdot n_x} (x^N - \hat{x}^N)^T \cdot (x^N - \hat{x}^N)} \end{aligned}$$

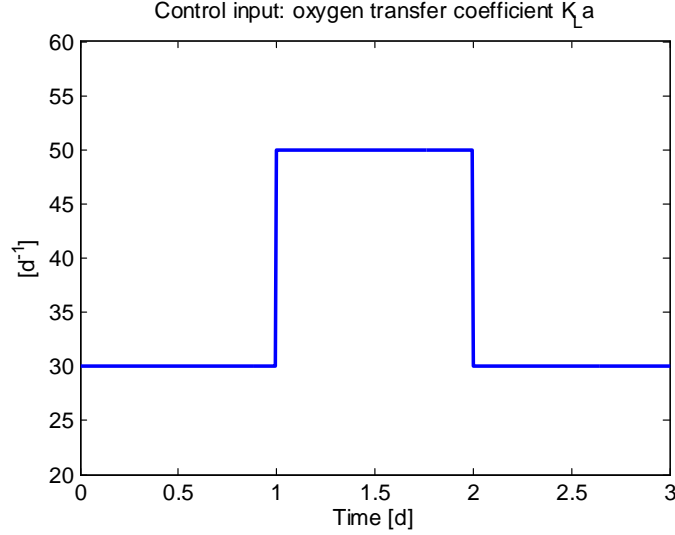


Figure 3.2: Control input for the N-removal process.

where N is the prediction horizon and we define the output sequence y^N , the estimated output sequence \hat{y}^N , the state sequence x^N , and the estimated state sequence \hat{x}^N as

$$\begin{aligned} y^N &\triangleq \{y_N^T, \dots, y_1^T\}^T \\ \hat{y}^N &\triangleq \{\hat{y}_N^T, \dots, \hat{y}_1^T\}^T \\ x^N &\triangleq \{x_N^T, \dots, x_1^T\}^T \\ \hat{x}^N &\triangleq \{\hat{x}_N^T, \dots, \hat{x}_1^T\}^T, \end{aligned}$$

respectively. The criterion V_x can only be evaluated for simulation studies when the true state sequence x^N is known.

3.3.2 Results and discussions

Figures 3.2 and 3.3 show the control input and the simulated output variables used to drive the nonlinear state estimation filters.

To compare the different estimation approaches, the tuning parameters for the EKF, the UKF, and the SRUKF are chosen identical. Three groups of simulations are carried out with same process noise covariances but different measurement noise covariances given as follows:

$$\begin{aligned} Q_k &= \text{diag}(x_o^2 \cdot 10^{-6}) \quad \text{and} \\ R_k &= \begin{cases} \text{diag}(y_o^2 \cdot 10^{-8}), & \text{in Test I} \\ \text{diag}(y_o^2 \cdot 10^{-6}), & \text{in Test II} \\ \text{diag}(y_o^2 \cdot 10^{-4}), & \text{in Test III} \end{cases}, \end{aligned}$$

where (x_o, y_o) is the operating point. For the UKF and SRUKF, the parameter α is also an important tuning parameter, which determines the spread of the sigma points around the current estimate. Here, α is chosen to be 1×10^{-2} which is also used in Paper E.

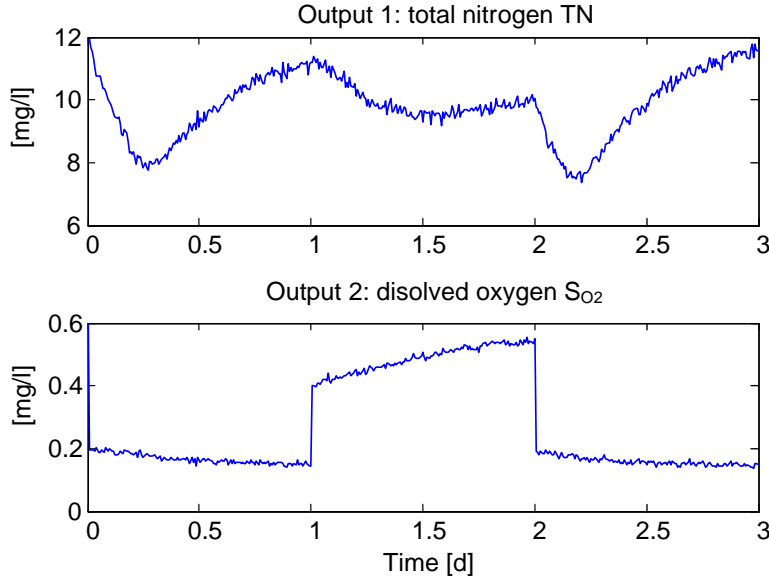


Figure 3.3: Simulated output variables for the N-removal process.

With the increase of the measurement noise covariance, the UKF generates negative state estimates. Thus, a constrained Kalman filtering technique described in Simon & Simon (2003) is applied to project the unconstrained state estimate \hat{x} onto the constraint surface. That is, if we have the constraints $\mathcal{B}x \leq b$, then the constrained estimate can be obtained by solving the problem

$$\hat{x}_{\text{con}} = \arg \min_{\hat{x}_{\text{con}}} (\hat{x}_{\text{con}} - \hat{x})^T W (\hat{x}_{\text{con}} - \hat{x}) \text{ such that } \mathcal{B}\hat{x}_{\text{con}} \leq b,$$

where W is a symmetric positive definite weighting matrix and W is set $W = I$ in this work. This inequality constrained problem can be formulated with a QP problem described in 3.A.2 and solved by the QP function in Matlab. The approach of incorporating state constraints into the UKF by projecting the sigma points which are outside the feasible region onto the constraint surface, is also described in Kandepu et al. (2008).

For comparison, the root squared errors (RSEs)³ of the states in Test I are shown in Figures 3.4–3.7, where four states that most clearly exhibit the difference of the nonlinear filters are chosen. With the UKF, the estimation results are quite noisy because of the numerical instability of the UKF. Due to the improved numerical properties, the SRUKF is more robust than the UKF.

To give an overview of the filter performance, the estimation results of the various filters are summarized in Table 3.6. All computations are performed on a 1.7 GHz Pentium M computer with 1 Gbyte RAM. Table 3.7 shows that the discrete-time EKF is the fastest filter. For instance, the discrete-time EKF requires about 38% of the computational burden of the unscented filters, because the calculation of the predictive sigma points in the unscented filters is time-consuming.

³The root squared error (RSE) of the state x at time t is defined as: $\text{RSE}_t = \sqrt{(x_t - \hat{x}_t)^2}$, where x_t is the actual value at time t and \hat{x}_t is the estimated value at time t .

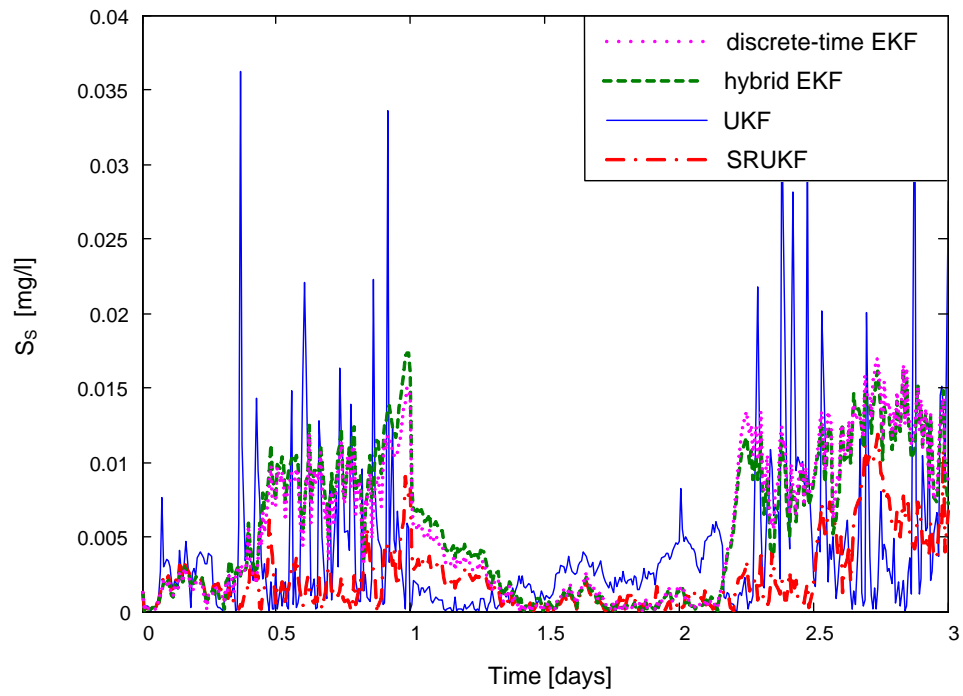


Figure 3.4: Comparison of the estimation error of the concentration of S_S in Test I.

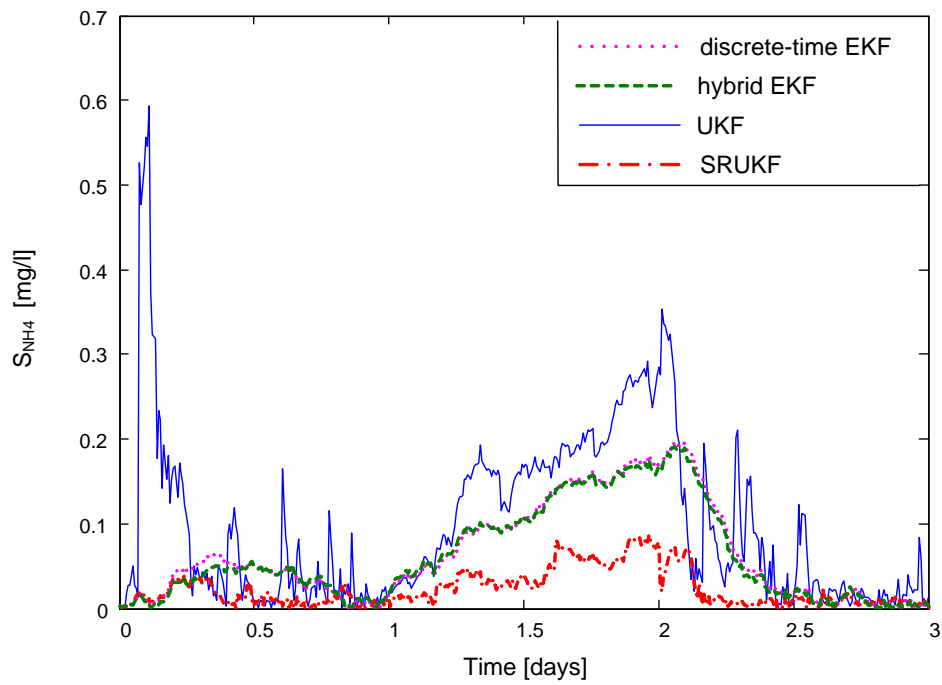


Figure 3.5: Comparison of the estimation error of the concentration of S_{NH_4} in Test I.

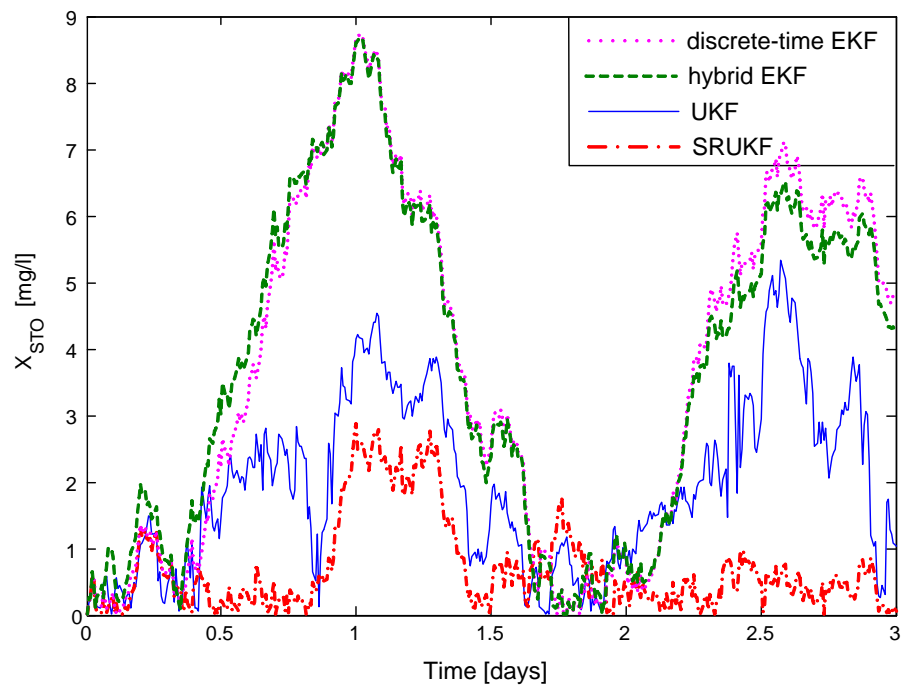


Figure 3.6: Comparison of the estimation error of the concentration of X_{STO} in Test I.

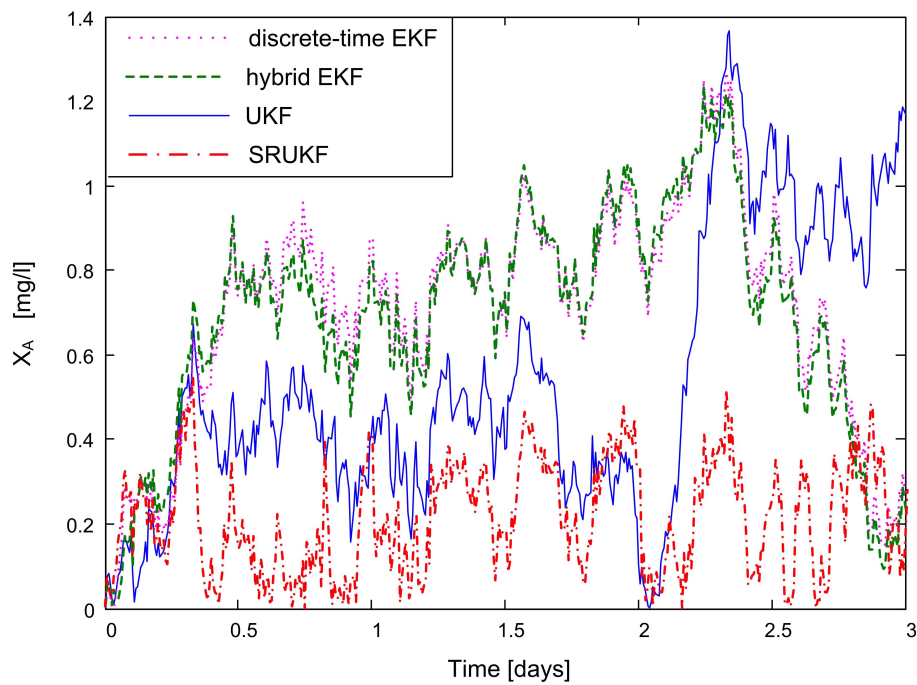


Figure 3.7: Comparison of the estimation error of the concentration of X_A in Test I.

Table 3.6: Performance results of the nonlinear estimation filters.

	Method	V_y	V_x
<i>Test I</i>	Discrete-time EKF	5.3230×10^{-4}	1.7368
	Hybrid EKF	5.3188×10^{-4}	1.7300
	UKF	7.5458×10^{-4}	0.9411
	SRUKF	1.7526×10^{-3}	0.3867
	MHE (filter horizon $N = 1$)	5.3230×10^{-4}	1.7368
<i>Test II</i>	Discrete-time EKF	6.9217×10^{-3}	1.7300
	Hybrid EKF	6.8751×10^{-3}	1.7230
	UKF	5.7945×10^{-3}	1.3613
	SRUKF	6.3563×10^{-3}	0.4759
<i>Test III</i>	Discrete-time EKF	9.4400×10^{-2}	1.4257
	Hybrid EKF	9.4278×10^{-2}	1.4204
	UKF(constrained)	1.6680×10^{-2}	2.5377
	SRUKF	8.2867×10^{-2}	0.6227

Table 3.7: CPU time for the nonlinear estimation filters.

	Discrete-time EKF	Hybrid EKF	UKF	SRUKF	MHE
CPU time	25 min	34 min	60 min	62 min	28 min

In Table 3.6, the difference between the real measurement and the predicted measurement, the criterion V_1 , is quite small. Therefore, the estimation error of the state, the criterion V_2 , is mainly considered for the comparison of the filters. The performance results demonstrate that the SRUKF produces the lowest estimation error for all the tests. The UKF is better than the extended Kalman filters in reducing the estimation errors in Tests I and II, whereas the performance of the UKF becomes poor when taking state constraints into account (see Test III). In fact, to handle state constraints, the MHE estimator is more straightforward and generally preferred.

In this work, a horizon size of 1 is used for the MHE with the update of the initial state estimate and the covariance based a first-order Taylor expansion (as in the discrete-time EKF). In this case, the MHE produces the identical performance as the discrete-time EKF. For this reason, the MHE result only appears in Test I.

Also, the comparison in Table 3.6 shows that the hybrid EKF provides slightly smaller estimation errors than the discrete-time EKF due to the more accurate time updates of the state estimate and the covariance, but the hybrid EKF takes the more CPU time.

3.A Optimization-based estimation formulated as QP problem

3.A.1 The optimization-based estimation problem

A linear system, approximated in discrete time is given as

$$x_k = Fx_{k-1} + Bu_{k-1} + Gw_{k-1} \quad (3.5)$$

$$y_k = Hx_k + Lv_k. \quad (3.6)$$

We still assume that the known information $\mathcal{I}_k = \{\hat{x}_{1|0}, y_1, \dots, y_k, u_1, \dots, u_k\}$. Here, we will introduce the notations for the measurement sequence $y^k \triangleq \{y_k^T, \dots, y_1^T\}^T$, and the input sequence $u^k \triangleq \{u_k^T, \dots, u_1^T\}^T$.

The estimation problem is then defined as

$$\min_{x_1, \dots, x_k} \frac{1}{2} (x_1 - \hat{x}_{1|0})^T M_P^1 (x_1 - \hat{x}_{1|0}) + \frac{1}{2} \sum_{i=1}^k v_i^T M_R^i v_i + \frac{1}{2} \sum_{i=1}^{k-1} w_i^T M_Q^i w_i \quad (3.7)$$

subject to

$$x_k = Fx_{k-1} + Bu_{k-1} + Gw_{k-1}$$

$$y_k = Hx_k + Lv_k$$

$$x^\ell \leq x_i \leq x^u$$

$$v^\ell \leq v_i \leq v^u$$

$$w^\ell \leq w_i \leq w^u.$$

3.A.2 Standard QP problem

This optimization problem can be posed as a quadratic programming (QP) problem of the form

$$\begin{aligned} \min_{\zeta_k} \mathcal{F}(\zeta) &= \min_{\zeta_k} \frac{1}{2} \zeta_k^T \mathcal{H} \zeta_k + g^T \zeta_k \\ \text{s.t. } \mathcal{A} \zeta_k &= a \\ \mathcal{B} \zeta_k &\leq b \\ \zeta^\ell &\leq \zeta_k \leq \zeta^u. \end{aligned} \quad (3.8)$$

The unknown variable ζ_k is defined as

$$\zeta_k \triangleq \begin{pmatrix} x^k \\ v^k \\ w^{k-1} \end{pmatrix}$$

where

$$\begin{aligned} x^k &\triangleq \{x_k^T, \dots, x_1^T\}^T \\ v^k &\triangleq \{v_k^T, \dots, v_1^T\}^T \\ w^{k-1} &\triangleq \{w_{k-1}^T, \dots, w_1^T\}^T. \end{aligned}$$

$x^k \in \mathbb{R}^{n_x \cdot k \times 1}$; $v^k \in \mathbb{R}^{n_v \cdot k \times 1}$; and $w^{k-1} \in \mathbb{R}^{n_w \cdot (k-1) \times 1}$. Matrix \mathcal{H} and vector g of Equation 3.8 are determined from the requirement that J_t of Equation 3.7 should equal $\mathcal{F}(\zeta)$ in Equation 3.8. The equality condition $\mathcal{A}\zeta_k = a$ contains the model in Equations 3.5 and 3.6. For the optimization problem defined here, inequality $\mathcal{B}\zeta_k \leq b$ is empty, while constraints are contained in ζ^ℓ and ζ^u . Thus, the following matrices result (\otimes denotes Kronecker product):

$$\mathcal{H} = \text{diag} \left(0_{n_x \cdot (k-1)}, M_P^1, I_k \otimes M_R, I_{(k-1)} \otimes M_Q \right), \quad (3.9)$$

$$g = \begin{pmatrix} 0_{n_x \cdot (k-1) \times 1} \\ -M_P^1 \hat{x}_{1|0} \\ 0_{(k \cdot n_v + (k-1) \cdot n_w) \times 1} \end{pmatrix}, \quad (3.10)$$

$$\mathcal{A} = \begin{pmatrix} I_k \otimes H & I_k \otimes L & 0_{k \cdot n_y \times (k-1) \cdot n_w} \\ I_{(k-1) \times k} \otimes I_{n_x} - I_{(k-1) \times k, 1} \otimes F & 0_{(k-1) \cdot n_x \times k \cdot n_v} & -I_{(k-1)} \otimes G \end{pmatrix}, \quad (3.11)$$

$$a = \begin{pmatrix} y^k \\ (I_{(k-1)} \otimes B) \cdot u^{k-1} \end{pmatrix}. \quad (3.12)$$

Next, the bounds are given as

$$\begin{aligned} \zeta^\ell &= \left(I_{k \cdot n_x \times 1} \cdot x^\ell, I_{k \cdot n_v \times 1} \cdot v^\ell, I_{(k-1) \cdot n_w \times 1} \cdot w^\ell \right)^T \\ \zeta^u &= \left(I_{k \cdot n_x \times 1} \cdot x^u, I_{k \cdot n_v \times 1} \cdot v^u, I_{(k-1) \cdot n_w \times 1} \cdot w^u \right)^T. \end{aligned}$$

To solve the QP problem, the QP solver `quadprog` of the Optimization Toolbox in Matlab can be used. The solution is in the form of the estimates

$$\zeta_k^* = \{ \hat{x}_{k|k}^T, \dots, \hat{x}_{1|k}^T, \hat{v}_{k|k}^T, \dots, \hat{v}_{1|k}^T, \hat{w}_{k-1|k}^T, \dots, \hat{w}_{1|k}^T \}^T.$$

Chapter 4

Parameter Estimation

4.1 Introduction

The ASM models provide a standardized set of basis models with nominal parameter values suggested in publications. However, the various plant operation, influent, and sludge characteristics of full-scale WWTPs require some of the model parameters to be adjusted in order to ensure good predictive capability of the model. In practice, the parameter estimation of the activated sludge plant models are characterized by the following features:

- The ASM models have many parameters, as compared to the number of available measurements, which makes parameters non-identifiable or poorly identifiable.
- The outputs are insensitive to the variation of many (possibly nonlinear) combinations of parameters.
- The numerical algorithms that perform the nonlinear parameter estimation suffer from convergence problems with large numbers of parameters, due to many local optima.
- Gathering good experimental data from the full-scale plant is difficult, mainly because of lack of cheap and reliable sensors and techniques for measurement.

Clearly, the quality of parameter estimation is directly related to the practical identifiability of parameters, which refers to the ability to obtain accurate parameter estimates from available experimental data. Therefore, the key question of the identifiability analysis is defined by Dochain et al. (1995) as follows:

Assume that a certain number of the state variables are available for measurement; on the basis of the type and quality of available data, can we expect to give a unique value to the model parameters via parameter estimation?

The identifiability analysis performed prior to the parameter estimation can provide answers to the key question. Parameter identifiability analysis of activated sludge process models has been studied in e.g. Vanrolleghem et al. (1995), Weijers & Vanrolleghem (1997), Lukasse et al. (1997), Brouwer et al. (1997). A good overview of identifiability of ASM1

kinetic parameters and experimental designs is given in Petersen (2000). Recently, a more systematic approach to parameter identifiability is introduced by Brun et al. (2001) and Brun et al. (2002), with applications in large environmental models and in an ASM2d model.

As regards parameter estimation of ASPs, estimation of kinetic parameters has been discussed in Papers A and B using the least square estimation technique. In addition, biological WWTPs are usually subject to large variations in both the influent pollutant concentrations and flow rates, which can strongly affect the operating conditions of the process. However, the large diversity and complexity of the components presented in the process (including biomass) makes it difficult to implement sensors at the entrance of the treatment plant. As a consequence, model-based estimators are applied in WWTPs to estimate the unmeasured input variables (i.e., disturbances from the process control point of view). Some disturbance estimation techniques are introduced in Darouach et al. (1995), Pina & Botto (2006), and Ha & Trinh (2004). In biological process applications, Aubrun et al. (2001) and Theilliol et al. (2003) propose to use the Kalman filter-based methods to estimate the input at the entrance of anaerobic wastewater treatment processes. In fact, disturbances can be considered as time-varying parameters, and parameter estimation techniques could be applicable to disturbance estimation. Therefore, disturbance estimation is discussed in this chapter in conjunction with parameter estimation.

The present chapter describes parameter estimation techniques for ASPs, with particular application to the Duvbacken P-removal plant. Section 4.2 describes the theoretical background and the procedure for parameter identifiability analysis. Estimation of kinetic parameters in the Duvbacken model and model validation is presented in section 4.3. State and disturbance estimation using the Kalman filter technique is discussed in section 4.4. In this chapter, the description of parameter identifiability follows the presentation given in Paper A and the Kalman filter-based algorithm uses the uniform notation as in chapter 3.

4.2 Parameter identifiability analysis

4.2.1 Theoretical background

Least squares estimation

A relatively standard method for model fitting is to select the model parameters $\theta \in \mathbb{R}^{n_\theta}$ such that the least squares criterion defined as

$$J = \frac{1}{2} \sum_{t=1}^T (y_t - y_t^m)^T (y_t - y_t^m)$$

is minimized, where $y_t \in \mathbb{R}^{n_y}$ is the output observed from the real plant at time t , $y_t^m \in \mathbb{R}^{n_y}$ is the model output at time t , while T is the number of available observations. That is, based on experimental data (output y_t and control input u_t , $t \in \{1, \dots, T\}$), the best parameter estimate $\hat{\theta}$ is sought in the sense that the model output $y_t^m(\hat{\theta})$ resembles the real output y_t as well as possible. The theoretical background of parameter identifiability analysis has also been discussed in Paper A.

Sensitivity and collinearity

We introduce the notation

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_T \end{pmatrix}, \quad y^m = \begin{pmatrix} y_1^m \\ \vdots \\ y_T^m \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ \vdots \\ u_T \end{pmatrix}, \quad e = \begin{pmatrix} e_1 \\ \vdots \\ e_T \end{pmatrix},$$

where e_t is the observation error $y_t - y_t^m$; hence $y, y^m, e \in \mathbb{R}^{T \cdot n_y \times 1}$, and $u \in \mathbb{R}^{T \cdot n_u \times 1}$. Clearly, model prediction y_t^m depends on the chosen parameter values. This dependence is quantified by the sensitive row vector \tilde{s}_t^T at time t , defined as

$$\tilde{s}_t^T = \frac{\partial y_t^m}{\partial \theta} = \left(\frac{\partial y_t^m}{\partial \theta_1} \quad \dots \quad \frac{\partial y_t^m}{\partial \theta_{n_\theta}} \right) \in \mathbb{R}^{n_y \times n_\theta}.$$

We can then build the sensitivity matrix S as

$$S = \begin{pmatrix} \tilde{s}_1^T \\ \vdots \\ \tilde{s}_T^T \end{pmatrix} = \begin{pmatrix} s_1 & \dots & s_{n_\theta} \end{pmatrix} \in \mathbb{R}^{T \cdot n_y \times n_\theta}, \quad s_j = \frac{\partial y^m}{\partial \theta_j} = \begin{pmatrix} \frac{\partial y_1^m}{\partial \theta_j} \\ \vdots \\ \frac{\partial y_T^m}{\partial \theta_j} \end{pmatrix} \in \mathbb{R}^{T \cdot n_y \times 1}.$$

Assume that we have some *a priori* estimate/idea θ^i of the true value θ^* of θ . The criterion J can be minimized by using Newton iteration:

$$\frac{\partial^2 J}{\partial \theta^2} \Big|_{\theta^i} (\theta - \theta^i) = - \frac{\partial J}{\partial \theta} \Big|_{\theta^i},$$

where

$$\begin{aligned} \frac{\partial J}{\partial \theta} &= - \sum_{i=1}^{n_y} \left(\sum_{t=1}^T (y_{t,i} - y_{t,i}^m) \frac{\partial y_{t,i}^m}{\partial \theta} \right) = - \sum_{i=1}^{n_y} \left(\sum_{t=1}^T e_{t,i} \frac{\partial y_{t,i}^m}{\partial \theta} \right) \\ \frac{\partial^2 J}{\partial \theta^2} &= \sum_{i=1}^{n_y} \sum_{t=1}^T \frac{\partial y_{t,i}^m}{\partial \theta} \left(\frac{\partial y_{t,i}^m}{\partial \theta} \right)^T - \sum_{i=1}^{n_y} \sum_{t=1}^T e_{t,i} \frac{\partial^2 y_{t,i}^m}{\partial \theta^2}. \end{aligned}$$

If θ is close to the correct parameter θ^* , it is common to assume that e_t will be close to zero on the average and hence use the Gauss-Newton approximation

$$\left\| \sum_{i=1}^{n_y} \sum_{t=1}^T \frac{\partial y_{t,i}^m}{\partial \theta} \left(\frac{\partial y_{t,i}^m}{\partial \theta} \right)^T \right\| \gg \left\| \sum_{i=1}^{n_y} \sum_{t=1}^T e_{t,i} \frac{\partial^2 y_{t,i}^m}{\partial \theta^2} \right\|.$$

Thus, the Newton iterator is reduced to the expression

$$\sum_{i=1}^{n_y} \sum_{t=1}^T \frac{\partial y_{t,i}^m}{\partial \theta} \left(\frac{\partial y_{t,i}^m}{\partial \theta} \right)^T \Big|_{\theta^i} (\theta - \theta^i) = \sum_{i=1}^{n_y} \left(\sum_{t=1}^T e_{t,i} \frac{\partial y_{t,i}^m}{\partial \theta} \right) \Big|_{\theta^i},$$

and we seek the value $\hat{\theta}$ such that the equality holds

$$\sum_{i=1}^{n_y} \sum_{t=1}^T \frac{\partial y_{t,i}^m}{\partial \theta} \left(\frac{\partial y_{t,i}^m}{\partial \theta} \right)^T \Big|_{\theta^i} (\hat{\theta} - \theta^i) = \sum_{i=1}^{n_y} \left(\sum_{t=1}^T e_{t,i} \frac{\partial y_{t,i}^m}{\partial \theta} \right) \Big|_{\theta^i}$$

or

$$(S^T S) \Big|_{\theta^i} \cdot \Delta \hat{\theta} = S^T e \Big|_{\theta^i}, \quad (4.1)$$

where $\Delta \hat{\theta} \triangleq \hat{\theta} - \theta^i$.

Clearly, matrix S must have full column rank $r \triangleq \text{rank } S = n_\theta$ in order to have a unique solution for Equation 4.1. In the simplest case, this requirement fails if $s_j \equiv 0$, i.e., output y^m is insensitive to parameter θ_j under the given experimental conditions. Furthermore, if $\|s_j\|$ is “small”, y^m is weakly sensitive to θ_j , indicating that in practice $\Delta \hat{\theta}_j$ can not be found.

Even when no columns of S are zero, there may be problems with linear dependence among the columns of S . Such a linear dependence is denoted *collinearity*, and one measure of nearness-to-collinearity in S is the smallest singular value $\sigma_{\min}(S)$. Since $\sigma_{\min}(S)$ is an absolute measure of nearness, it is difficult to interpret. To avoid this problem of interpretation, one can either scale S before computing $\sigma_{\min}(S)$, or else scale $\sigma_{\min}(S)$ afterwards (Stewart 1987). In the first strategy, one possible scaling strategy is to normalize every column in S . For the second strategy, it has become standard practice in linear algebra to scale $\sigma_{\min}(S)$ by the largest singular value $\sigma_{\max}(S)$, and present the inverse: this is the *condition number* $\kappa(S)$, which is

$$\kappa(S) = \frac{\sigma_{\max}(S)}{\sigma_{\min}(S)}.$$

In the sequel, a procedure for parameter identifiability and estimation will be introduced, with special emphasis on two identifiability measures: sensitivity measure and collinearity index. The first is designed to assess individual parameter importance in a least squares parameter estimation context, and the second accounts for the degree of collinearity among parameters for various parameter subset.

4.2.2 Procedure for selecting identifiable parameters

As mentioned in section 4.1, calibration of ASM models may be a challenging tasks. ASM models are most often highly complex, and they are usually high-order nonlinear systems incorporating a large number of parameters. In addition, models usually are overparameterized or have poorly identifiable parameters with respect to the available experimental data. Parameter estimation of ASM models is therefore addressed by *ad hoc* selecting and tuning approaches. The identifiability procedure described in Brun et al. (2002) seems well suited to WWTPs, since it handles large simulation models and provides systematic selection and tuning of parameter subsets, see also Brun et al. (2001), Dueñas Díez et al. (2006). The approach of Brun and co-workers is rooted in basic linear algebra and least squares parameter estimation, as discussed above. Also, for automated analysis of sensitivity of a large number of parameters, Lund & Foss (2008) present an easily implemented method on parameter ranking by successive orthogonalization, which may be of interest for on-line estimation.

Figure 4.1 illustrates the general steps in the procedure for parameter identifiability analysis and estimation. The procedure is composed of three main stages: experiments, identifiability analysis, and estimation and validation. Here, we give a brief discussion of each of the steps, and more details of the approach can be found in Dueñas Díez (2004) and Dueñas Díez et al. (2006).

1. *Experimental design.* Designing experiments which provide as much information as possible to determine and estimate the regarded parameters usually reward us with good parameter estimates.
2. *Experimental campaign.*
3. *Prior analysis.* Scaling of outputs and parameters is essential. For the numerical analysis to make sense, the quantities should be dimensionless. Thus, it is recommended to use dimensionless, scaled quantities

$$\tilde{y}_{t,j} = \frac{y_{t,j}}{y_j^\circ}, \quad \tilde{\theta}_j = \frac{\theta_j}{\theta_j^\circ}$$

where y_j° is a scale factor with the same physical dimension as $y_{t,j}$, typically indicating the variation in $y_{t,j}$ or a nominal value, while θ_j° is a scale factor with the same physical dimension as θ_j , typically indicating the uncertainty in θ_j . In the sequel, scaling parameters and outputs as part of the preprocessing of the data is highly recommended.

4. *Calculate model output.* Identifiability analysis as discussed in this approach requires only simulated model output.
5. *Compute sensitivity.* Based on the definition of sensitivities in 4.2.1, the column s_j of the scaled sensitivity matrix S is computed as

$$s_j \approx \text{sc}_j \frac{y^m(\theta^0 + e_j \cdot \delta\theta_j) - y^m(\theta^0)}{\delta\theta_j},$$

where sc_j is the proper scaling factor and $\delta\theta_j$ is the perturbation value of θ_j . $y^m(\theta^0)$ and $y^m(\theta^0 + e_j \cdot \delta\theta_j)$ represent the nominal output and the perturbed output, respectively, while e_j is column j of the identity matrix $I \in \mathbb{R}^{n_\theta \times n_\theta}$.

6. *Parameter importance ranking.* With Equation 4.1, the first step is to consider whether the model output is insensitive to any of the parameters, i.e. to compare $\|s_j\|$ for the various parameters θ_j . Brun et al. (2001) propose to scale $\|s_j\|$ by the square-root of the amount of data $n_y T$ and introduce the *sensitivity measure* δ_j^{msqr} defined as

$$\delta_j^{\text{msqr}} \triangleq \frac{1}{\sqrt{n_y T}} \|s_j\|_2,$$

instead of using $\|s_j\|_2$ directly; they also consider other norms $\|s_j\|$. The strategy is then to sort the various parameters according to the value of δ_j^{msqr} . A high value of δ_j^{msqr} means that a change in the parameter θ_j has an important influence on the model output.

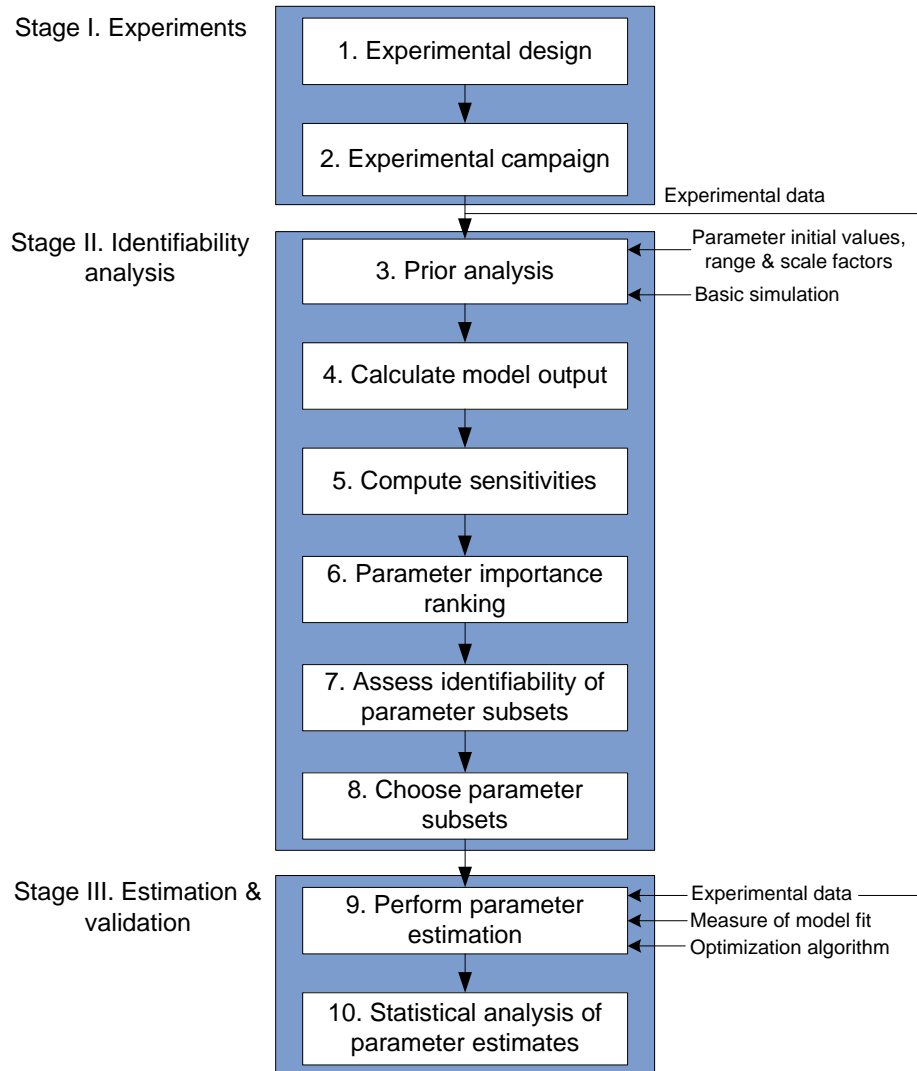


Figure 4.1: Procedure for parameter identifiability analysis and estimation. Taken from Dueñas Díez et al. (2006).

7. *Assess identifiability of parameter subsets.* In considering Equation 4.1, the next step is to look for collinearity. Brun et al. (2001) defines a *collinearity index* γ_K as

$$\gamma_K \triangleq \frac{1}{\sigma_{\min}(\tilde{S}_K)} = \frac{1}{\sqrt{\sigma_{\min}(\tilde{S}_K^T \tilde{S}_K)}},$$

where \tilde{S}_K is a $T \cdot n_y \times K$ submatrix of \tilde{S} containing those columns that correspond to the K parameter subset of θ . \tilde{S} is a slightly modified sensitivity matrix where the columns have been normalized as noted above, and $\sigma_{\min}(\tilde{S}_K)$ is the smallest singular value of \tilde{S}_K . A high value of γ_K indicates that the parameter subset K is poorly identifiable even if the individual parameters of K are among the top parameters of the parameter importance ranking.

8. *Choose parameter subsets.* An identifiable parameter subset is chosen based on the numerical analysis described in steps 3–8.
9. *Perform parameter estimation.* At this step, the measured data from the real plant are used. Various estimation techniques can be applied, e.g. weighted least squares estimation, maximum likelihood estimation, the Kalman filter, moving horizon estimation, asymptotic observers, etc., (Crassidis & Junkins 2004), (Gelb 1992), (Lewis 1986), (Dochain & Vanrolleghem 2001).
10. *Statistical analysis of parameter estimates.* Finally, a statistical analysis should be carried out to analyze the uncertainties of the estimates, and to evaluate the quality of the fitting.

4.3 Estimation of kinetic parameters in the Duvbacken model

4.3.1 Parameter identifiability and estimation

Identifiability analysis and estimation of the kinetic parameters for the Duvbacken plant are discussed in the Papers A and B. The online measurements of the Duvbacken plant are taken in the period of October 24–29, 2005. The influent characteristics and on-line and off-line measurements are given in Paper A.

The methodology presented in subsection 4.2.2 is applied to the Duvbacken model, and two studies are carried out:

- As an initial study before doing parameter estimation based on real data from the plant, we check whether suitable estimates can be obtained based on simulation data with known true parameter values, see Paper A. The model validation result shows that the model fit is satisfactory, especially for those outputs which are important to the plant operation (e.g. S_{PO_4} and TP). However, in that study, the model implementation in Matlab is restrictive with respect to possible input signals, and components of the model are difficult to modify and reuse.

- The model is thus rewritten in the more flexible modeling language Modelica, as described in section 2.4. The procedure shown in Figure 4.1 is used to analyze the identifiability again, but this time we use the same influent flowrate as in the real plant. Eventually, the selected parameter subset is estimated from real data, see Paper B.

Comparing the identifiable parameter subsets obtained in the studies, it can be seen that most parameters in the optimally identifiable parameter set are the same. This confirms that these parameters are indeed identifiable.

Based on the identifiability analysis, nonlinear least squares estimation is used, with the following optimization criterion

$$\hat{\theta} = \arg \min_{\theta} J = \arg \min_{\theta} \frac{1}{2} e^T W e. \quad (4.2)$$

Here, the computation of the model output y_t^m in the criterion J is based on ballistic simulation, i.e., using multiple-step-ahead prediction errors. This is in contrast to applying recursive estimation techniques (e.g. the Kalman filter-based approaches), which minimize the one-step-ahead prediction error. To solve the optimization problem, two different algorithms are applied considering convergence speed of optimization. The Finkel-Kelley `Direct` optimization algorithm is proven to be more robust but slower than Matlab's `lsqnonlin` algorithm.

4.3.2 Model validation

Once parameter estimation has been carried out, the model with the estimated parameters should be tested on a new set of experimental data. The measurements taken from the Duvbacken plant over the period of October 17–24, 2005 are used for model validation, with a periodically varying flowrate shown in Figure 4.2. A sudden increase of the feed flowrate (as well as COD loadings) occurs at the last day, while the oxygen supply to the process is insufficient. This results in the dramatic increasing of the phosphate concentrations, see the solid lines in the last two subplots of Figure 4.3. The simulation results in Figure 4.3 show that the model with the estimated parameters (dashed line) captures the large variations in S_{PO_4} and P_{tot} , whereas the model with the nominal parameters (dotted line) gives poor prediction. This indicates again the conclusion obtained in the papers: the resulting model gives good prediction of the measurements related to phosphorus.

4.3.3 Confidence interval of parameter estimates

Once the model is fitted to experimental data, we need to find out how accurate we have estimated the parameters and conclude whether this accuracy is sufficient for our purpose. Typically, the confidence interval of the parameter estimates are used for assessing the accuracy.

A confidence interval can be obtained by using the relation (Rawlings & Ekerdt 2002):

$$\left(\theta - \hat{\theta}\right)^T H|_{\theta=\hat{\theta}} \left(\theta - \hat{\theta}\right) \leq 2s^2 n_p F(n_p, n_d - n_p, \alpha)$$

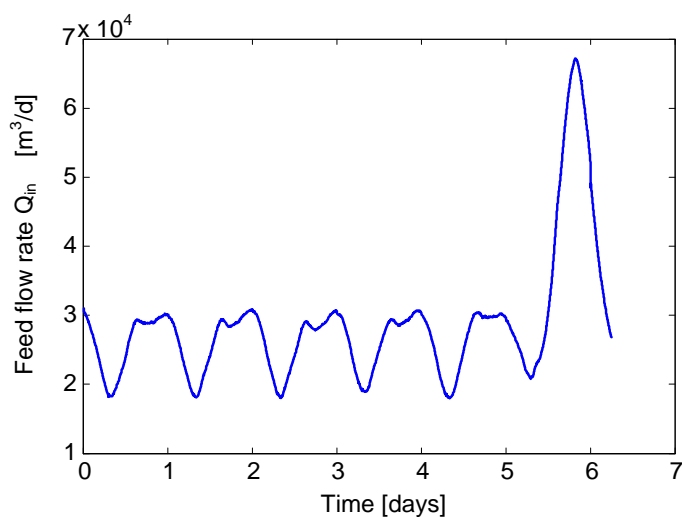


Figure 4.2: Measured flowrate of the process feed in the period of October 17–24, 2005.

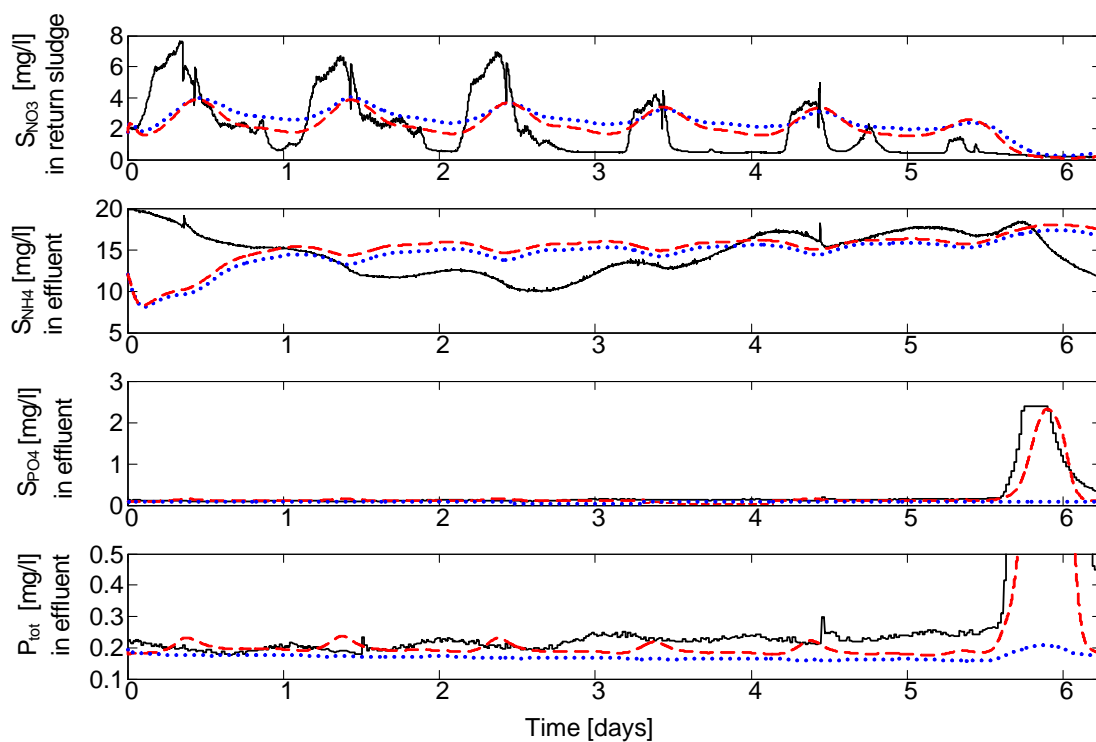


Figure 4.3: Comparison of the model outputs using the estimated parameters (dashed) and using the nominal parameters (dotted) with the real measurements (solid) taken in the period of October 17–24, 2005.

in which θ is the real value of the parameter, $\hat{\theta}$ is the estimated value of the parameter, n_p is the number of estimated parameters, n_d is the number of data points, α is the desired confidence level, and F represents the F -distribution which can be read from statistical tables. It should be mentioned that θ and $\hat{\theta}$ only consist of the chosen identifiable parameters. H is the Hessian of the objective function in Equation 4.2, and as discussed in 4.2.1, H can be calculated with the following approximate expression

$$H = S_K^T W S_K,$$

where S_K is a submatrix of S containing those columns that correspond to the chosen parameter subset K , while W is the weight matrix. s^2 is the sample variance

$$s^2 = \frac{2J(\hat{\theta})}{n_d - n_p}.$$

For the parameter estimation with 10 estimated parameters, the 95% confidence interval is:

$$(\theta - \hat{\theta})^T (S_K^T W S_K)|_{\theta=\hat{\theta}} (\theta - \hat{\theta}) \leq 1.4178. \quad (4.3)$$

The confidence interval is often reported with plus/minus limits, i.e., $\hat{\theta} = \hat{\theta} \pm \theta_\Delta$, in which θ_Δ can be given as

$$\theta_\Delta = \sqrt{2s^2 n_p F(n_p, n_d - n_p, \alpha) \cdot \tilde{H}|_{\theta=\hat{\theta}}},$$

where \tilde{H} is the diagonal vector of H^{-1} . Then, the corresponding confidence limits to Equation 4.3 is found to be:

$$\hat{\theta} = \begin{pmatrix} K_h \\ K_X \\ q_{fe} \\ b_{HET} \\ q_{PP} \\ \mu_{PAO} \\ K_{PS,PAO} \\ K_{PHA} \\ \mu_{AUT} \\ K_{P,AUT} \end{pmatrix} = \begin{pmatrix} 3.000 \\ 0.122 \\ 2.889 \\ 0.477 \\ 1.000 \\ 1.000 \\ 0.193 \\ 0.010 \\ 0.852 \\ 0.014 \end{pmatrix} \pm \begin{pmatrix} 0.076 \\ 0.082 \\ 0.086 \\ 0.133 \\ 0.092 \\ 0.107 \\ 0.102 \\ 0.078 \\ 0.305 \\ 0.108 \end{pmatrix}.$$

The most parameters are determined with narrow confidence intervals. However, some parameters (e.g. K_{PHA} , $K_{P,AUT}$) have very wide confidence intervals. It should be noticed that the confidence limits can give an overly pessimistic picture since they define a rectangle which can encompass quite a bit more area than the actual confidence region, as illustrated in Figure 4.4. This is particularly the case when there is a strong correlation between the parameters.

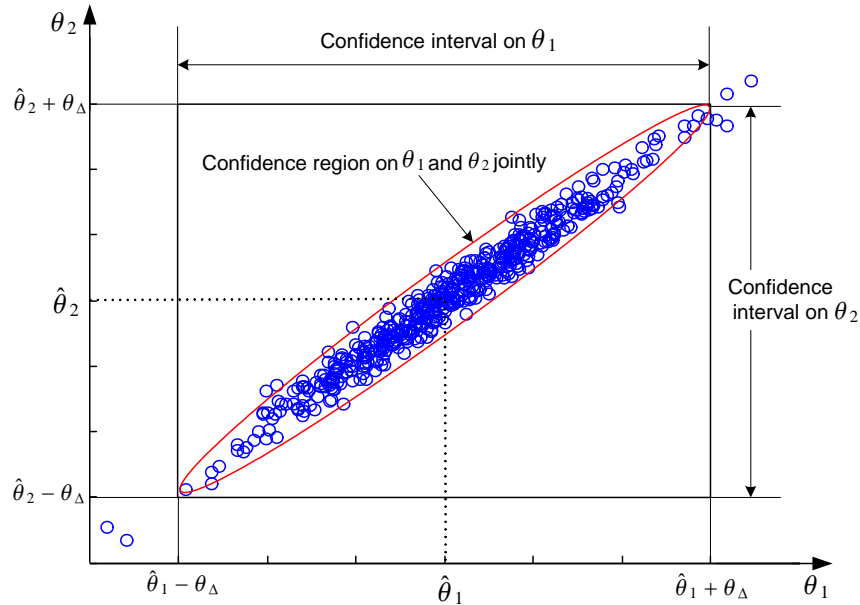


Figure 4.4: An example of confidence region and confidence intervals in two-dimensional case. Adapted from Rawlings & Ekerdt (2002).

4.4 Disturbance estimation using the Kalman filter

To have a good prediction capability of the Duvbacken model, we need information about the input signals. The on-line measurement of the influent flowrate of the plant is available, but there are no sensors to measure the influent compositions. Indeed, a good characterization of the COD, nitrogen, and phosphorus components in the influent is very important to the performance of the ASM models (Olsson & Newell 1999). Therefore, in this section, we accept model parameter values provided in the ASMs literatures, and instead estimate the influent compositions. The objective is to develop a method for simultaneously estimating both state and disturbance for the Duvbacken plant. This topic has not been presented in any papers in Part II of the thesis.

4.4.1 State and disturbance estimation

The Kalman filter is widely applied for on-line implementation of state and disturbance estimation. Also, in section 3.3, we have shown that the discrete-time EKF provides fast and satisfactory state estimation for an ASP. Therefore, the discrete-time EKF is the chosen algorithm for state and disturbance estimation here.

For the general dynamic system

$$\begin{aligned} x_k &= f(x_{k-1}, u_{k-1}, w_{k-1}, d_{k-1}) \\ y_k &= h(x_k, v_k), \end{aligned}$$

a standard approach to disturbance estimation using the Kalman filter is to augment the state vector $x \in \mathbb{R}^{n_x}$ with the disturbance $d \in \mathbb{R}^{n_d}$, and to describe the time variance of

disturbances with a state-space model:

$$\begin{aligned}\xi_k &= f_d(\xi_{k-1}) + r_{k-1} \\ d_k &= h_d(\xi_k),\end{aligned}$$

where $\xi \in \mathbb{R}^{n_\xi}$ is the state of the disturbance model and r_k is white noise. It is assumed that w_k , v_k , and r_k are independent, zero-mean Gaussian noise processes of covariance matrices Q_k , R_k , and D_k , respectively, i.e.,

$$\begin{aligned}w_k &\sim \mathcal{N}(0, Q_k) \\ v_k &\sim \mathcal{N}(0, R_k) \\ r_k &\sim \mathcal{N}(0, D_k).\end{aligned}$$

The choice of which disturbances to estimate should be guided by an identifiability analysis, as described in section 4.2.

We first combine the state and disturbance vectors to form an augmented state such that the dynamics may be represented by

$$\tilde{x}_k \triangleq \begin{bmatrix} x_k \\ \xi_k \end{bmatrix} = \begin{bmatrix} f(x_{k-1}, u_{k-1}, w_{k-1}, h_d(\xi_{k-1})) \\ f_d(\xi_{k-1}) + r_{k-1} \end{bmatrix} \quad (4.4)$$

$$y_k = h(x_k, v_k). \quad (4.5)$$

Table 4.1 gives the algorithm steps for simultaneous state and disturbance estimation, which correspond directly to the state estimation EKF steps except with larger matrix operations.

4.4.2 Observability analysis

Observability analysis for state and disturbance estimation

Let us now consider a linear system with unmeasured disturbances

$$\frac{dx(t)}{dt} = Fx(t) + Cd(t) \quad (4.6)$$

$$y(t) = Hx(t), \quad (4.7)$$

in which the disturbance can be described as a state-space model

$$\begin{aligned}\frac{d\xi(t)}{dt} &= F_d\xi(t) \\ d(t) &= H_d\xi(t),\end{aligned}$$

where $\xi \in \mathbb{R}^{n_\xi}$ is the state of the disturbance model.

Then, the system can be rewritten as an augmented system in the form of

$$\frac{d\tilde{x}(t)}{dt} = \tilde{F}\tilde{x}(t) \quad (4.8)$$

$$y(t) = \tilde{H}\tilde{x}(t), \quad (4.9)$$

Table 4.1: Summary of the extended Kalman filter for simultaneous state and input estimation.

Nonlinear state-space model

$$\begin{bmatrix} x_k \\ \xi_k \end{bmatrix} = \begin{bmatrix} f(x_{k-1}, u_{k-1}, w_{k-1}, h_d(\xi_{k-1})) \\ f_d(\xi_{k-1}) + r_{k-1} \end{bmatrix} \quad \text{or} \quad \begin{aligned} \tilde{x}_k &= \tilde{f}(\tilde{x}_{k-1}, u_{k-1}, \tilde{w}_{k-1}) \\ y_k &= \tilde{h}(\tilde{x}_k, v_k) \end{aligned}$$

$$y_k = h(x_k, v_k),$$

where $w_k \sim \mathcal{N}(0, Q_k)$, $v_k \sim \mathcal{N}(0, R_k)$ and $r_k \sim \mathcal{N}(0, D_k)$.

We let $\tilde{x}_k \triangleq [x_k^T, \xi_k^T]^T$, $\tilde{w}_k \triangleq [w_k^T, r_k^T]^T$, and $Q_{\tilde{w}} = \text{diag}(Q_k, D_k)$

Definition:

$$F_k = \left. \frac{\partial \tilde{f}}{\partial \tilde{x}} \right|_{\hat{\tilde{x}}_{k|k}, u_k, 0}, \quad G_k = \left. \frac{\partial \tilde{f}}{\partial \tilde{w}} \right|_{\hat{\tilde{x}}_{k|k}, u_k, 0}, \quad H_k = \left. \frac{\partial \tilde{h}}{\partial \tilde{x}} \right|_{\hat{\tilde{x}}_{k|k-1}, u_k, 0}, \quad L_k = \left. \frac{\partial \tilde{h}}{\partial v} \right|_{\hat{\tilde{x}}_{k|k-1}, u_k, 0}$$

Initialization: for $k = 0$, set $\tilde{x}_0 \sim \mathcal{N}(\hat{\tilde{x}}_{0|0}, P_{0|0}^{\tilde{x}})$

Computation: for $k = 1, 2, \dots$, compute

State time update	$\hat{\tilde{x}}_{k k-1} = \tilde{f}(\hat{\tilde{x}}_{k-1 k-1}, u_{k-1}, 0)$
Covariance time update	$P_{k k-1}^{\tilde{x}} = F_{k-1} P_{k-1 k-1}^{\tilde{x}} F_{k-1}^T + G_{k-1} Q_{k-1}^{\tilde{w}} G_{k-1}^T$
Output estimate	$\hat{y}_{k k-1} = \tilde{h}(\hat{\tilde{x}}_{k k-1}, 0)$
Kalman gain	$K_k = P_{k k-1}^{\tilde{x}} H_k^T (H_k P_{k k-1}^{\tilde{x}} H_k^T + L_k R_k L_k^T)^{-1}$
State measurement update	$\hat{\tilde{x}}_{k k} = \hat{\tilde{x}}_{k k-1} + K_k (y_k - \hat{y}_{k k-1})$
Covariance measurement update	$P_{k k}^{\tilde{x}} = (I - K_k H_k) P_{k k-1}^{\tilde{x}} (I - K_k H_k)^T + K_k L_k R_k L_k^T K_k^T$

where $\tilde{x}(t) \triangleq \begin{pmatrix} x(t) \\ \xi(t) \end{pmatrix} \in \mathbb{R}^{(n_x+n_\xi) \times 1}$, $\tilde{F} = \begin{pmatrix} F & CH_d \\ 0 & F_d \end{pmatrix}$, and $\tilde{H} = \begin{pmatrix} H & 0 \end{pmatrix}$.

According to the PBH observability test (Theorem 6 in section 3.2), this augmented system is said to be observable if

$$\text{rank} \begin{pmatrix} \tilde{H} \\ sI - \tilde{F} \end{pmatrix} = \text{rank} \begin{pmatrix} H & 0 \\ sI - F & -CH_d \\ 0 & sI - F_d \end{pmatrix} = n_x + n_\xi. \quad (4.10)$$

Before discussing the conditions for meeting Equation 4.10, we need the concept of transmission zeros.

Definition 7 (Transmission zeros) Consider a system (F, C, H) described in Equations 4.6 and 4.7 and its linear pencil $\Sigma(s) = \begin{pmatrix} H & 0 \\ sI - F & -C \end{pmatrix}$, the transmission zeros¹ of the system are those s such that

$$\text{rank}(\Sigma(s)) = \text{rank} \begin{pmatrix} H & 0 \\ sI - F & -C \end{pmatrix} < n_x + \min[n_y, n_d].$$

Now we consider the rank condition in Equation 4.10. For all $s \in \text{eig}(F_d)$, where $\text{eig}(F_d)$ denotes the eigenvalues of F_d , in order to fulfill the rank condition, the matrix $\begin{pmatrix} H & 0 \\ sI - F & -CH_d \end{pmatrix}$ needs to have full rank. Hence, the number of states in the disturbance model should be less than or equal to the number of outputs (i.e. $n_\xi \leq n_y$), and there must not exist any transmission zeros at $s \in \text{eig}(F_d)$ for the system (F, CH_d, H) . For all $s \notin \text{eig}(F_d)$, the n_ξ last columns are linearly independent, i.e. $\text{rank}(sI - F_d) = n_\xi$. If the system without disturbance is observable, i.e. $\text{rank} \begin{pmatrix} H \\ sI - F \end{pmatrix} = n_x$, then the augmented system is observable.

These observation can be summarized in the following theorem:

Theorem 8 The augmented system described in Equations 4.8 and 4.9 is observable if the system has the following properties:

- The original system (F, H) is observable.
- The number of states in the disturbance model is less than or equal to the number of outputs, i.e. $n_\xi \leq n_y$.
- The system (F, CH_d, H) has no transmission zeros at any $s \in \text{eig}(F_d)$.

Example 9 (Constant disturbance) Let us reconsider the linear state equations 4.6 and 4.7. Now we assume that the disturbance is constant

$$\frac{d}{dt}d(t) = 0.$$

¹In Matlab, the function `zero` can be used to compute transmission zeros of LTI models.

The augmented system becomes

$$\begin{aligned}\frac{d\tilde{x}(t)}{dt} &= \tilde{F}\tilde{x}(t) \\ y(t) &= \tilde{H}\tilde{x}(t),\end{aligned}$$

where $\tilde{x}(t) \triangleq \begin{pmatrix} x(t) \\ d(t) \end{pmatrix} \in \mathbb{R}^{(n_x+n_d) \times 1}$, $\tilde{F} = \begin{pmatrix} F & C \\ 0 & 0 \end{pmatrix}$, and $\tilde{H} = \begin{pmatrix} H & 0 \end{pmatrix}$. The augmented system is observable if

$$\text{rank} \begin{pmatrix} H & 0 \\ sI - F & -C \\ 0 & sI \end{pmatrix} = n_x + n_d.$$

According to Theorem 8, the state equations 4.6 and 4.7 with constant disturbance is observable if

- the original system (F, H) is observable,
- the number of unknown disturbances is less than or equal to the number of outputs, i.e. $n_d \leq n_y$, and
- the system (F, C, H) has no transmission zeros at $s = 0$.

Structural observability

For the nonlinear WWTP model, in order to guarantee the (arbitrarily chosen) exponential convergence of the filter, the process must be locally observable, i.e., the linearized model must be observable and fulfill the observability test discussed above. This requires that the measurements must be sufficient in number and in their relationship to the predicted states and disturbances. To give a clue that helps us to find necessary measurements, a graph approach is performed for structural observability analysis (Lunze 1992). Based on the model structure, one can find system observability without knowing the precise system parameters. A system is said to be *output-connectable* (or *output-reachable*) if there exists for every state vertex a path from state vertex to at least one output vertex. The reachability test is a necessary condition of investigating the observability.

The structure of a system described by ASM2d model is shown in appendix 4.A, with the outputs that are available in practice. We can see that the states S_I , S_{N_2} , and X_I do not connected to the outputs, thus, the system is unobservable. In other words, to make the system output-connectable, these states have to be considered as available measurements.

4.4.3 Application to the Duvbacken plant

Process description

As introduced in Dochain & Vanrolleghem (2001), the general mathematical model of the Duvbacken plant can be formulated as

$$\begin{aligned}\frac{dx(t)}{dt} &= f(x(t), u(t)) + Cd(t) + w(t) \\ y &= h(x(t)) + v(t)\end{aligned}$$

where $x \in \mathbb{R}^{n_x}$ is the state, $u \in \mathbb{R}^{n_u}$ is the model input, $d \in \mathbb{R}^{n_d}$ is the disturbance that typically represents the feed composition, and the matrix C is known. The process noise and the measurement noise are assumed to be $w(t) \sim \mathcal{N}(0, Q)$ and $v(t) \sim \mathcal{N}(0, R)$, respectively.

In general terms, disturbance models have to be chosen according to the expected load. In the case of WWTPs, the feed composition of domestic wastewater may typically vary during the course of a day. However, wastewater characteristics vary from plant to plant, and there are quite few measurements taken from the feed. Therefore, an accurate disturbance model is difficult to find. It is more realistic to assume a constant disturbance

$$\frac{d}{dt}d(t) = r,$$

where r is the white noise.

Considering the application of the discrete-time EKF, the augmented model described in equations 4.4 and 4.5 becomes

$$\begin{aligned} \tilde{x}_k &= \begin{bmatrix} x_k \\ d_k \end{bmatrix} = \begin{bmatrix} f(x_{k-1}, u_{k-1}, w_{k-1}, d_{k-1}) \\ d_{k-1} + r_{k-1} \end{bmatrix} \\ y_k &= h(x_k, v_k). \end{aligned}$$

In contrast to the states x , the additional states d are considered as constants during the propagation phase in the filter. However, the values of d will be corrected in the update steps of the Kalman filter algorithms at discrete-time instants. A significant advantage of state augmentation for disturbance estimation is the fact that the algorithm can track time-varying disturbances.

A dynamic model of the Duvbacken plant has been described in section 2.3. In the model, each of the 4 steps in the plant (anaerobe, aerobe 1–3) includes 17 states and this gives a model with 68 states. The model input is the volumetric feed flowrate Q_{in} , while the feed compositions of 17 variables are the model disturbance. Due to the practical focus of this study, only the 4 most interesting feed concentrations are estimated: readily biodegradable substrate $S_{F,in}$, fermentation products $S_{A,in}$, ammonium $S_{NH_4,in}$, and slowly biodegradable substrate $X_{S,in}$. The other feed concentrations are chosen to be constant.

In the plant, 13 measurements listed in Table 4.2 are available with the sampling time of 1 min. However, according to the observability analysis in subsection 4.4.2, we find that at least the following 16 state variables² need to be available for measurement to render the Duvbacken system locally observable:

In anaerobe reactor: $S_{PO_4}, S_I, S_{N_2}, S_{ALK}, X_I, X_{TSS}$

In aerobe reactor 1: S_{O_2}, S_{PO_4}

In aerobe reactor 2: $S_{O_2}, S_{NH_4}, S_{NO_3}, S_{PO_4}$

In aerobe reactor 3: $S_{O_2}, S_{NH_4}, S_{NO_3}, S_{PO_4}$.

Consider the problem of constructing estimates of the unmeasured state variables and disturbances from the available outputs listed in Table 4.2. The resulting augmented system is observable if the conditions in Theorem 8 are fulfilled. It is obvious that the original system is unobservable due to the lack of measurements. Therefore, the augmented system is unobservable as well. In the sequel, we first choose to check whether the suitable state and

²Unfortunately, it is not realistic to measure all these states in practice.

Table 4.2: Measurements of the Duvbacken WWTP.

Output	Component	Description
y_1	$S_{\text{NH}_4,\text{a}1}$	Ammonium in outlet of aerobe reactor 1
y_2	$S_{\text{NO}_3,\text{a}1}$	Nitrate in outlet of aerobe reactor 1
y_3	$X_{\text{TSS},\text{a}1}$	Total suspended solids in outlet of aerobe reactor 1
y_4	$S_{\text{O}_2,\text{a}2}$	Dissolved oxygen in outlet of aerobe reactor 2
y_5	$S_{\text{NO}_3,\text{a}2}$	Nitrate in outlet of aerobe reactor 2
y_6	$X_{\text{TSS},\text{a}2}$	Total suspended solids in outlet of aerobe reactor 2
y_7	$S_{\text{O}_2,\text{a}3}$	Dissolved oxygen in outlet of aerobe reactor 3
y_8	$S_{\text{NO}_3,\text{rs}}$	Nitrate in return sludge chamber
y_9	$X_{\text{TSS},\text{rs}}$	Total suspended solids in return sludge chamber
y_{10}	$S_{\text{NH}_4,\text{eff}}$	Ammonium in effluent of the plant
y_{11}	$S_{\text{PO}_4,\text{eff}}$	Phosphate in effluent of the plant
y_{12}	TP_{eff}	Total phosphorus in effluent of the plant
y_{13}	$X_{\text{TSS},\text{eff}}$	Total suspended solids in effluent of the plant

disturbance estimates can be obtained for the observable process by the simulation study, then the application to the plant based on real data is investigated.

Simulation results

The proposed estimation approach is evaluated for the observable process based on responses from the simulation model. The input Q_{in} is obtained from the real plant over a 40 h period, and the time evolution of Q_{in} is plotted in Figure 4.5. The feed wastewater composition and the simulation results over the 40 h period for the plant are shown in Table 4.3.

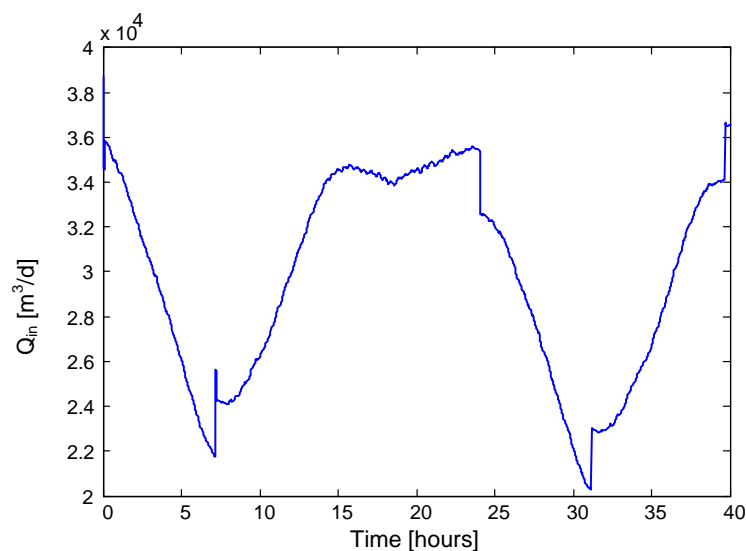


Figure 4.5: Time evolution of the measured feed flowrate over a 40 h period.

Table 4.3: Simulated concentrations of ASM2d components.

ASM2d components	Feed [g/m ³]	Effluent [g/m ³]	Recycled sludge [g/m ³]
S_{O_2}	0.1	2.91	2.91
S_F	30	0.47	0.47
S_A	20	0.01	0.01
S_{NH_4}	25	18.37	18.37
S_{NO_3}	0.5	4.05	4.05
S_{PO_4}	7	0.03	0.03
S_I	30	30.06	30.06
S_{ALK}	5	5.79	5.79
S_{N_2}	2	8.34	8.34
X_I	10	1.50	2436.31
X_S	75	0.05	78.10
X_H	10	1.98	3231.77
X_{PAO}	0	1.16	1885.27
X_{PP}	0	0.13	215.34
X_{PHA}	0	0.13	215.34
X_{AUT}	0.2	0.06	91.01
X_{TSS}	100	3.76	6120.39

Figure 4.6 shows the good convergence properties of the estimator for the disturbances. Also, the estimation of unmeasured states are plotted in Figure 4.7. Due to the large quantity of the state variables, only some of the states for the last reactor (aerobe reactor 3) are shown in the figure. The time evolutions of both the unmeasured states and the inputs illustrate the potential of the proposed approach to simultaneously estimate unmeasured states and disturbances.

Experimental results

Although the system is unobservable based on the experimental data, it is still of great interest to investigate the practical performance of the disturbance estimator. Figure 4.8 shows estimated values of the unmeasured feed concentrations based on real data, in comparison to the disturbance values used in the simulation model. However, since the real value of the disturbance is not known, this makes the initialization and tuning of the filter more difficult. In order to evaluate the performance of the filter, the evolution of some key outputs are displayed in Figures 4.9 and 4.10: the real output from the plant (solid line) and the estimated output with the EKF (dashed line). For verification of the disturbance estimate, we also include the model output with nominal disturbance values (dotted line) and the model output after disturbance estimation (dash-dotted line). The comparison of the simulations will indicate whether the model provides smaller prediction errors through disturbance estimation. We can see that the output prediction is quite good using the EKF. However, the estimated disturbances do not provide a significant improvement in the prediction capability of the model outputs.

From a practical point of view, it should be stressed that the obtained results are inter-

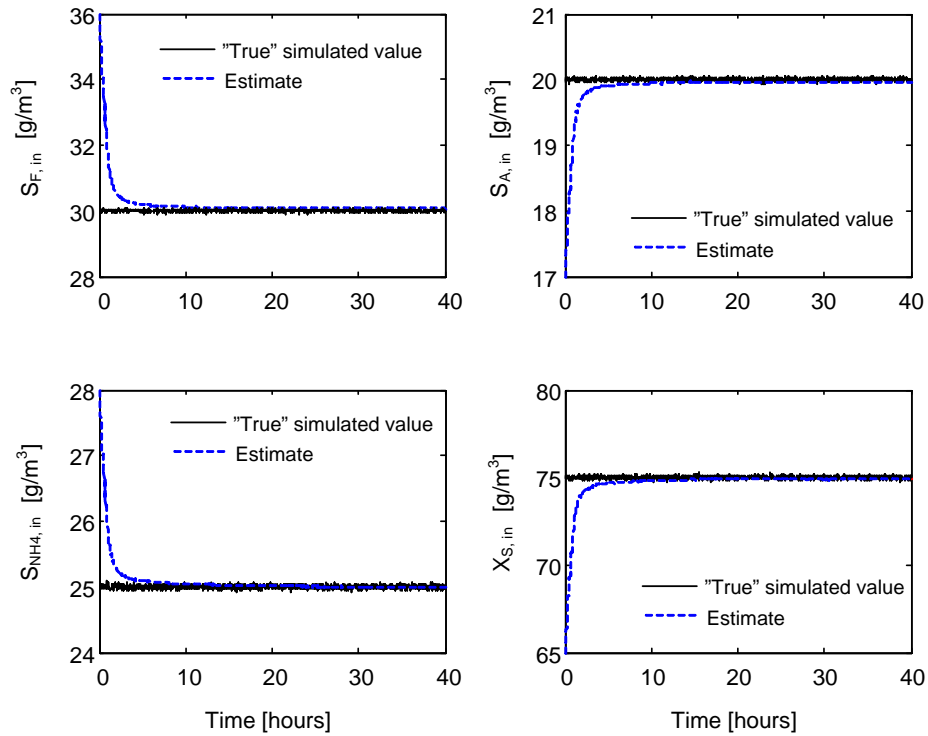


Figure 4.6: Time evolution of the disturbances for the observable process.

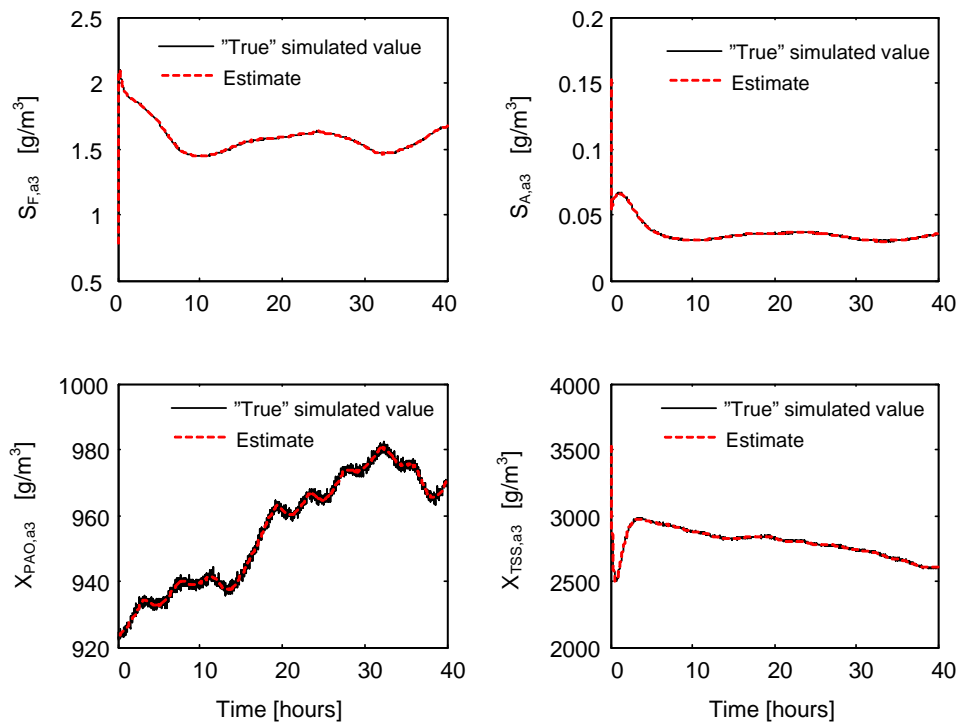


Figure 4.7: Time evolution of the unmeasured states for the observable process.

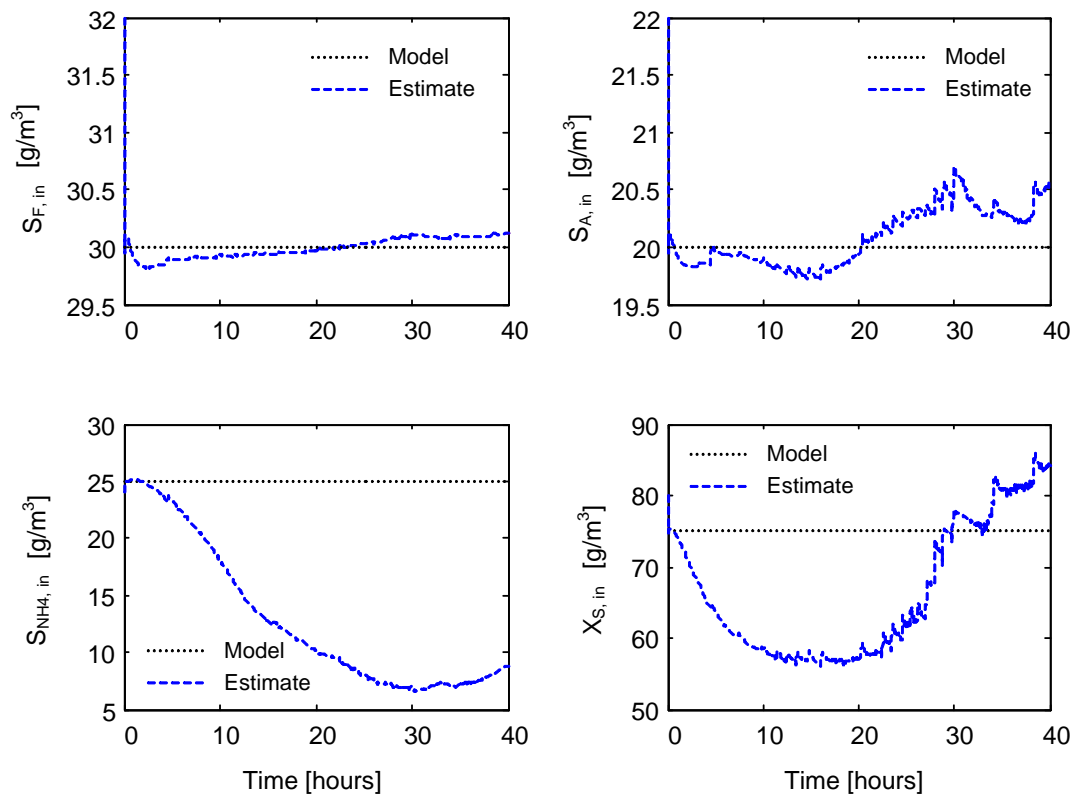


Figure 4.8: Estimated values of the disturbances based on the experimental data vs. nominal disturbance values used in the simulation model.

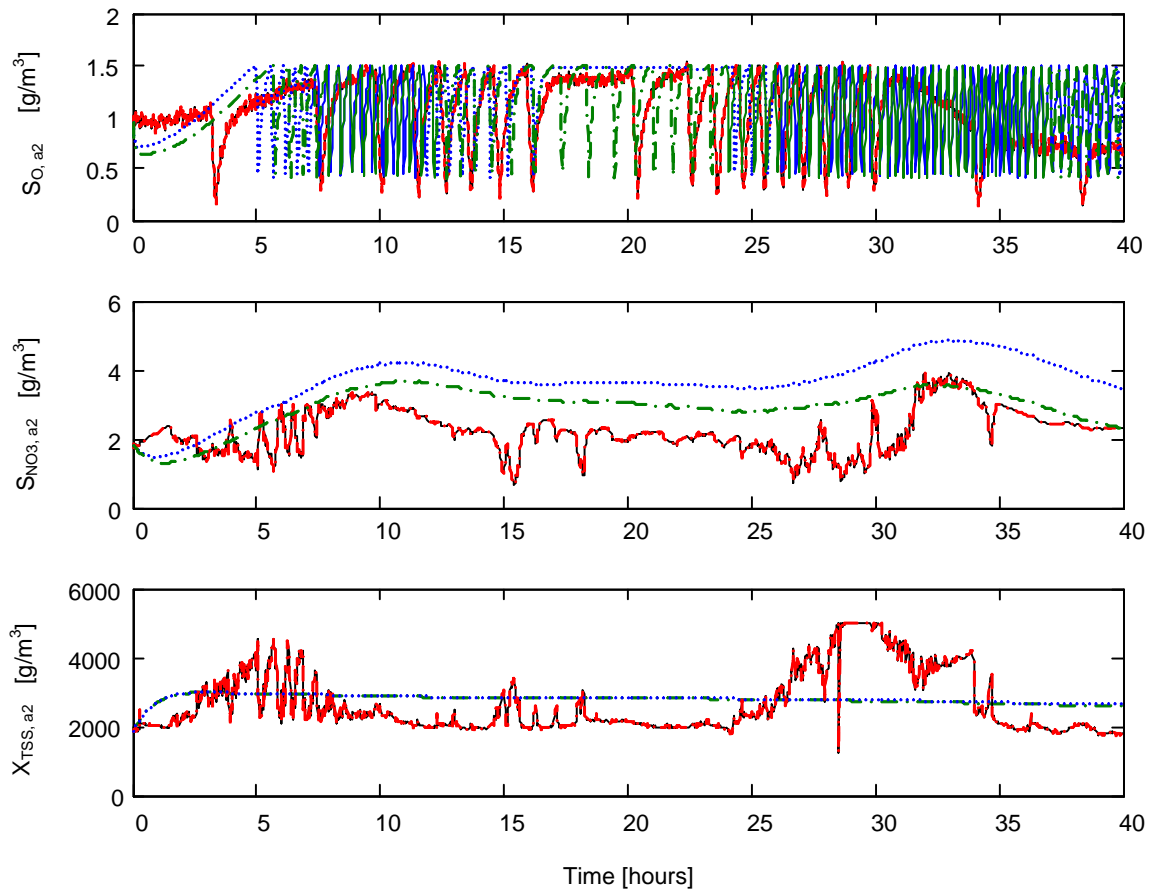


Figure 4.9: Comparison of the outputs in the aerobic reactor 2: experimental data (black, solid), estimated outputs (red, dashed), simulation results with the nominal values for the disturbances (blue, dotted), and simulated results with the estimated values for the disturbances (green, dash-dotted).

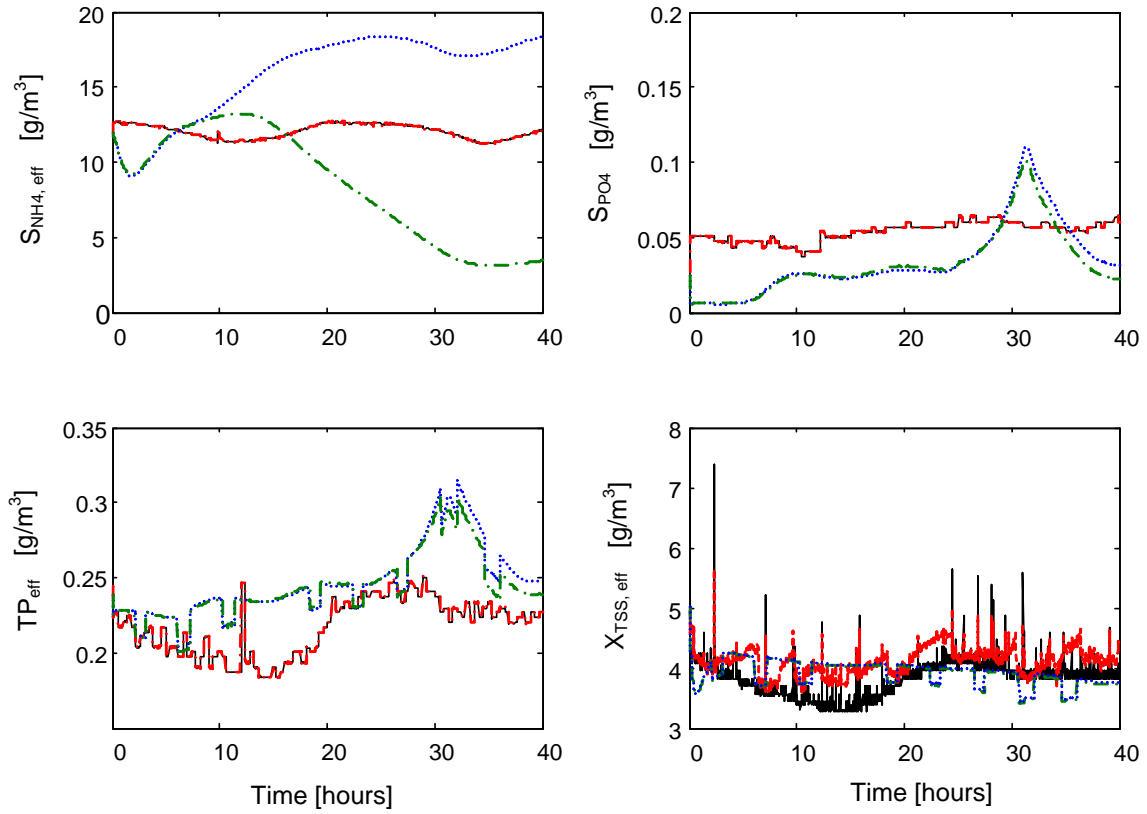


Figure 4.10: Comparison of the outputs in the effluent: experimental data (black, solid), estimated outputs (red, dashed), simulation results with the nominal values for the disturbances (blue, dotted), and simulated results with the estimated values for the disturbances (green, dash-dotted).

esting. Indeed, as it has been pointed out in the introduction, the estimation of the feed concentrations of WWTPs is a very challenging problem. The processes suffer from a lack of sensors, and their investment and operating costs together with difficulties of measuring gas and biomass concentrations limit practical implementation of sensors. The proposed approach allows us to estimate both the unmeasured model states and the disturbances, using a number of available measurements.

According to the application of the EKF to the state and disturbance estimation of the Duvbacken process, we would like to draw attention to the following conclusions:

- The EKF requires that a reliable process model is available and model fitting and validation should preferably be carried out. However, perfect models rarely exist and uncertainty of e.g. initial conditions and kinetics often appear in WWTPs. In this case, observability of the system is particularly important to ensure the dynamics of the EKF to be assigned arbitrarily.
- To render the Duvbacken system locally observable, the following additional state variables need to be available for measurement based on the measured output listed in Table 4.2: S_{PO_4} , S_{I} , S_{N_2} , S_{ALK} , X_{I} , and X_{TSS} in anaerobe reactor, S_{O_2} and S_{PO_4} in aerobe reactor 1, S_{NH_4} and S_{PO_4} in aerobe reactor 2, and S_{NO_3} in aerobe reactor 3. However, measurement of some of the state variables (e.g. S_{I} and S_{N_2}) is not realistic with today's technology.
- The EKF should be carefully tuned in the real-life application, otherwise biased results might appear. The main parameters to tune are the covariance matrices Q and R . Also, the choice of initial value of state error covariance P_0 plays an important role in the convergence properties of the EKF.

4.A Structural observability analysis of ASM2d model

The structure of the system described by the ASM2d model is illustrated in Figure 4.11.

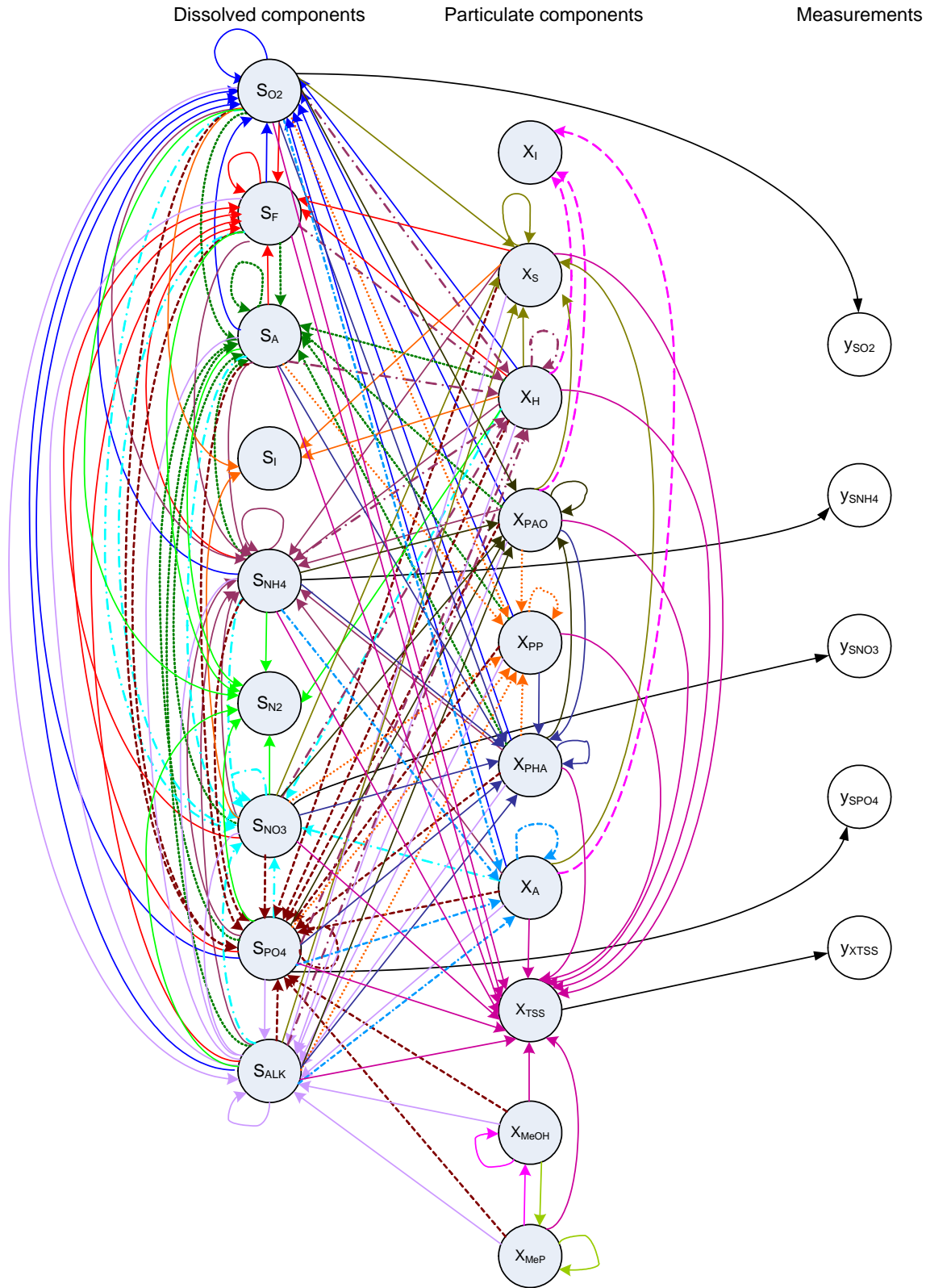


Figure 4.11: Structure of the system described by ASM2d model.

Chapter 5

Control Strategies

5.1 Introduction

Applying advanced control strategies to ASPs is a challenging task. Compared to other industrial processes, some of the challenges that are particular to the control of the ASP are stated in Olsson & Newell (1999):

- The disturbances of the ASP are extremely large.
- The process has significant nonlinearities limiting the usefulness of simple controllers.
- The process has stiff dynamics, i.e., there is a wide range of time constants (from a few minutes to several days).
- There are few manipulated variables.
- Sensors do not exist for many states and are expensive for others.
- The strict EU Directive 91/271/EEC on Urban Wastewater Treatment strongly constraints the maximum pollutant concentrations allowed in the effluent.

For ASPs, many control techniques have been proposed in the literature from classical control (e.g. PID control, cascade control, and ratio control), rule-based control (e.g. knowledge-based control and fuzzy control), model based control (e.g. optimal control, MPC, and adaptive control), to plant-wide control. For a thorough review of the techniques, the reader is referred to Weijers (2000). Also, state-of-the-art of control methods and practical cases for wastewater treatment are presented in Olsson et al. (2005). Some examples of control strategies for ASPs are given in Lindberg (1997), Lukasse, Keesman, Klapwijk & van Straten (1999), and Samuelsson (2005).

In this work, MPC is chosen for advanced control of ASPs. MPC is by far the most commonly applied advanced control technique in the chemical process industry (Bequette 2003). The reasons why MPC looks appropriate for control of ASPs are:

- MPC handles constraints in an optimal fashion.

- The MPC algorithm allows us to anticipate and remove the effects of disturbances. Indeed, disturbances are a major reason why control is required for ASPs.
- The optimization-based procedure is a natural way of handling multivariable control problems.
- MPC provides a general framework for complicated processes with time-delays, inverse-responses, and interactions.

In this chapter, control strategies of the aeration system for the intermittently aerated ASP introduced in section 2.2 are studied. We consider both simple rule-based controllers in section 5.2 and advanced model-based controllers (e.g. optimal control and MPC) in section 5.3. In the application of MPC, we first choose to study linear MPC for a nitrogen removal ASP with continuous aeration in subsection 5.3.1, which provides us with guidelines for further MPC development. Next, nonlinear MPC is considered due to the high nonlinearity of ASPs and discussed in subsection 5.3.2.

5.2 Rule-based control strategies

In practice, many rule-based operating modes have been investigated and applied to intermittently aerated ASPs, and can be categorized as

- open-loop control using fixed lengths of aerobic and anoxic phases, or
- feedback control using varying phase lengths by establishing a switch point for each phase. The switch point is typically the concentration of dissolved oxygen, nitrate, or ammonia.

These rule-based control strategies are often used for evaluation of process performance (Hao & Huang 1996), (Kimochi et al. 1998) and for comparison with advanced model-based control strategies (Isaacs & Thornberg 1998), (Lukasse, Keesman, Klapwijk & van Straten 1999), (Chachuat et al. 2005).

In Paper F, three rule-based strategies are presented for the nitrogen removal process with intermittent aeration:

- fixed lengths of aerobic and anoxic phases,
- relay control of the blower (on/off) state by the upper level of dissolved oxygen (DO) whilst the aeration cycle length is fixed, and
- relay control of the blower state by the upper and lower bounds of ammonia concentration (S_{NH_4}).

Another widely-used strategy is relay control by both the upper and lower bounds of DO, which is the approach applied to the laboratory process in chapter 6. The reactor is then intermittently aerated to restrict the value of DO between DO_{min} and DO_{max} . By the

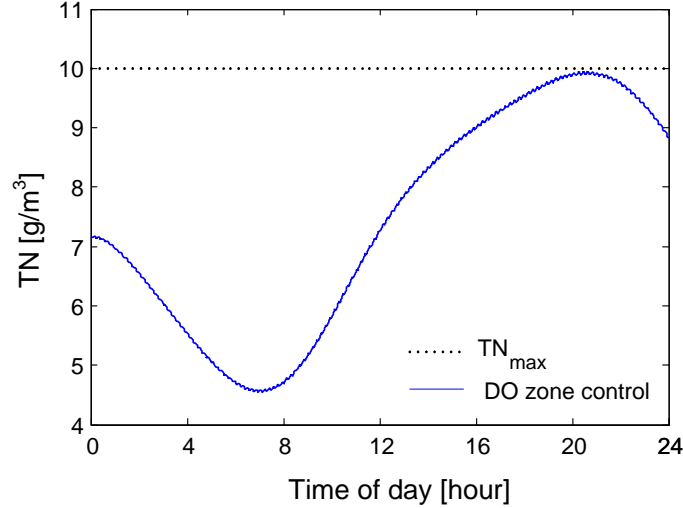


Figure 5.1: Simulation result of the DO zone control strategy: TN in the effluent.

simulation study, the reasonable values of DO_{\min}^1 and DO_{\max} are found to be 0.02 g m^{-3} and 1.1 g m^{-3} , which results in the total aeration fraction of 53.95%. The effluent TN obtained with the DO zone control strategy is shown in Figure 5.1. The controller has a similar performance as the fixed phase length controller and the DO upper level controller shown in Paper F, but the aeration switching frequency is higher (the number of aeration cycles $N_c = 129 \text{ d}^{-1}$).

Common to all these strategies is that the control variable is selected so as to satisfy the following criterion over a 24 h control horizon:

$$\begin{aligned} \min J &= \frac{\sum d}{T_h} \\ \text{subject to } \frac{dx_a}{dt} &= \mathcal{F}(x_a, u_a) \\ 0 &\leq \text{TN}_{\max} - \text{TN}. \end{aligned}$$

The objective is to determine the aeration profile that minimizes the aeration fraction, i.e., the total aeration time ($\sum d$) divided by the control horizon (T_h), with adherence to the effluent limit for the concentration of total nitrogen (TN). The objective function J is subject to the nonlinear process model $\frac{dx_a}{dt} = \mathcal{F}(x_a, u_a)$. The rule-based control strategies handle the usual daily variations of influent flowrate and TN load quite well and keep the effluent TN within the constraint during the 24 h control horizon. However, the rule-based approaches scheduled on a daily basis, are limited in that predetermined time schedules of aeration cycles with fixed phase lengths or fixed values of switch points cannot compensate for unanticipated loading variations (see also Potter et al. (1996)).

¹In practice, the accuracy of the DO measurement may not be guaranteed when the value of DO is getting too low. Thus, the tuning of the DO levels should be performed carefully.

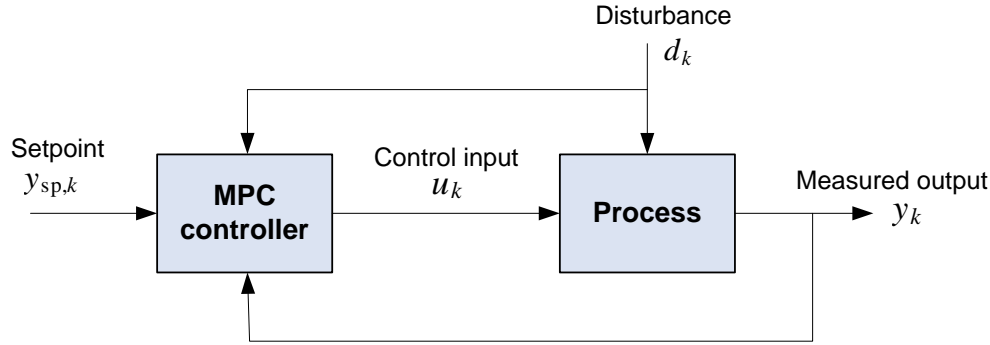


Figure 5.2: Predictive control scheme.

Next, the more advanced MPC controller will be discussed with comparison to the rule-based controllers.

5.3 Model predictive control (MPC)

The development of mathematical models for ASPs motivates the application of model-based optimal control, see e.g. Kabouris & Georgakakos (1990), Zhao et al. (1994b), Isaacs (1997), Kim et al. (2000). However, the open-loop optimal control should not be implemented in practice because the open-loop operation is very sensitive to model error and unknown disturbances. The ideal solution is to apply the dynamic optimization on-line with a receding horizon and include feedback in MPC. Figure 5.2 shows the block diagram of MPC. Readers not familiar with MPC may consult one of the many texts on the subject, e.g. Maciejowski (2002), Rossiter (2003), Qin & Badgwell (2003), and Camacho & Bordons (2004).

Applications of optimal control and MPC in intermittently aerated ASPs has been the subject of a number of research studies. The MPC technique is studied in Zhao et al. (1994a), Lukasse et al. (1998), Lukasse, Keesman & van Straten (1999a), and Holenda et al. (2007) using simplified models. In recent research, for example Chachuat et al. (2001), Chachuat et al. (2005) and Balku & Berber (2006), some efficient optimization algorithms are investigated for alternating aerobic-anoxic processes based on complete ASMs. In addition, Balku (2007) presents a comparison between conventional and intermittently aerated systems in terms of nitrogen removal and aeration time with an optimization algorithm.

For a complicated system like the ASP, a simple model is only possible by neglecting large parts of the process dynamics. Large simplifications of dynamics result in poor predictive capability of the model, and the simplified model is often only suitable for the specific process. The main hindrance for the implementation of MPC in wastewater treatment has always been the large computational demand for solving the optimization problem in MPC. However, with the increase in speed of computing hardware and improvements in optimization algorithms, interest in application of MPC using complete ASMs is growing.

5.3.1 Linear MPC for continuously aerated ASP

Application of MPC for a SNdN process with continuous aeration is investigated in Paper D. Linear MPC with constraints, which is a standard technique, is applied as a starting point of the MPC study. For linear MPC, a typical mathematical formulation of the criterion may be, see Hauge (2003):

$$\min_{u_k, \dots, u_{k+N}} J_k = \frac{1}{2} \sum_{i=0}^N (e_{k+i}^T Q e_{k+i} + u_{k+i}^T R u_{k+i} + \Delta u_{k+i}^T S \Delta u_{k+i}), \quad (5.1)$$

constrained by the definitions

$$\begin{aligned} e_k &= r_k - y_k \\ \Delta u_k &= u_k - u_{k-1}, \end{aligned}$$

the linear model of the process, e.g.

$$\begin{aligned} x_{k+1} &= Fx_k + Bu_k \\ y_k &= Hx_k + Du_k, \end{aligned}$$

and the bounds

$$\begin{aligned} u_k^\ell &\leq u_k \leq u_k^u \\ y_k^\ell &\leq y_k \leq y_k^u \\ \Delta u_k^\ell &\leq \Delta u_k \leq \Delta u_k^u, \end{aligned}$$

in which u , y , r , and x are the control input, controlled output, output reference, and state, respectively. The optimization horizon N and the weighting matrices Q , R , and S are the tuning parameters of MPC. The use of the linear model together with the quadratic objective function gives rise to a convex problem (Quadratic Programming, QP) whose solution is well studied with many commercial products available (e.g. the QP solver `quadprog` of the Optimization Toolbox in Matlab).

The state-space model for the controller design in Paper D is generated by the linearization of the process model (based on the ASM3 model) at a steady-state operating point of the plant. The steady-state is reached by assuming constant influent flowrate and concentrations, and the reactor condition at the steady-state is also considered initial conditions for the optimization.

The objective of the controller is to regulate the concentration of total nitrogen (TN) around a specified set-point by operating on the aeration system. Since the oxygen transfer process can be described with the oxygen mass transfer coefficient ($K_L a$), we choose $K_L a$ as the manipulated input, which is constrained to $K_L a \in [0, 240] \text{ d}^{-1}$. Moreover, no process and measurement noises are taken in consideration.

The simulation result shows that the performance of the MPC controller is satisfactory: the concentration of TN is controlled to track its set-point with good convergence properties.

5.3.2 Nonlinear MPC for intermittently aerated ASP

The main advantage of nonlinear MPC over linear MPC is obviously the possibility of dealing with nonlinear dynamics. In fact, the extension of linear MPC ideas to nonlinear processes is straightforward, at least conceptually. However, there are some difficulties derived from the use of nonlinear models such as (Camacho & Bordons 2004):

- The optimization problem is nonconvex, and its resolution is much more difficult than the QP problem.
- The difficulty of the optimization problem translates into an important increase in computation time.
- Nonlinear system identification is more difficult than the linear case, and applications of efficient identification techniques, as described in section 4.2, need to be investigated.

Some of these problems have partially been solved, and nonlinear MPC is becoming a field of intense research, see Henson (1998) and Kouvaritakis & Cannon (2001).

For the characteristics of ASPs, application of nonlinear MPC would seem essential for improved performance. There are two main reasons for this: first, ASPs are usually highly nonlinear. Secondly, ASPs are hardly ever in steady state, but is subject to transient behavior all the time, because the wastewater influent typically varies both in its flowrate and in its composition. In addition, by controlling reaction times of nitrifying and denitrifying in a single tank, the process offers energy savings and low investments of instruments.

The most straightforward approach to the nonlinear MPC problem is to retain the formulation of the optimization problem as we have discussed in linear MPC, simply replacing the linear model by a nonlinear one, and rely on suitable optimization methods. Therefore, the optimization problem for the aeration control of the intermittently aerated ASP is formulated as

$$\min_{f_1, \dots, f_{N_c}} J = \frac{1}{N_c} \sum_{k=1}^{N_c} f_k,$$

constrained by the nonlinear process model

$$\frac{dx_a}{dt} = \mathcal{F}(x_a, u_a)$$

and the bounds

$$\begin{aligned} 0 &\leq \text{TN}_{\max} - \text{TN} \\ \frac{t_{\text{on},\min}}{a} &\leq f_k \leq \frac{t_{\text{on},\max}}{a} \\ 1 - \frac{t_{\text{off},\max}}{a} &\leq f_k \leq 1 - \frac{t_{\text{off},\min}}{a}. \end{aligned}$$

Here, we consider the value of $K_L a$ constant, instead, the aeration fraction f_k is defined as the manipulated input. Also, f_k is constrained to its lower and upper limits depending on maximum and minimum air on/off periods, while the output variable (TN) is controlled to be less than its limit TN_{\max} . As the defined problem is a constrained nonlinear optimization

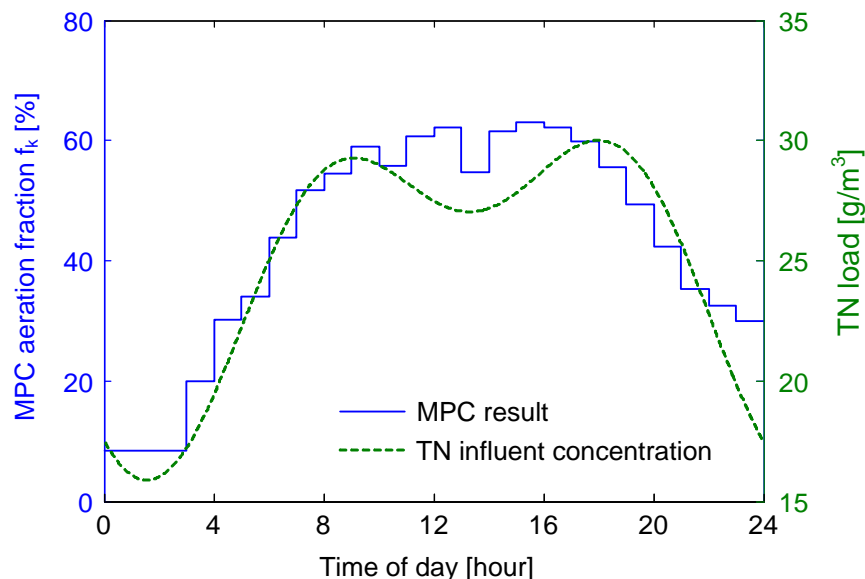


Figure 5.3: MPC aeration fraction and TN concentration in the influent vs. time.

problem, several powerful commercial solvers (e.g. the `SNOPT` and the `conSolve` solvers in the TOMLAB optimization toolbox for Matlab) can be used to solve this kind of problem.

The most important source of disturbances is the wastewater influent as described in section 4.4. In order to test the MPC performance under process disturbances, in Paper F, typical daily variations of dry weather conditions are simulated using some weighting functions defined in the literature for the influent wastewater flowrate and composition. Figure 5.3 illustrates that the variation of the aeration fraction obtained with MPC is closely related to the variation of the feed concentration: the high TN load induces the high aeration fraction, whereas the low TN results in the low aeration fraction. In addition, the simulation over a long-term period (10 d) shows that the MPC controller guarantees that the strict effluent requirements are fulfilled and the energy saving is lasting.

Finally, to confirm the benefits that may be realized using the proposed MPC strategy, comparisons with conventional rule-based controllers introduced in section 5.2 is considered in Paper F, and the MPC controller presents the following advantages:

- MPC provides most reduction of the aeration time.
- It is easier to deal with the inequality constraints in both effluent requirements and operating conditions.
- The tuning of the MPC controller is relatively easy; in this study, the main tuning parameter is the MPC horizon.

In the computation, MPC takes 22 min for a 24 h control horizon. Although this computation time is much longer than the other simple controllers, it is still feasible for the on-line implementation of the MPC controller.

Chapter 6

Laboratory Case Study

A new laboratory-scale plant was built by an N-removal project group at Telemark University College in 2007¹. Unfortunately, the author could only participate in the design and the plant construct phases of the project due to the time limit of the PhD research period. In this chapter, an introduction to the laboratory-scale plant is given, and then some possible future research directions for the project are suggested.

6.1 Experimental set-up

The N-removal plant is schematically depicted in Figure 6.1. Experiments are performed on two different types of reactors: the biofilm reactor² (Reactor 1 in the figure) and the ASP reactor (Reactor 2 in the figure).

Reactor system

The reactors are made of Plexiglass, and each reactor has a working volume of 10l, with 7.7l of the reactor volume and 2.3l of the settler volume. The settled sludge in the settler is recycled internally. For both reactors, the sludge retention time (SRT) is maintained at 12 ~ 24 d, i.e., the wasted sludge flowrate is $Q_w \in [0.32, 0.64]$ l/d. The hydraulic retention time (HRT) is chosen to be $HRT \in [12, 24]$ h, which results in the influent flowrate of $Q_{in} \in [7.7, 15.4]$ l/d.

Sludge and feed wastewater

The seeding sludge for the laboratory reactors was taken from the WWTP of city of Risør, Norway.

¹The project group included a number of professors, researchers, and master students at Telemark University College (TUC), as well as two visiting researchers from China. They are professor Rune Bakke, professor Bernt Lie, lecturer Morten Pedersen, associate professor Finn Haugen, PhD fellow Deshai Both-eju, Yuan Song, Shadi Lotfi, and Lei Zhao. Visiting researchers Dr. Zhiqiang Chen from Harbin Institute of Technology, China, and professor Tao He from Hubei University of Technology, China, has also given tremendous help and many suggestions during their stay at TUC.

²For the biofilm reactor system, a new model is required.

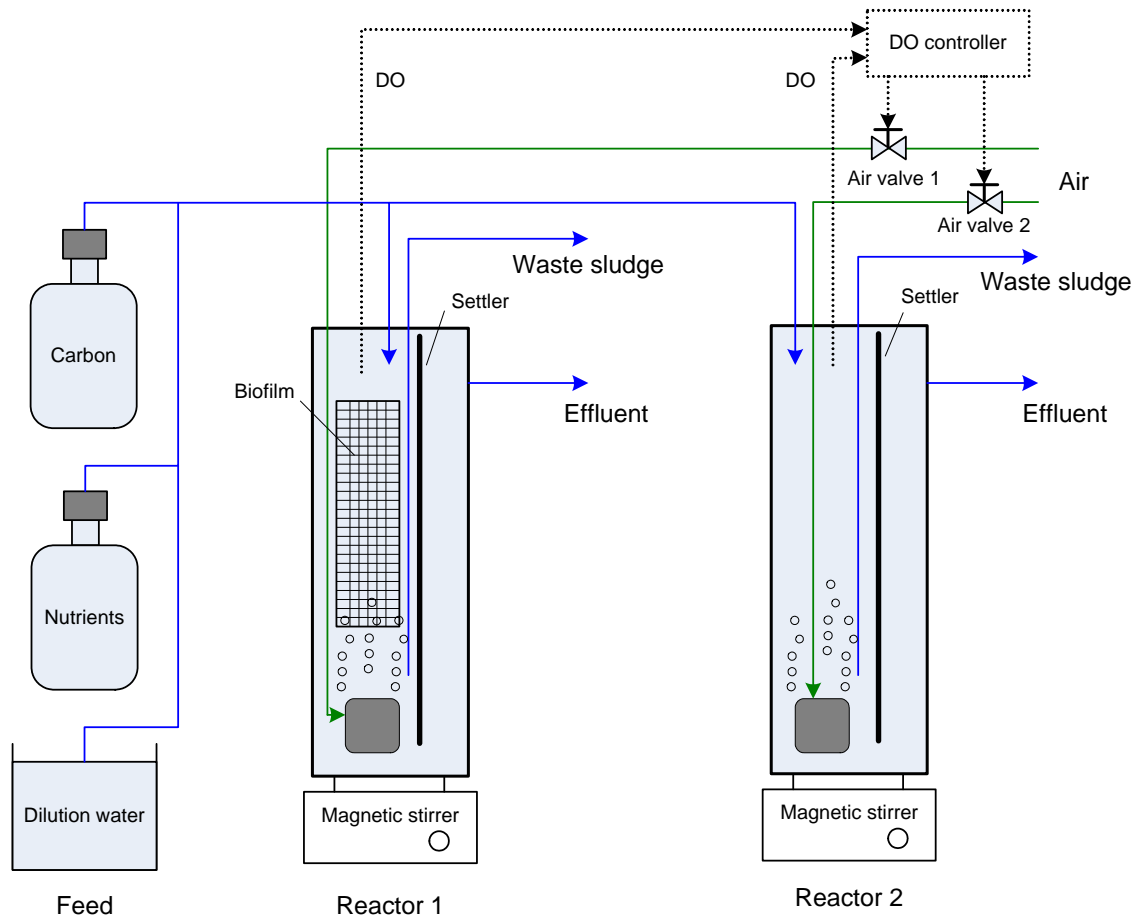


Figure 6.1: Schematic diagram of the laboratory N-removal plant.

Table 6.1: Composition and flowrate of the influent feed for the laboratory study.

Feed source	Composition	Concentration	Flowrate
Nutrient (nitrogen) source			0.21/ d
	(NH ₄) ₂ SO ₄	4438 mg/ l	
	K ₂ HPO ₄	344 mg/ l	
	KH ₂ PO ₄	344 mg/ l	
	Na HCO ₃	7445 mg/ l	
Carbon source			0.171/ d
	Starch (potatoes)	3899 mg/ l	
	Yeast extract	100 mg/ l	
Dilution water			8.641/ d
Total feed flowrate is 9.011/ d.			

The influent feed consists of nutrient source, carbon source, and dilution water. In order to reduce replenishment times for the influent feed, the nutrient and carbon sources are concentrated, and they are stored in two 10 l glass reservoirs. The wastewater in the reservoirs is transferred to the reactors by a peristaltic pump (Cole-Parmer Masterflex 7520-47) that can be controlled on/off from the PC. An example of the composition and the flowrate of the feed wastewater is given in Table 6.1. In this case, the feed wastewater consists of total chemical oxygen demand (COD) of 74 mg/ l, total nitrogen (TN) of 21 mg/ l, and total phosphorus (TP) of 3 mg/ l.

Data acquisition and control system

The data acquisition system used in this project consists of a computer (2.8 GHz Pentium 4, 512 MB RAM) with an interface card (National Instruments PCI-6024E), a terminal panel, and a user interface software. Signals of the DO sensors (Oxi340 and Oxi340i models, WTW) are sent to the computer through the series port connection. The electric control valves (5821 model, SAMSON) are used to regulate air supply to the reactors, by receiving signals from the controller designed in the control software (LabVIEW version 8.2). The reactors are aerated by intermittent aeration strategies. That is, the air valves are controlled by the upper and lower bounds of DO based on the on-line measurement of DO: the aeration is turned on when $DO \geq DO_{\max}$ and is turned off when $DO \leq DO_{\min}$. The DO sensors have a 1 min time delay and 1 min sampling time.

LabVIEW is also used for real-time database management and real-time display of process variables, as shown in Figure 6.2.

Off-line measurements

The influent wastewater samples are collected from the outlet of the feed pipes. Effluent samples for each reactor are taken from the respective effluent outlet. Process performance variables are routinely analyzed, included TN, COD, ammonium (S_{NH_4}), nitrite (S_{NO_2}), nitrate (S_{NO_3}), and total suspended solids (X_{SS}). In addition, pH and temperature are also recorded.

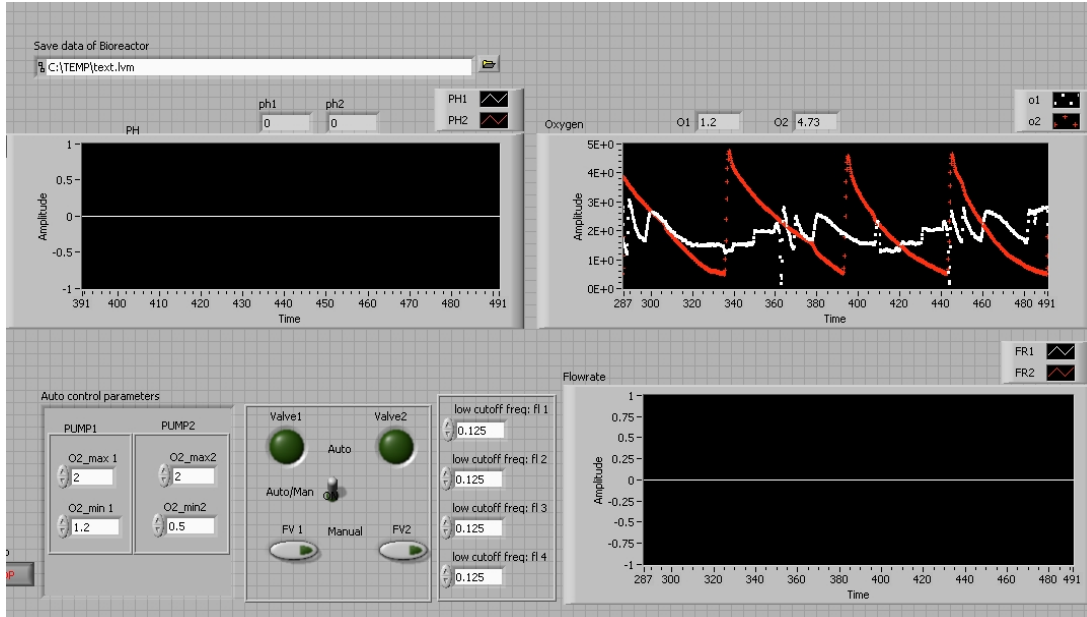


Figure 6.2: Controller screen of the LabVIEW user interface; under development.

6.2 Recommendations to the N-removal project

Based on the current research results described in this thesis, recommendations for future research directions of the N-removal project are given, with emphasis on the activated sludge reactor.

Modeling

1. First of all, the model of the N-removal process should be developed based on the ASM3 model which may be the best choice in the modeling of the biological processes. The object-oriented modeling approach (see section 2.4) can provide flexible and efficient model representation.
2. Once a process model is chosen, the model needs to be validated, not only in a qualitative but also in a quantitative way by comparing its predictive capabilities with data measured from the real system. The systematic approach to parameter identifiability and estimation, which is presented in chapter 4, can be applied.
3. The developed model can be used for many different purposes, for instance, system analysis by simulation, process optimization, state estimator, process control, see Dueñas Díez (2004).
4. As only few of the ASM3 components in the feed influent are available by measurement, the unmeasured feed composition can be estimated using the approach proposed in section 4.4.

Process control

1. The performance of the zone control strategy based on the available DO measurement, may be improved by considering the time delay of DO sensors.
2. The linear and the nonlinear MPC algorithms described in section 5.3 can be implemented for the ASP to achieve on-line application of the model-based optimal control.
3. When testing MPC on the laboratory plant, the current states must be known. Since we cannot measure all the states, the state estimation techniques discussed in chapter 3 can be used to estimate the model states.

In order to implement the aforementioned modeling and process control issues, it is of interest to have an open implementation architecture for efficiently integrating control algorithms and software tools.

6.3 Implementation architecture — future perspective

In the EU project INCOOP (INtegrated process COntrol and plantwide OPtimization)³, an architecture for advanced, model based control is proposed, see Figure 6.3. The reference architecture is developed to give a good software development environment where process control and instrumentation system, process simulator (e.g. gPROMS), optimizer, and algorithm prototyping environment (e.g. Matlab) can be integrated. The main advantages of the architecture are:

- The architecture is set up in a distributed modular structure, and various modules can easily be added, modified, or removed.
- A powerful computing tool, such as Matlab, is used to achieve the freedom in algorithmic development.
- Open communication standards make the integration of different tools easy.

In the wastewater treatment area, Olsson & Jeppsson (2006) point out the motivation of using open architectures for process control:

The controllers, I/O and buses are built to accommodate third-party modules and to integrate easily for fast setup. This feature gives facilities the freedom to choose the best and latest technologies from manufacturers without having to stay within product lines or fear product obsolescence.

³The INCOOP project is an RTD project carried out from 2000–2003 aiming at full integration and realization of the model based control and real time optimization in the process industry. Researchers from chemical process industries (Bayer, Shell International Chemicals, IPCOS Technology, MDC Technology) and academia (Eindhoven University of Technology, Delft University of Technology, RWTH Aachen, and Carnegie Mellon University) have participated in this project; see http://www.lpt.rwth-aachen.de/Research/Completed/co_incoop.php.

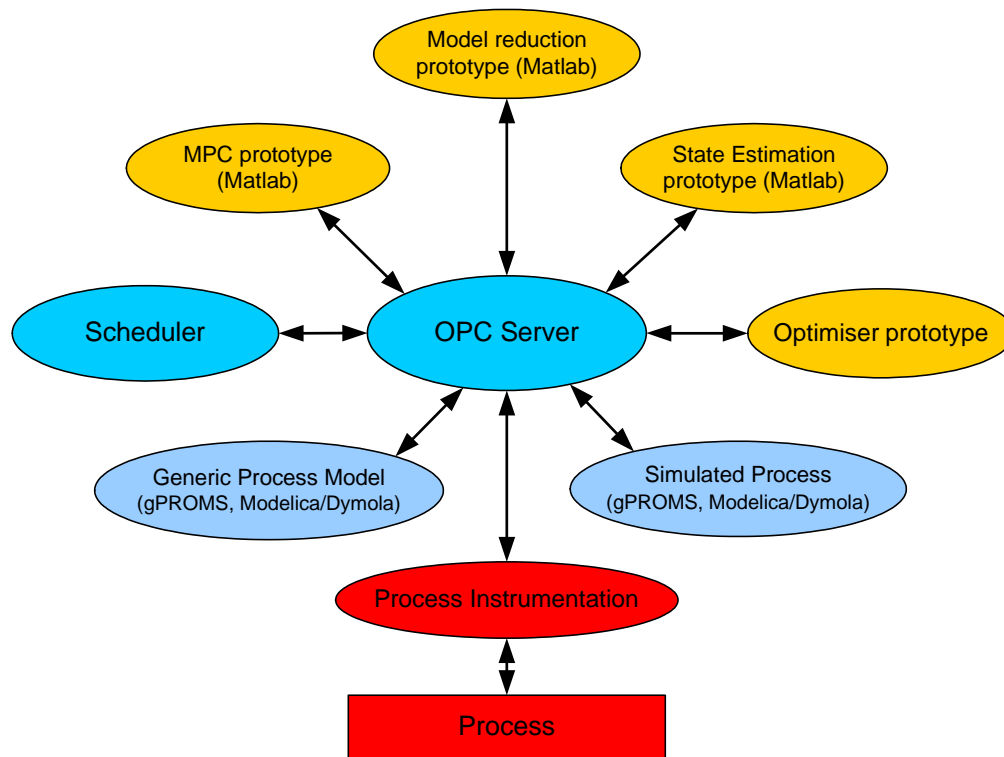


Figure 6.3: Software architecture for advanced control. Redrawing of figure from INCOOP workshop (INCOOP 2007).

The INCOOP architecture provides us with a methodology for efficiently integrating a flexible object-oriented modeling approach with advanced control techniques for WWTPs. Let us comment on each of the modules of the architecture.

OPC server is used as a key component for building the software environment. As an open connectivity via open standards, OPC (OLE for process control) is widely used in industrial automation. Most of control systems offer OPC servers in order to openly communicate with any OPC client, see <http://www.opcfoundation.org>. In addition, Matlab provides the OPC Toolbox that is an OPC Data Access client application, capable of connecting to any OPC Data Access compliant server (e.g. National Instruments LabVIEW OPC server (Chai 2005)). Dymola version 6.1 also provides an interface to OPC Data Access.

Scheduler is used to coordinate the various tools.

Generic process model is important for model-based control and optimization. In INCOOP, it is proposed to use gPROMS (PSE 2006); the choice is based on several characteristics for gPROMS, such as strong support for process models and an equation based system which makes it possible to introduce manipulation of the models (e.g. analytic model reduction, linearization). Another general, equation-based modeling language is Modelica. Currently, Modelica has less support for process models than gPROMS, but the modeling scope is more general.

Simulated process indicates the same equation-based simulation tools (i.e., gPROMS and Modelica/Dymola).

Model reduction includes elimination of nonlinear models, elimination of implicit expressions, reduction of the number of states in the model, etc. The approximate model should be suited specially for the application in the optimization and control modules (modules *Optimizer prototype* and *MPC prototype (Matlab)* in Figure 6.3).

Optimiser prototype is required to generate optimal control trajectories for the process. Efficient algorithms for optimization are necessary to enable the optimization techniques to be applied on-line, and software tools are often used to solve optimization problems.

State estimation prototype is used to find unmeasured states since optimization of a system requires that the states of the system are available. The common techniques for state estimation have been discussed in chapter 3.

MPC prototype is developed to control the process based on the optimized trajectories obtained by optimizer. MPC predicts the future optimal control actions and send them to the plant that includes a primary control system.

Process instrumentation is a tool for communication with the process, which enables the use of data from the process and puts out actuator signals to the process.

In this research, Matlab is used as the integration platform due to its support to Dymola and powerful numerical capabilities. Recently, LabVIEW version 8.20 adds a functionality to support the communication with Dymola. With the development of the interface between LabVIEW and Dymola/Modelica, the use of LabVIEW with MathScript⁴ could be an excellent integration platform. Also, the open source programming language Python described in appendix 2.A, could replace Matlab in a “poor man solution”.

⁴MathScript is a new feature for LabVIEW 8 that adds math-oriented, textual programming to LabVIEW, and is generally compatible with m-file script syntax used by e.g. Matlab. For more information see <http://www.ni.com>.

Chapter 7

Conclusions and Future Directions

7.1 Conclusions

The main objective of this thesis has been to investigate possible improvements in operation and performance of activated sludge plants through applications of advanced modeling and control techniques. The activated sludge process (ASP) is generally applied in biological wastewater treatment due to its efficient elimination of organic matter and nutrients such as nitrogen and phosphorus. Although ASPs have been introduced in a large number of studies, there are still many challenges associated with the application of modern modeling methods and control strategies for such processes.

Modeling

Special emphasis has been put on the investigation of approaches that implement efficiently the complex activated sludge plant models. A laboratory-scale nitrogen removal process and a full-scale phosphorus removal plant: the Duvbacken WWTP, have been used as case studies. For ASPs, the family of Activated Sludge Models (ASMs) presented by the IWA are generally accepted as the standard models of ASPs. The models of the N-removal process and the Duvbacken WWTP have then been developed based on the ASM3 model and the ASM2d model, respectively. The object-oriented modeling approach providing open component-based architectures and reuse of components, is suggested in this thesis in order to achieve more flexible model representation. The object-oriented modeling language Modelica with one of its commercial simulation environments, Dymola, is chosen as model representation tool for activated sludge systems. Special consideration has been given to the Dymola interfaces to other programs, in order to build a good software development environment. Two Dymola interfaces are described: the Dymola-Matlab interface provided by Dymola and the Dymola-Python interface developed by the author, where both Matlab and Python are powerful algorithm prototyping environments for advanced control. Within the Dymola simulation environment, the models of the N-removal process and the Duvbacken WWTP are easily constructed by using the free Modelica library `WasteWater`. The applications of the modeling approach in this thesis have shown that the combination of Modelica/Dymola for efficient object-oriented modeling and Matlab for further analysis is a powerful tool for control system design. However, currently, information transfer in Dymola-Matlab interaction is via

files. It would presumably dramatically decrease the execution time if the simulator could be connected to Matlab as a function.

State estimation

As regards state estimation and parameter estimation, the attention is focused on the application of the different estimation techniques to ASPs. An overview of the most generally applied techniques for state estimation such as *standard Kalman filter (KF)*, *extended Kalman filter (EKF)*, *unscented Kalman filter (UKF)*, *square-root unscented Kalman filter (SRUKF)*, and *moving horizon estimation (MHE)*, has been given, with consideration of systematic presentation of the algorithms for easy implementation. Although these techniques have been widely studied in the literature, the performances of the filters for the complete ASM models have seldom been discussed. In this thesis, the nonlinear estimation techniques are compared for state estimation of the N-removal process. The comparison results show that the discrete-time EKF gives the fastest performance with a satisfactory estimation result. The slightly variant version of the discrete-time EKF: the hybrid EKF, produces slightly lower estimation errors than the discrete-time EKF due to the more accurate time updates of the state and covariance. Among all the estimation techniques, the SRUKF improves numerical properties of the UKF and provides best estimation performance; and one main advantage of the unscented filters is the fact that no Jacobian matrices of the system need to be calculated. To handle state constraints, the MHE estimator is most straightforward and generally preferred. When inequality constraints are not considered, the MHE with a horizon size of 1 and with the update of the initial state estimation and the covariance as in the EKF, gives equivalent results to that of the EKF.

Parameter estimation

The large variations of plant operation, influent, and sludge characteristics of ASPs require some of the model parameters to be adjusted in order to ensure good predictive capabilities of the model. However, parameter estimation of ASP models is a challenging task, since ASP models are in general large-scale, nonlinear, and overparameterized. In this thesis, parameter estimation is carried out for both kinetic parameters of the ASM model and compositions in the influent which is considered model disturbances from the process control point of view.

A procedure for parameter identifiability analysis is discussed and tested with the Duvbacken model. Such a method proves to be very useful, since it reduces a large number of parameters to a small subset of parameters that can be identified from the available experimental data. Estimation of the kinetic parameters is carried out using the least squares method, based on real data taken from the industrial plant. After parameter estimation, the fitting of the model to a new set of experimental data is not great but satisfactory, especially for those outputs which are important to the plant operation (e.g. S_{PO_4} and TP).

Since a good characterization of the main compositions in the influent is very important to the performance of the ASM models, a method for simultaneously estimating both model state and influent composition has been developed using the discrete-time EKF. A method to study observability for state and disturbance estimation is discussed. The EKF-based approach is illustrated with the Duvbacken process; however, the Duvbacken system

is unobservable due to lack of measurements. As an initial study before doing state and disturbance estimation based on real data from the plant, we check whether suitable estimates can be obtained for the observable process based on responses from the simulation model; and this study shows the estimator has a good convergence property. Next, when the filter is evaluated based on the experimental data from the real plant, the performance of the filter is relatively poor. As argued, the lack of the necessary measurements renders the system unobservable. The use of the additional measurements, such as S_{PO_4} , S_{I} , S_{N_2} , S_{ALK} , X_{I} , and X_{TSS} in anaerobe reactor, S_{O_2} and S_{PO_4} in aerobe reactor 1, S_{NH_4} and S_{PO_4} in aerobe reactor 2, and S_{NO_3} in aerobe reactor 3, may improve the estimation of states and disturbances, and hopefully the development of measuring technology would give us probabilities to obtain all the necessary measurements. In addition, since the real value of the disturbance is not known, it makes the initialization and tuning of the EKF more difficult. Although the estimation capability of the filter is limited, the application still demonstrates that the proposed approach provides us with a general methodology to estimate unmeasured states and disturbances using a number of available measurements.

Process control

The development towards tighter effluent requirements and better dynamic models have motivated the application of advanced control to ASPs, where the focus has been put on the MPC algorithm because it provides an optimization-based procedure with an elegant way of handling constraints and with capabilities of removing process disturbances. In this thesis, the MPC algorithm has been evaluated with the N-removal process: first of all, a linear MPC controller is developed to regulate the concentration of total nitrogen (TN) around a specified set-point by operating on the *continuous* aeration system. Secondly, the MPC technique using the nonlinear process model is investigated and applied to controlling the *intermittent* aeration system. The aim of the control is to design an aeration strategy (air-on and air-off periods) which minimizes the energy consumption of the aeration system, with adherence to the EU effluent standards and the operating constraints. The simulation over a long-term period (10 d) shows that the nonlinear MPC controller guarantees that the strict effluent requirements are fulfilled and the energy saving is lasting.

Finally, to confirm the benefits that may be realized using the proposed nonlinear MPC strategy, comparisons with conventional rule-based controllers used in practice and in the literature, have been carried out, where MPC had the most reduction of the aeration time/energy consumption and offered the easiest way to handle the inequality constraints. Based on the results presented in this thesis, it can be concluded that the application of the nonlinear MPC to the N-removal process is promising.

7.2 Future directions

In this work, the particular emphasis has been to study advanced modeling and control techniques for improving performances of ASPs. Although the modeling method and control strategies are evaluated with simulation studies, the implementation of them is straightforward. It is thus of great interest to test the proposed approaches to the laboratory plant and

eventually to real full-scale WWTPs. Based on the current research results, the most obvious expectation for future work which might hopefully be carried out by research activities at Telemark University College (TUC), are that:

- Model development of the activated sludge plant will be improved, with particular emphasis on modeling of settling and clarification which is an important element in determining the overall process dynamics.
- Sensitivity and correlation problems depend strongly on the experiments performed, the measurements taken, and the experimental conditions. Thus, careful experimental designing and reasonable implementation of experiments would help us get promising estimation of model parameters.
- A more general overview of state estimation techniques will be given. Other estimation techniques such as ensemble Kalman filter and particle filter, will also be evaluated. In this work, the MHE has been tested with a horizon size of 1; clearly, it will be of interest to study the behavior of the MHE using a larger horizon size.
- A general model for describing dynamics of ASP disturbances will be developed. Since the feed composition of domestic wastewater typically varies significantly during the course of a day, the process disturbance is a strong periodic signal. This requires a proper formulation (e.g. an oscillator model) to describe the periodic variation.
- When applying MPC to the real plant, the estimation techniques are expected to be combined with the control strategy to obtain the unmeasured current states or/and process disturbances, based on the available measurements.
- Advantages of advanced, model-based controllers over simple rule-based controllers will be studied further. Besides for nominal performance, different control techniques will be compared for robustness against model mismatch and large process disturbances.
- Since lack of cheap and reliable sensors for measurement is a big challenge in parameter and state estimation of WWTPs, investigation of simple and efficient software sensors will be of great interest. It is recommended to cooperate with the sensor technology group at TUC.

Bibliography

- Alcaraz-González, V., Harmand, J., Rapaport, A., Steyer, J., González-Alcaraz, V. & Pelayo-Ortiz, C. (2002), ‘Software sensors for highly uncertain wwtps: a new approach based on interval observers’, *Water Research* **36**, 2515–2524.
- Andrews, J. (1974), ‘Dynamic models and control strategies for wastewater treatment processes’, *Water Research* **8**, 261–289.
- Aubrun, C., Theilliol, D., Harmand, J. & Steyer, J. (2001), ‘Software sensor design for cod estimation in an anaerobic fluidized bed reactor’, *Water Science and Technology* **43**(7), 115–120.
- Baksaas, H. (1996), Analyse av avløpsrensaneanlegget på Esso raffineriet, Slagentangen for optimalisering, Master’s thesis, Telemark University College, P.O. Box 203, 3901 Porsgrunn, Norway.
- Balchen, J. & Mummé, K. (1988), *Process control: structures and applications*, ISBN 0-442-21155-4, Van Nostrand Reinhold, New York.
- Balku, S. (2007), ‘Comparison between alternating aerobic-anoxic and conventional activated sludge systems’, *Water Research* **41**, 2220–2228.
- Balku, S. & Berber, R. (2006), ‘Dynamics of an activated sludge process with nitrification and denitrification: Start-up simulation and optimization using evolutionary algorithm’, *Computers and Chemical Engineering* **30**, 490–499.
- Bastogne, T. (2004), ‘A multiport diagram for batch system modeling: Methodology and implementation’, *Simulation Modelling Practice and Theory* **12**, 425–449.
- Bequette, B. (2003), *Process control: modeling, design, and simulation*, ISBN 0-13-353640-8, Prentice Hall, New Jersey.
- Boaventura, K., Roqueiro, N., Coelho, M. & Araújo, O. (2001), ‘State observers for a biological wastewater nitrogen removal process in a sequential batch reactor’, *Bioresource Technology* **79**, 1–14.
- Bogaerts, P. & Wouwer, A. (2004), ‘Parameter identification for state estimation—application to bioprocess software sensors’, *Chemical Engineering Science* **59**, 2465–2476.

- Bortolin, G., Borg, S. & Gutman, P. (2004), ‘Modeling of the wet end part of a paper mill with dymola’, *Mathematics and Computers in Simulation* **65**, 31–38.
- Brosilow, C. & Joseph, B. (2002), *Techniques of model-based control*, ISBN 0-13-028078-X, Prentice Hall PTR, Upper Saddle River, New Jersey.
- Brouwer, H., Klapwijk, A. & Keesman, K. (1997), ‘Identification of activated sludge and wastewater characteristics using respirometric batch-experiments’, *Wat. Res.* **32**(4), 1240–1254.
- Brun, R., Kühni, M., Siegrist, H., Gujer, W. & Reichert, P. (2002), ‘Practical identifiability of ASM2d parameters - systematic selection and tuning of parameter subsets’, *Water Research* **36**(16), 4113–4127.
- Brun, R., Reichert, P. & Künsch, H. R. (2001), ‘Practical identifiability analysis of large environmental simulation models’, *Water Resour. Res.* **37**(4), 1015–1030.
- Bubnicki, Z. (2005), *Modern control theory*, ISBN 3-540-23951-0, Springer, Berlin.
- Camacho, E. & Bordons, C. (2004), *Model predictive control*, ISBN 1-85233-694-3, second edn, Springer, London.
- Chachuat, B., Roche, N. & Latifi, M. (2001), ‘Dynamic optimisation of small size wastewater treatment plants including nitrification and denitrification processes’, *Computers & Chemical Engineering* **25**, 585–593.
- Chachuat, B., Roche, N. & Latifi, M. (2005), ‘Long-term optimal aeration strategies for small-size alternating activated sludge treatment plants’, *Chemical Engineering and Processing* **44**, 593–606.
- Chai, Q. (2005), Application of the MATLAB OPC Toolbox. Technical report (unpublished), Telemark University College, Norway.
- Chen, C. (1999), *Linear system theory and design*, ISBN 0-19-511777-8, 3rd edn, Oxford university press, New York.
- Collivignarelli, C. & Bertanza, G. (1999), ‘Simultaneous nitrification-denitrification processes in activated sludge plants: performance and applicability’, *Water Science Technology* **40**(4–5), 187–194.
- Crassidis, J. & Junkins, J. (2004), *Optimal estimation of dynamic systems*, ISBN 1-58488-391-X, Chapman & Hall/CRC.
- Darby, M. & Nikolaou, M. (2007), ‘A parametric programming approach to moving-horizon state estimation’, *Automatica* **43**, 885–891.
- Darouach, M., Zasadzinski, M., Onana, A. & Nowakowski, S. (1995), ‘Kalman filtering with unknown inputs via optimal state estimation of singular systems’, *International Journal of Systems Science* **26**(10), 2015–2028. Available at <http://hal.archives-ouvertes.fr/hal-00143941/en/>, accessed 1 June 2007.

- Dempsey, M. (2003), Automatic translation of Simulink models into Modelica using Simelica and the AdvancedBlocks library, *in* P. Fritzson, ed., ‘Proceedings of the 3rd International Modelica Conference’, Linköping, Sweden, pp. 115–124.
- Dochain, D. (2003), ‘State and parameter estimation in chemical and biochemical processes: a tutorial’, *Journal of Process control* **13**, 801–818.
- Dochain, D. & Chen, L. (1992), ‘Local observability and controllability of stirred tank reactors’, *Journal of Process Control* **2**(3), 139–144.
- Dochain, D. & Vanrolleghem, P. (2001), *Dynamical modelling and estimation in wastewater treatment processes*, ISBN 1900222 50 7, IWA Publishing, London, UK.
- Dochain, D., Vanrolleghem, P. & van Daele, M. (1995), ‘Structural identifiability of biokinetic models of activated sludge respiration’, *Wat. Res.* **29**(11), 2571–2578.
- Droste, R. (1997), *Theory and practice of waste and wastewater treatment*, ISBN 0-471-12444-3, John Wiley & Sons, USA.
- Dueñas Díez, M. (2004), Population balance modeling and passivity-based control of particulate processes, applied to the Silgrain process, PhD thesis, Telemark University College, P.O. Box 203, N-3901 Porsgrunn, Norway.
- Dueñas Díez, M., Fjeld, M., Andersen, E. & Lie, B. (2006), ‘Validation of a compartmental population balance model of an industrial leaching process: The Silgrain(R) process’, *Chem. Engng. Sci.* **61**, 229–245.
- Dynasim (2004), *Dymola Multi-Engineering Modeling and Simulation: Dymola user manual*, Dynasim AB, Lund, Sweden.
- Dynasim (2007), ‘Dymola release notes’. <http://www.dynasim.se/update>, accessed 10 July 2007.
- Ferrero, C. (2005), Parameter estimation for model of biological wastewater treatment, Master’s thesis, Telemark University College, Porsgrunn, Norway.
- Fritzson, P. (2004), *Principles of Object-Oriented Modeling and Simulation with Modelica 2.1*, IEEE Press, Piscataway, NJ. ISBN 0-471-47163-1.
- Gauld, A. (2000), *Learn to program using Python*, ISBN 0-201-70938-4, Addison-Wesley, Upper Saddle River, NJ 07458.
- Gelb, A., ed. (1992), *Applied optimal estimation*, ISBN 0-262-20027-9, The M.I.T. Press, Massachusetts.
- Gernaey, K. V., Loosdrecht, M. C. V., Henze, M., Lind, M. & Jørgensen, S. B. (2004), ‘Activated sludge wastewater treatment plant modelling and simulation: State of the art’, *Environmental Modelling and Software* **19**, 763–783.

- Grewal, M. & Andrews, A. (2001), *Kalman filtering: theory and practice using MATLAB*, ISBN 0-471-39254-5, second edn, John Wiley & Sons, New York.
- Gujer, W., Henze, M., Mino, T., Matsuo, T., Wentzel, M. & Marais, G. (1995), 'The activated sludge model no. 2: Biological phosphorus removal', *Wat. Sci. Tech.* **31**(2), 1–11.
- Gujer, W., Henze, M., Mino, T. & van Loosdrecht, M. (1999), 'Activated Sludge Model No. 3', *Wat. Sci. Tech.* **39**(1), 183–193.
- Ha, Q. & Trinh, H. (2004), 'State and input simultaneous estimation for a class of nonlinear systems', *Automatica* **40**, 1779–1785.
- Hadj-Sadok, M. & Gouzé, J. (2001), 'Estimation of uncertain models of activated sludge processes with interval observers', *Journal of Process Control* **11**, 299–310.
- Hao, O. & Huang, J. (1996), 'Alternating aerobic-anoxic process for nitrogen removal: process evaluation', *Water Environ. Res.* **68**, 83–93.
- Haseltine, E. & Rawlings, J. (2005), 'Critical evaluation of extended Kalman filtering and moving-horizon estimation', *Eng. Chem. Res.* **44**, 2451–2460.
- Hauge, T. (2003), Roll-out of model based control with application to paper machines, PhD thesis, Telemark University College, Porsgrunn, Norway. ISBN 82-471-5581-8.
- Heck, B. S., Wills, L. M. & Vachtsevanos, G. J. (2003), 'Software technology for implementing reusable, distributed control systems', *IEEE Control Systems Magazine* **23** (1), 19–35.
- Henson, M. (1998), 'Nonlinear model predictive control: current status and future directions', *Computers and Chemical Engineering* **23**, 187–202.
- Henson, M. A. & Seborg, D. E. (1997), *Nonlinear Process Control*, Prentice Hall PTR, Upper Saddle River, NJ. ISBN 0-13-625179-X.
- Henze, M., Grady, C. J., Marais, W. G. G. & Matsuo, T. (1987), Activated Sludge Model No. 1, IAWPRC Scientific and Technical Report No. 1, London.
- Henze, M., Gujer, W., Mino, T., Matsuo, T., Wentzel, M. C., Marais, G. V. R. & van Loosdrecht, M. C. M. (1999), 'Activated Sludge Model No. 2d, ASM2d', *Wat. Sci. Tech.* **39**(1), 165–182.
- Henze, M., Gujer, W., van Loosdrecht, M. C. M. & Mino, T. (2000), *Activated Sludge Models ASM1, ASM2, ASM2d and ASM3.*, Vol. 9 of *Scientific and Technical Report*, IWA Publishing.
- Holenda, B., Domokos, E., Rédey, Á. & Fazakas, J. (2007), 'Dissolved oxygen control of the activated sludge wastewater treatment process using model predictive control', *Computers and Chemical Engineering*. In press.
- Hu, Z., Wentzel, M. & Ekama, G. (2003), 'Modelling biological nutrient removal activated sludge systems – a review', *Water Research* **37**, 3430–3444.

- INCOOP (2007), 'Referat fra INCOOP workshop 23-24 jan 2003'. http://www.nt.ntnu.no/users/skoge/prost_mail/latest/0004.html, accessed July 2007.
- Ip, S., Bridger, J. & Mills, N. (1987), 'Effect of alternating aerobic and anaerobic conditions on the economics of the activated sludge system', *Wat. Sci. Tech.* **19**(Rio), 911–918.
- Isaacs, S. (1997), 'Automatic adjustment of cycle length and aeration time for improved nitrogen removal in an alternating activated sludge process', *Wat. Sci. Tech.* **35**, 225–232.
- Isaacs, S. & Thornberg, D. (1998), 'A comparison between model and rule based control of a periodic activated sludge process', *Wat. Sci. Tech.* **37**(12), 343–351.
- Jenkins, D. (2007), 'Principle of biological phosphorus removal', Presentation on IWPCA 89th Annual Conference. Available at <http://www.iawpca.org>, accessed 31 January 2008.
- Jeppsson, U. (1996), Modelling aspects of wastewater treatment processes, PhD thesis, Lund Institute of Technology, Lund, Sweden. Available at <http://www.iea.lth.se/ielulf/publications/phd-thesis/PhD-thesis.pdf> (accessed July 2007).
- Julier, S. & Uhlmann, J. (2004), 'Unscented filtering and nonlinear estimation', *Proceedings of the IEEE* **92**(3), 401–422.
- Julier, S., Uhlmann, J. & Durrant-Whyte, H. (1995), 'A new approach for filtering nonlinear systems', *American control conference* pp. 1628–1632.
- Kabouris, J. & Georgakakos, A. (1990), 'Optimal control of the activated sludge process', *Wat. Res.* **24**(10), 1197–1208.
- Kandepu, R., Foss, B. & Imsland, L. (2008), 'Applying the unscented Kalman filter for nonlinear state estimation', *Journal of Process Control* **xxx**, xxx–xxx. In Press.
- Keesman, K. (2002), 'State and parameter estimation in biotechnical batch reactors', *Control Engineering Practice* **10**, 219–225.
- Kim, H., McAvoy, T., Anderson, J. & Hao, O. (2000), 'Control of an alternating aerobic-anoxic activated sludge system — part 2: optimization using a linearized model', *Control Engineering Practice* **8**, 279–289.
- Kimochi, Y., Inamori, Y., Mizuochi, M., Xu, K. & Matsumura, M. (1998), 'Nitrogen removal and n₂o emission in a full-scale domestic wastewater treatment plant with intermittent aeration', *Journal of Fermentation and Bioengineering* **86**(2), 202–206.
- Kiusalaas, J. (2005a), *Numerical methods in engineering with MATLAB*, ISBN 978-0-521-85288-3, Cambridge University Press, Cambridge.

- Kiusalaas, J. (2005*b*), *Numerical methods in engineering with Python*, ISBN 978-0-521-85287-6, Cambridge University Press, Cambridge.
- Kouvaritakis, B. & Cannon, M., eds (2001), *Nonlinear predictive control: theory and practice*, ISBN 0 85296 984 8, The Institute of Electrical Engineers, Stevenage, England.
- Kurtz, M. & Henson, M. (1998), ‘State and disturbance estimation for nonlinear systems affine in the unmeasured variables’, *Computers Chem. Engng* **22**, 1441–1459.
- Langtangen, H. P. (2006), *Python scripting for computational science*, ISBN 3-540-29415-5, second edn, Springer, Berlin.
- Lewis, F. (1986), *Optimal estimation*, ISBN 0-471-83741-5, Jon Wiley & Sons, Georgia.
- Lie, B., Dueñas Díez, M. & Hauge, T. A. (2005), ‘A comparison of implementation strategies for MPC’, *Modeling, Identification and Control* **26**, 39–50.
- Lindberg, C. (1997), Control and estimation strategies applied to the activated sludge process, PhD thesis, Uppsala University.
- Ljung, L. (1999), *System Identification. Theory for the User*, ISBN 0-13-656695-2, second edn, Prentice Hall, Upper Saddle River, New Jersey.
- Lukasse, L., Keesman, K., Klapwijk, A. & van Straten, G. (1998), ‘Optimal Control of N-removal in ASPs’, *Wat. Sci. Tech.* **38**, 255–262.
- Lukasse, L., Keesman, K., Klapwijk, A. & van Straten, G. (1999), ‘A comparison of NH₄/NO₃ control strategies for alternating activated sludge processes’, *Wat. Sci. Tech.* **39**, 93–102.
- Lukasse, L., Keesman, K. & van Straten, G. (1997), Identification for model predictive control of biotechnological processes, case study: nitrogen removal in an activated sludge process, in ‘Proc. 11th IFAC Symp. on Syst. Ident.’, Fukuoka, Japan, pp. 1525–1530.
- Lukasse, L., Keesman, K. & van Straten, G. (1999*a*), ‘l₁-norm optimal control of n-removal in an activated sludge process’, *Control Engineering Practice* **7**, 63–70.
- Lukasse, L., Keesman, K. & van Straten, G. (1999*b*), ‘A recursively identified model for short-term predictions of nh₄/no₃-concentrations in alternating activated sludge processes’, *Journal of Process Control* **9**, 87–100.
- Lund, B. & Foss, B. (2008), ‘Parameter ranking by orthogonalization—applied to nonlinear mechanistic models’, *Automatica* **44**, 278–281.
- Lunze, J. (1992), *Feedback control of large scale systems*, ISBN 0-13-318353-X, Practice Hall International, Salisbury, England.
- Lutz, M. & Ascher, D. (2003), *Learning Python*, ISBN 0-596-00281-5, second edn, O’Reilly Media.

- Maciejowski, J. (2002), *Predictive control: with constraints*, ISBN 0 201 39823 0 PPR, Prentice Hall, Harlow, England.
- Maffezzoni, C., Ferrarini, L. & Carpanzano, E. (1999), ‘Object-oriented models for advanced automation engineering’, *Control Engineering Practice* **7**, 957–968.
- MathWorks (2007), ‘MATLAB and Simulink for technical computing’. <http://www.mathworks.com>, accessed 10 July 2007.
- Metcalf & Eddy, I. (2003), *Wastewater Engineering. Treatment and Reuse*, fourth edn, McGraw Hill, New York.
- Nakajima, J. & Kaneko, M. (1991), ‘Practical performance of nitrogen removal in small-scale sewage treatment plants operated in intermittent aeration mode’, *Wat. Sci. Tech.* **23**(Kyoto), 709–718.
- Olsson, G. & Jeppsson, U. (2006), ‘Plant-wide control: dream, necessity or reality?’, *Wat. Sci. Tech.* **53**(3), 121–129.
- Olsson, G. & Newell, B. (1999), *Wastewater Treatment Systems. Modelling, Diagnosis and Control*, IWA Publishing, London.
- Olsson, G., Nielsen, M., Yuan, Z., Jensen, A. & Steyer, J. (2005), *Instrumentation, control and automation in wastewater systems*, ISBN 1900222833, IWA publishing, London.
- Perrier, M., de Azevedo, S., Ferreira, E. & Dochain, D. (2000), ‘Tuning of observer-based estimators: theory and application to on-line estimation of kinetic parameters’, *Control Engineering Practice* **8**, 377–388.
- Petersen, B. (2000), Calibration, identifiability and optimal experimental design of activated sludge models, PhD thesis, Department of Applied Mathematics Biometrics and Process Control (BIOMATH), University of Gent, Available at <https://archive.ugent.be/retrieve/738/19957980.pdf>, accessed 7 June 2007.
- Pina, L. & Botto, M. (2006), ‘Simultaneous state and input estimation of hybrid systems with unknown inputs’, *Automatica* **42**, 755–762.
- Plett, G. (2006a), ‘Sigma-point Kalman filtering for battery management systems of LiPB-based HEV battery packs — Part 1: introduction and state estimation’, *Journal of Power Sources* **161**, 1356–1368.
- Plett, G. (2006b), ‘Sigma-point Kalman filtering for battery management systems of LiPB-based HEV battery packs — Part 2: simultaneous state and parameter estimation’, *Journal of Power Sources* **161**, 1369–1384.
- Potter, T., Koopman, B. & Svoronos, S. (1996), ‘Optimization of a periodic biological process for nitrogen removal from wastewater’, *Water Research* **30**(1), 142–152.
- PSE (2006), ‘gPROMS’. <http://www.psenterprise.com/gproms/>, accessed 11 July 2007.

- Qin, S. & Badgwell, T. (2003), 'A survey of industrial model predictive control technology', *Control Engineering Practice* **11**, 733–764.
- Rawlings, J. & Bakshi, B. (2006), 'Particle filtering and moving horizon estimation', *Computer & Chemical Engineering* **30**, 1529–1541.
- Rawlings, J. & Ekerdt, J. (2002), *Chemical reactor analysis and design fundamentals*, ISBN 0-615-11884-4, Nob Hill Publishing, Madison.
- Ray, W. (1981), *Advanced process control*, ISBN 0-07-051250-7, McGraw-Hill Book Company, New York.
- Reichl, G. (2003), 'WasteWater a library for modelling and simulation of wastewater treatment plants in modelica', *Proceedings of the 3rd International Modelica Conference* pp. 171–176.
- Robertson, D. & Lee, J. (1995), 'A least squares formulation for state estimation', *Journal of Process Control* **5**(4), 291–299.
- Rossiter, J. (2003), *Model-based predictive control: a practical approach*, ISBN 0-8493-1291-4, CRC Press, Florida.
- Rugh, W. (1996), *Linear system theory*, ISBN 0-13-441205-2, 2nd edn, Prentice-Hall, New Jersey.
- Russell, S., Robertson, D., Lee, J. & Ogunnaike, B. (2000), 'Model-based quality monitoring of batch and semi-batch processes', *Journal of Process Control* **10**, 317–332.
- Samuelsson, P. (2005), Control of nitrogen removal in activated sludge processes, PhD thesis, Uppsala University, Uppsala, Sweden.
- Schmitz, G., ed. (2005), *MOSILAB: development of a Modelica based generic simulation tool supporting model structural dynamics*, the 4th International Modelica Conference, the 4th International Modelica Conference, Hamburg.
- Seborg, D., Edgar, T. & Mellichamp, D. (2004), *Process dynamics and control*, ISBN 0-471-45246-7, second edn, Wiley, New Jersey.
- Simon, D. (2006), *Optimal state estimation*, ISBN-13 978-0-471-70858-2, John Wiley & Sons, Hoboken, New Jersey.
- Simon, D. & Simon, D. (2003), Kalman filtering with inequality constraints for turbofan engine health estimation, Technical Memorandum NASA/TM-2003-212111, National Aeronautics and Space Administration (NASA), Washington, DC 20546-0001.
- Söderström, T. & Stoica, P. (1989), *System identification*, ISBN 0-13-881236-5, Prentice Hall International, New York.
- Sontag, E. (1998), *Mathematical control theory: deterministic finite dimensional*, ISBN 0-387-98489-5, 2nd edn, Springer, New York.

- Soroush, M. (1998), 'State and parameter estimations and their applications in process control', *Computers & Chemical Engineering* **23**, 229–245.
- Stewart, G. (1987), 'Collinearity and least squares regression', *Statistical Science* **2**(1), 68–84.
- Sunan, H., Kiong, T. & Heng, L. (2002), *Applied predictive control*, ISBN 1-85233-338-3, Springer, London.
- Takács, I., Patry, G. & Nolasco, D. (1991), 'A dynamic model of the clarification — thickening process', *Wat. Res.* **25**(10), 1263–1271.
- Theilliol, D., Ponsart, J., Harmand, J., Join, C. & Gras, P. (2003), 'On-line estimation of unmeasured inputs for anaerobic wastewater treatment processes', *Control Engineering Practice* **11**, 1007–1019.
- Tiller, M. (2001), *Introduction to physical modeling with Modelica*, ISBN 0-7923-7367-7, Kluwer Academic Publishers, Massachusetts.
- Tummescheit, H. (2002), Design and Implementation of Object-Oriented Model Libraries Using Modelica, PhD thesis, Lund Institute of Technology, Lund, Sweden.
- van der Merwe, R. & Wan, E. (2001), 'The square-root unscented Kalman filter for state and parameter estimation'. <http://speech.bme.ogi.edu/publications/ps/merwe01a.pdf>.
- VanDyke, M., Schwartz, J. & Hall, C. (2004), 'Unscented Kalman filter filtering for spacecraft attitude state and parameter estimation'. Available at <http://www.aoe.vt.edu/~cdhall/papers/AAS04-115.pdf>, accessed 31 May 2007.
- Vanrolleghem, P., van Daele, M. & Dochain, D. (1995), 'Practical identifiability of a biokinetic model of activated sludge respiration', *Wat. Res.* **29**, 11.
- Videla, J. (2007), Dymola-Dymosim: functionalities and interfaces with external environments. Technical report (unpublished), Telemark University College, Norway.
- Walter, E. & Pronzato, L. (1997), *Identification of Parametric Models from Experimental Data*, ISBN 3-540-76119-5, Springer, Berlin.
- Wan, E. & van der Merwe, R. (2000), The unscented Kalman filter for nonlinear estimation, in 'Proc. of IEEE Symposium 2000 (AS-SPCC)', Lake Louise, Alberta, Canada.
- Weijers, S. (2000), Modelling, identification and control of activated sludge plants for nitrogen removal, PhD thesis, Technische Universiteit Eindhoven.
- Weijers, S. & Vanrolleghem, P. (1997), 'A procedure for selecting best identifiable parameters in calibrating activated sludge model No. 1 to full-scale plant data', *Wat. Sci. Tech.* **36**(5), 69–79.
- Zhao, H., Isaacs, S., Sørensen, H. & Kümmel, M. (1994a), 'A Novel Control Strategy for Improved Nitrogen Removal in an Alternating Activated Sludge Process — Part II. Control Development', *Wat. Res.* **28**, 535–542.

- Zhao, H., Isaacs, S., Sørensen, H. & Kümmel, M. (1994b), 'Nonlinear optimal control of an alternating activated sludge process in a pilot plant', *Journal of Process Control* **4**(1), 33–43.

Part II

Published and Submitted Papers

Paper A: Systematic Analysis of Parameter Identifiability for Improved Fitting of a Biological Wastewater Model to Experimental Data

This paper was first presented at the conference and published in the Proceedings of the Scandinavian Conference on Simulation and Modeling 2005 (SIMS 2005) held in Trondheim, Norway, 13 – 14 October 2005. Extension of the paper was published in *Modeling, Identification and Control* Volume 27, Number 4, pp. 219 – 238.

The topic of this paper was studied by Ferrero (2005) in the master thesis during her stay at Telemark University College in spring 2005. The author of this PhD thesis actively participated in the entire process of this study, and further developed applications of the proposed method in her PhD research activities (see Paper B).

Systematic Analysis of Parameter Identifiability for Improved Fitting of a Biological Wastewater Model to Experimental Data¶

CRISTINA SARMIENTO FERRERO†, QIAN CHAI†,
MARTA DUEÑAS DÍEZ‡, SVERRE H. AMRANI§ and BERNT LIE*†

Keywords: *ASM2d, identifiability analysis, parameter estimation, model fitting*

In this paper, a general description of wastewater treatment based on activated sludge is given, with emphasis on the ASM2d model. Particular emphasis has been given to make the presentation readable without a too detailed prior knowledge of wastewater treatment. Next, a method for experimental parameter identifiability analysis is described. After a presentation of the wastewater treatment plant at Duvbacken in Gävle, Sweden, the experimental identifiability of the dynamic model is analyzed. Out of the 45 original parameters in the model, the analysis indicates that with the given experimental conditions, 12 parameters can be identified.

1. Introduction

Water is one of our most precious resources, and wastewater treatment is becoming more and more important in a highly populated, industrialized world. Wastewater contains organic matter and microorganisms, and the microorganisms thus deplete the water for oxygen. Furthermore, nitrogen and phosphorous content leads to algae growth. It is thus of importance to reduce the content of organic matter, as well as nitrogen and phosphorous. There are also safety regulations on the content of microorganisms, as well as on other content in the wastewater. See e.g. Metcalf & Eddy (2003) for an overview of characteristics of wastewater, and treatment of wastewater.

Both for design and operation of wastewater treatment plants, it is of interest to develop models of how the plants transform the feed. One characteristic of wastewater feed is its highly varying flow rate and composition (Olsson & Newell 1999). This implies that dynamic models are highly relevant for wastewater treatment plants. One such series of models are the Activated Sludge Models (ASM) (Henze, Harremoës, la Cour Jansen & Arvin 1996), which have been developed to include a description of organic content and decomposition, microorganisms, nitrogen processes, and phosphorous processes. In particular, ASM2d attempts to describe the important processes involved in the transformation of organic matter, nitrogen, and phosphorous (Henze, Gujer, van Loosdrecht & Mino 2000).

The ASM models are parametric, and nominal parameter values are suggested in publications. However, the models are based on averages of microorganism and organic matter populations, and wastewater characteristics vary from plant to plant. It is thus necessary to fit the model parameters to experimental data from a specific wastewater

*Corresponding author: Bernt.Lie@hit.no.

†Telemark University College, P.O. Box 203, N-3901 Porsgrunn, Norway.

‡Hydro Corporate Research Center, Porsgrunn, Norway.

§BioTek AS, Porsgrunn, Norway.

¶An early version of this paper was presented at SIMS 2005, the 46th Conference on Simulation and Modeling, Trondheim, Norway, October 13–14, 2005.

stream in order to ensure good predictive properties of a model (Jeppson 1993), (Brun, Kühni, Siegrist, Gujer & Reichert 2002).

In this paper, we consider the Duvbacken wastewater treatment plant in Gävle, Sweden, which is designed to treat the municipal wastewater from 100.000 pe¹. This plant has recently been modified to remove phosphorous, nitrogen, and organic matter, primarily using bacteria. In situations where the biology represented by the bacteria struggle to perform according to governmental effluent criteria, chemicals are added as a supplement. The particular emphasis is to study the identifiability of parameters in a dynamic model of the plant, based on available experimental data. In this paper, real experimental data were unavailable, and the study is therefore based on simulated data. However, the procedure can be used without modification with real data.

In section 2 of the paper, an overview of principles for modeling biological wastewater treatment (BWWT) is given. In section 3, basic principles of parameter identifiability are discussed. In section 4, a plant at Duvbacken, Gävle, is presented, and a model of the plant is verified. In section 5, identifiability analysis for the Duvbacken model is carried out, and some results of model fitting are given. Finally, some conclusions are drawn.

2. Biological Wastewater Treatment

2.1. Wastewater characterization and treatment

Due to the importance of clean water, wastewater treatment is a large and growing industry throughout the world. Microorganisms in water tend to use dissolved oxygen (DO) while consuming organic and inorganic compounds, resulting in harmfully low concentrations of DO to the aquatic environment. Thus, a primary task in wastewater treatment is to remove the oxygen consuming pollutants and their feed. Other pollutants consisting of nitrogen and phosphorous lead to eutrophication, which is the accelerated ageing of lakes and estuaries due to excessive plant and algal growth. It is thus in addition important to remove nitrogen and phosphorous compounds from wastewater.

The composition of wastewater depends on its source; important classes of wastewaters are industrial and domestic wastes. The latter, sewage, consists of substances such as ground garbage, laundry water, and excrement. More than 99% of wastewater is water, and about 0.5 g/l is suspended solids such as cellulose and organic matter in the form of fatty acids, carbohydrates, and proteins. The bad odor of sewage stems from protein decomposition under *anaerobic*² conditions. Sewage contains a varied population of microorganisms such as bacteria, which must not be mixed into drinking water. Still, microorganisms are useful in transforming dissolved substances into particulate substances which can then be removed from the wastewater.

As mentioned, one problem with wastewater is that it contains carbonaceous matter which through chemical reactions depletes water for oxygen. The simplest measure of the content of such matter, is the *biochemical oxygen demand* (BOD): the BOD is equal to the amount of dissolved oxygen that is consumed by the wastewater when incubated for a specified length of time at 20°C, typically for 5 or 7 days (BOD₅ or BOD₇, respectively); BOD measures the oxidation of organic matter. A better measure is the

¹ pe = person equivalents.

² Anaerobic, i.e., without the presence of oxygen or oxides.

Table 1. Typical composition of domestic waste water, with indication of effluent standards. From Bailey & Ollis (1986).

Parameter	Influent raw wastewater	Effluent in an acceptable plant
BOD, mg/l	100 – 250	5 – 15
COD, mg/l	200 – 700	15 – 75
Suspended solids, mg/l	100 – 400	10 – 25
Nitrogen, mg/l	20 – 30	2 – 5
Total phosphorous, mg/l	6 – 10	0.2 – 0.6

chemical oxygen demand (COD), both because it can be measured routinely in a couple of hours³ (Bailey & Ollis 1986), but also because it gives a more accurate measure of the carbonaceous content: COD measures the oxidation of both organic and inorganic matter.

Other undesirable components in wastewater are nitrogen, typically in the form of ammonia (NH₃), and phosphorous. In addition, wastewater holds both larger objects and readily settleable solids, as well as oil films and other components such as heavy metals and toxic organics. Table 1 illustrates typical feed compositions and effluent standards for sewage.

A wastewater treatment plant typically consists of two or three stages. In the *primary stage*, easily separable particles (boulders—floc particles, etc.) are removed. In the *secondary stage*, suspended particles and soluble materials are removed, while in the *tertiary stage*, remaining contaminants (subcolloidal and soluble materials) are removed, the water is treated chemically to remove bacteria, etc.

In this paper, we will concentrate on the secondary stage, where the wastewater has a muddy consistence and where oxygen is added. Under such oxygenated conditions, microorganisms thrive, forming an active suspension of biological solids (mostly bacteria) called an *activated sludge*.

2.2. Wastewater species

Essentially, microorganisms metabolize and breed while consuming *substrate*⁴. In this process, the microorganisms synthesize and secrete a *polysaccharide gel* causing the microbes to agglomerate into flocs (Bailey & Ollis 1986). These flocs have a high affinity for suspended matter (substrate, etc.), and flocculation thus enables the removal of microorganisms and substrate through sedimentation.

Wastewater contains a large number of microorganisms and substrate, and it is necessary to simplify the description of the components. In the sequel, the components that are detailed are those used in the Activated Sludge Model 2d (ASM2d) (Henze, Gujer, Mino, Matsuo, Wentzel, Marais & van Loosdrecht 1999), (Henze *et al.* 2000). The structure of the presentation is similar to that of Jeppson (1993), although his is for ASM1. A minor simplification of the presentation here compared to the full ASM2d, is that we neglect the possibility of phosphorous to bind to metal-hydroxides, and the associated chemical processes. In the presentation, we will follow the naming conventions of the Activated Sludge Models (ASM), which distinguish between dissolved

³ Or in a few minutes using sophisticated instruments.

⁴ By substrate is meant nutrients (organic and inorganic matter, including oxygen) and energy.

component S and particulate component X . Subscripts are attached to distinguish between the various dissolved and particulate components. As wastewater contains a wide variety of e.g. organic matter, a precise model would need to use population balances. In the ASM, average species are considered in order to simplify the models. In the description, S or X (with subscript) refer both to a specie, and to the concentration of the specie.

Organic substrate is considered to be either *biodegradable*, or *unbiodegradable*. The unbiodegradable matter can be either soluble (small species) or particulate (larger species); in both cases, it is considered *inert* matter, and named S_I and X_I , respectively. The biodegradable matter is substrate for the organisms, and can be either soluble (small species/molecules) or particulate (larger species). Particulate biodegradable matter X_S is too large to pass through the cell walls of the microorganisms, and must first be decomposed—which is time consuming, thus X_S is denoted slowly biodegradable substrate. Soluble biodegradable matter is readily biodegradable. ASM2d distinguishes between fermentable biodegradable substrate S_F , and the biodegradable fermentation product S_A (assumed to be acetate CH_3CO_2^- in later analysis). In ASM1, there was no distinction between S_F and S_A , and $S_S = S_F + S_A$ was used there.

Inorganic substrate consists of small molecules, which are soluble. In addition to oxygen S_{O_2} (O_2), these are nitrous components such as nitrogen S_{N_2} (N_2), ammonium S_{NH_4} (some ammonia NH_3 , but mainly ammonium ion NH_4^+ ; assumed to be 100% ammonium ion in stoichiometric computations), nitrogen oxides S_{NO_3} (nitrate NO_3^- and nitrite NO_2^- ; for stoichiometric computations assumed to be 100% nitrate), and inorganic soluble phosphorous S_{PO_4} (primarily ortho-phosphates, which are anions of the inorganic chemical compound of phosphoric acid H_3PO_4 ; for stoichiometric computations assumed to consist of 50% H_2PO_4^- and 50% HPO_4^{2-}).

To facilitate the computation of pH and charge balance, the alkalinity of wastewater is described by S_{ALK} (for stoichiometric computations assumed to be bicarbonate HCO_3^-).

Several types of microorganisms exist in wastewater; all of these are considered to be particulate. *Heterotrophic* organisms X_H require organic compounds as their substrate; they are responsible for *hydrolysis*⁵ of slowly biodegradable substrate X_S into smaller components, and grow *aerobically*⁶ and *anoxically*⁷, as well as are active in *anaerobic*⁸ fermentation. *Autotrophic* organisms X_{AUT} require inorganic compounds as their substrate; they are responsible for *nitrification* under aerobic conditions, where ammonium S_{NH_4} is oxidized directly into nitrogen oxides S_{NO_3} . Finally, there are phosphate-accumulating organisms (PAO), X_{PAO} ; these may grow in both anoxic and aerobic environments.

In addition to the phosphorous accumulating organisms X_{PAO} mentioned above, two types of cell internal storage products associated with PAO are considered: X_{PHA} includes primarily organic poly-hydroxyalkanoates (assumed to be poly-hydroxy-butyrate $(\text{C}_4\text{H}_6\text{O}_2)_n$ in stoichiometric computations), and inorganic poly-phosphates X_{PP} (assumed to be

⁵ Chemical decomposition of matter by the splitting of a bond and the addition of the hydrogen cation H^+ and the hydroxide anion OH^- of water.

⁶ Aerobic = requires the *presence* of oxygen; $S_{\text{O}_2} > 0$.

⁷ Anoxic = extracts oxygen from nitrate oxides; $S_{\text{O}_2} \approx 0$, $S_{\text{NO}_3} > 0$. In this process, nitrate NO_3 is converted to N_2 , hence this process is also denoted *denitrification*.

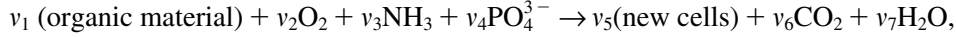
⁸ Anaerobic = requires the *absence* of oxygen and nitrate oxides; $S_{\text{O}_2} \approx 0$, $S_{\text{NO}_3} \approx 0$.

($K_{0.33} Mg_{0.33} PO_3)_n$ in stoichiometric computations). The phosphorous contents of X_{PHA} and X_{PP} are not included in X_{PAO} .

Finally, since phosphorous removal and precipitation introduces mineral fractions in the wastewater which are not included in the other components, a total suspended solids X_{TSS} is introduced.

2.3. Wastewater processes

The overall chemical process taking place in BWWT can be described as (Metcalf & Eddy 2003):

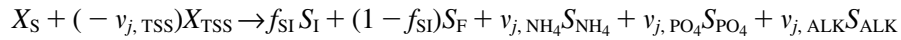


where the v_j -s are stoichiometric coefficients and it is assumed that microorganism enzymes have a catalytic effect on the reaction, but also that microorganisms grow and breed (new cells). Essentially, organic material is transformed into new cells under the consumption of oxygen, ammonia, and phosphorous components, producing water, CO_2 , and new cells which flocculate and may sediment.

In a realistic mathematical model, a more detailed set of stoichiometric reactions must be incorporated. In the following description of the chemical processes taking place between substrate and microorganisms, coefficients v_j , i_j , f_j , and Y_j constitute a parameterization of the stoichiometric coefficients (Henze *et al.* 1999); in all chemical reactions below, the parameterized stoichiometric coefficients are positive. Four groups of chemical processes are introduced: Hydrolysis processes (decomposition of slowly biodegradable organic substrate X_S), Processes of facultative heterotrophic organisms⁹ (metabolism and breeding of X_H , etc.), Processes of phosphorous-accumulating organisms (involving X_{PAO} and their cell-internal storages X_{PHA} and X_{PP}), and Nitrification processes (involving S_{NH_4} , S_{NO_3} , and S_{N_2}).

A final comment before going into the some details of the chemical processes in wastewater treatment: there are some inconsistencies in Henze *et al.* (1999). As an example, in their Table 3, the stoichiometric coefficients for ‘‘Aerobic growth on S_F ’’ are given, involving a number of species. However, in their summary Table 11, more species appear to be involved in the ‘‘reaction’’ than in their Table 3. The same is true for other processes. In the discussion below, the species that are discussed in their section ‘‘Biological processes, stoichiometry and kinetics’’ are underlined in the reaction schemes (the main species), while the additional species (as of their Table 11) are kept without underlining, and are given an arbitrary stoichiometric coefficient v_j .

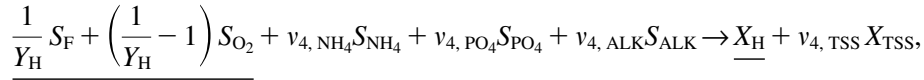
Hydrolysis: The following hydrolysis (decomposition) reactions take place on slowly biodegradable substrate X_S , catalyzed by enzymes in three classes of microorganisms, and split into soluble unbiodegradable S_I and fermentable S_F matter with a splitting factor f_{SI} :



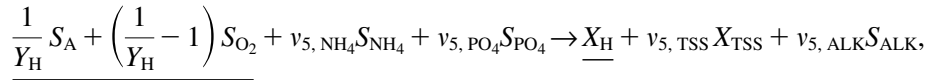
where, $j = 1$ is *aerobic* hydrolysis ($S_{O_2} > 0$), $j = 2$ is *anoxic* hydrolysis ($S_{O_2} \approx 0$, $S_{NO_3} > 0$), and $j = 3$ is *anaerobic* hydrolysis ($S_{O_2} \approx 0$, $S_{NO_3} \approx 0$).

⁹ Facultative = exhibiting a certain behavior under some environmental conditions, but not under other conditions.

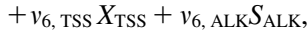
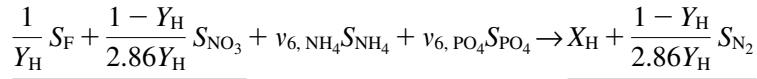
Facultative heterotrophic organisms: Facultative heterotrophic organisms metabolize and breed on substrates. Six chemical processes are considered. In oxygen based *aerobic growth* of X_H on fermentable substrate S_F :



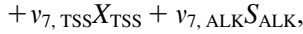
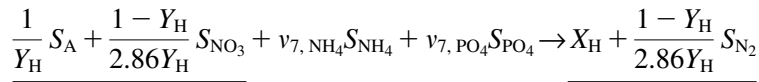
in oxygen based *aerobic growth* of X_H on fermentation products S_A :



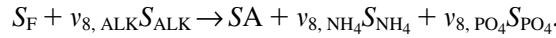
in nitrate based *anoxic growth* of X_H on fermentable substrate S_F :



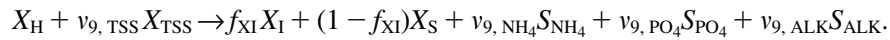
while in nitrate based *anoxic growth* of X_H on fermentation products S_A :



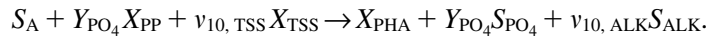
Under *anaerobic* conditions ($S_{O_2} \approx 0$, $S_{NO_3} \approx 0$), the heterotrophic microorganisms do not grow. Instead X_H catalyze *fermentation* of S_F leading to fermentation products S_A :



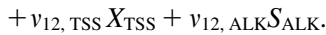
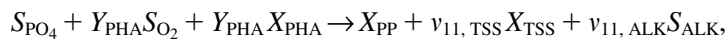
Finally, *lysis* reactions (decomposition/decay) occur in the heterotrophic organisms, where essentially X_H is split into unbiodegradable matter X_I and slowly biodegradable substrate X_S with a certain splitting factor f_{XI} :



Phosphorous-accumulating organisms: Organisms X_{PAO} have the potential to accumulate phosphorous in the form of poly-phosphate X_{PP} . Eight chemical processes are considered. Poly-hydroxy-alkanoates X_{PHA} are created from fermentation products S_A and stored¹⁰:

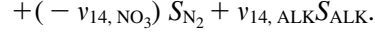
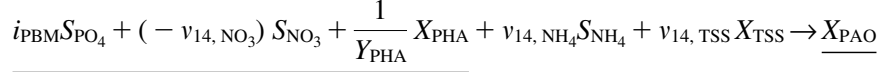
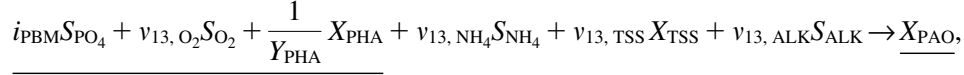


Poly-phosphates X_{PP} grow *aerobically* and *anoxically*:

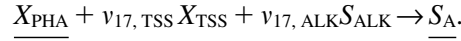
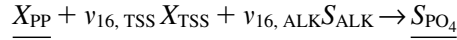
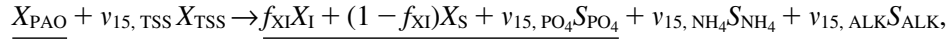


¹⁰ In Table 4 of Henze *et al.* (1999), X_{PHA} is assigned the stoichiometric coefficient 0 by misprint.

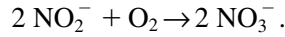
Likewise, the organisms X_{PAO} grow *aerobically* and *anoxically*:



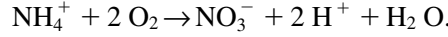
Finally, each of the phosphorous particles X_{PAO} , X_{PP} , and X_{PHA} are prone to lysis (decomposition):



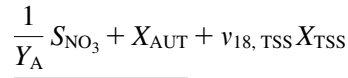
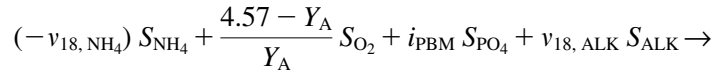
Nitrifying organisms: In anoxic processes, *autotrophic* organisms utilize oxygen from nitrate ions (NO_3^- , denitrification), producing nitrogen N_2 . This nitrate is produced in the nitrification process, which is a two step process, typically:



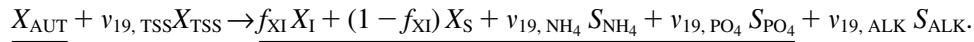
Since the reaction of nitrite (NO_2^-) into nitrate is very fast, nitrite will hardly be present, and it is convenient to operate with a simplified overall stoichiometric reaction:



Utilizing the ASM notation, we write this nitrification process as:



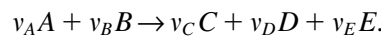
Also, the autotrophic organisms are prone to *lysis*:



In Henze *et al.* (1999), two processes for chemical precipitation of phosphates are included; we have neglected these processes in our model.

2.4. Wastewater kinetics and dynamic models

A stoichiometric equation for a chemical process is typically written as



If A , B , C , D , and E are unrelated, we can not say much more about the stoichiometric coefficients v_j . However, if $A = \text{NH}_4^+$, $B = \text{O}_2$, $C = \text{NO}_3^-$, $D = \text{H}^+$, and $E = \text{H}_2\text{O}$,¹¹ conservation of atoms and charge dictates that $v_A = k$, $v_B = 2k$, $v_C = k$, $v_D = 2k$, and $v_E = k$, where k conveniently can be chosen as 1 (Smith & Missen 1991). We can use conservation of atoms and charge to relate the stoichiometric coefficients of the ASM2d, hence parametric description of the coefficients $v_i(Y_j, f_k, i_l)$ can be found; see Henze *et al.* (1999) for typical parameters.

With known stoichiometric equations and coefficients, kinetic rate expressions $\rho_j \left[\frac{\text{matter}}{\text{volume} \cdot \text{time}} \right]$ are needed, one for each chemical process. For all rates ρ_j , the rate is proportional to the concentration of the involved microorganism:

1. Hydrolysis and heterotroph growth depend on the presence of heterotroph organisms X_H , hence $\forall j \in \{1, \dots, 9\} : \rho_j \propto X_H$.
2. The processes depending on phosphorous-accumulating organisms X_{PAO} have rates $\forall j \in \{10, \dots, 17\} : \rho_j \propto X_{\text{PAO}}$.
3. Processes governed by nitrifying organisms (autotroph organisms) have rates $\forall j \in \{18, 19\} : \rho_j \propto X_{\text{AUT}}$.

Next, the hydrolysis process rates depend on the simultaneous presence of substrate X_S and microorganism X_H ; these rates depend on how well X_S covers the surface of X_H , hence

$$\forall j \in \{1, 2, 3\} : \rho_j \propto \frac{X_S/X_H}{K_X + X_S/X_H} \rightarrow \begin{cases} 0, & \frac{X_S}{X_H} \rightarrow 0 \\ 1, & \frac{X_S}{X_H} \rightarrow \infty \end{cases}.$$

Similar rate expressions are used for other processes where microorganisms catalyze transformation of material, e.g. for storage of X_{PHA} (process 10) and aerobic storage of X_{PHA} and X_{PP} (processes 11 and 13).

Finally, some switching functions are introduced in order to model how the rates depend on the presence or absence of substrates. As an example, aerobic processes are assumed to depend on the *presence* of oxygen, hence the rates are made proportional to

$$\forall j \in \{1, 4, 5, 11, 13, 18\} : \rho_j \propto \frac{S_{\text{O}_2}}{K_{\text{O}_2} + S_{\text{O}_2}} \rightarrow \begin{cases} 0, & S_{\text{O}_2} \rightarrow 0 \\ 1, & S_{\text{O}_2} \rightarrow \infty \end{cases}.$$

On the other hand, for anoxic and anaerobic processes that depend on the *absence* of oxygen, the rates are made proportional to either

$$\forall j \in \{2, 3, 6, 7, 8\} : \rho_j \propto \frac{K_{\text{O}_2}}{K_{\text{O}_2} + S_{\text{O}_2}} \rightarrow \begin{cases} 1, & S_{\text{O}_2} \rightarrow 0 \\ 0, & S_{\text{O}_2} \rightarrow \infty \end{cases}$$

or

$$\forall j \in \{12, 14\} : \rho_j \propto \frac{K_{\text{O}_2}}{S_{\text{O}_2}} \rightarrow \begin{cases} \infty, & S_{\text{O}_2} \rightarrow 0 \\ 0, & S_{\text{O}_2} \rightarrow \infty \end{cases}.$$

Similar switching functions are used to model the required presence or absence of nitrate S_{NO_3} .

¹¹ The simplified nitrification process.

The full set of process rate equations ρ_j are given in Table 7 of Henze *et al.* (1999). The final step in the modeling phase is to formulate specie balances, which typically have the form

$$\frac{d}{dt} m_j = \dot{m}_{i,j} - \dot{m}_{e,j} + \dot{m}_{g,j}$$

where m_j is the accumulated mass of specie j (e.g. S_F , X_H , etc.), $\dot{m}_{i,j}$ is the mass influx of specie j into a system, $\dot{m}_{e,j}$ is the mass efflux of specie j out of a system, and $\dot{m}_{g,j}$ is the mass of specie j generated in a chemical reaction per time unit. For a well mixed system of volume V ,

$$\dot{m}_{g,j} = V \cdot \sum_{k=1}^K v_{k,j} \rho_k,$$

where K chemical processes are taking place; $K = 19$ in our slightly simplified ASM2d description. In our description, we use either 18 or 17 species, depending on whether we include $S_S = S_F + S_A$ or not. Thus, we will have 18 (or 17) differential equation for mass accumulation for each well mixed system volume in our biological wastewater treatment plant.

3. Parameter Identifiability

3.1. System, model, and parameter

For a system \mathcal{S} , assume that we have developed a model structure \mathcal{M} with parameter $\theta \in \mathbb{R}^{n_\theta}$; the corresponding model is denoted $\mathcal{M}(\theta)$. The (parametric) model may be based on a mechanistic understanding of the system \mathcal{S} , or be a generic mathematical mapping. By varying θ over a feasible set Θ , we get a set of models $\mathcal{M}(\Theta)$, i.e. a model structure \mathcal{M} . In reality, the model behavior depends on the experimental conditions \mathcal{K} ; \mathcal{K} is a description of how an experiment has been carried out, including a sequence of inputs $u_t \in \mathbb{R}^{n_u}$, $t \in \{1, \dots, T\}$ to, and outputs $y_t \in \mathbb{R}^{n_y}$, $t \in \{1, \dots, T\}$ from the real system \mathcal{S} . With a given model $\mathcal{M}(\theta)$ and experimental conditions \mathcal{K} , we can compute a model output $y_t^m(\mathcal{M}(\theta), \mathcal{K})$. We will simplify the notation for the model output to $y_t^m(\theta)$.

With available model structure $\mathcal{M}(\Theta)$ and experimental conditions \mathcal{K} , we can attempt to find a specific parameter $\hat{\theta}$ which is such that the model output $y_t^m(\hat{\theta})$ is close to the system output y_t in some sense. To achieve this, we can choose from a set of identification methods \mathcal{I} ; the estimate $\hat{\theta}$ will also depend on the chosen identification method \mathcal{I} . We will simplify the notation and write $y_t^m(\hat{\theta})$ as \hat{y}_t .

Usually, it is assumed that the real system \mathcal{S} is an element of the model structure $\mathcal{M}(\Theta)$ and has a true parameter θ^* such that \mathcal{S} is (input-output) identical to $\mathcal{M}(\theta^*)$. It is thus of interest to study whether and to what degree it is possible to find the true system parameters θ^* from the given model structure $\mathcal{M}(\Theta)$, experimental conditions \mathcal{K} , and identification method \mathcal{I} . We will denote this study an *experimental* parameter identifiability study, since the result depends on the experimental conditions \mathcal{K} (and to some degree on \mathcal{I}). Another important study could be whether and to what degree θ^* can be found from $\mathcal{M}(\Theta)$ if we allow *any* (conceivable) experimental conditions \mathcal{K} ; this we will denote a *theoretical* parameter identifiability study.

Clearly, the set of experimentally identifiable parameters will be a subset of the set of theoretically identifiable parameters. Examples of methods for studying theoretical parameter identifiability, are given in e.g. Pohjanpalo (1978) and Holmberg (1982), and in Ljung & Glad (1994); see also Walter & Pronzato (1997). However, for many

industrial processes it will be difficult to get accept for carrying out active experiments, and it may be necessary to rely on the natural perturbation in the system. Thus, we may not be able to identify all the theoretically identifiable parameters in practice, and the experimental identifiability may give a better measure of the parameters which can be found.

After a set of identifiable parameters have been found, these can be estimated using some identification technique \mathcal{I} . Then, the statistical quality of the model parameters $\hat{\theta}$ and prediction model \hat{y}_t should be studied.

Parameter identifiability and parameter estimation in mechanistic models has been widely studied in engineering (Bard 1974), (Beck & Arnold 1977), (Walter & Pronzato 1997), (Rawlings & Ekerdt 2002), (Ljung 1999), (Söderström & Stoica 1989). The systematic approach described in Brun, Reichert & Künsch (2001) seems well suited to Biological Waste Water Models, since it handles large simulation models, and provides identifiability diagnosis for parameter subsets. In the sequel, some background to the method of Brun *et al.* (2001) is given.

3.2. Practical identifiability analysis

We assume that the output $y_t \in \mathbb{R}^{n_y \times 1}$ and the model output $y_t^m(\theta | \mathcal{K})$ are related by $y_t = y_t^m(\theta | \mathcal{K}) + e_t$, where e_t describes model error/uncertainty. Furthermore, both y_t and u_t are available for $t \in \{1, \dots, T\}$ (experimental conditions \mathcal{K}) and we introduce the notation $y = y^m(\theta | \mathcal{K}) + e$ to describe all the resulting equations; $y, y^m, e \in \mathbb{R}^{n_y \cdot T \times 1}$. With a perfect model structure and the correct parameter vector θ^* , the error e will be $e \equiv 0$ and $y \equiv y^m(\theta^* | \mathcal{K})$. In the real world, e is unknown. A possible strategy is to set $e = 0$, and hope for a solution θ such that $y = y^m(\theta | \mathcal{K})$. We base the analysis on a linearized approximation of this equation, giving $T \cdot n_y$ linear equations in n_θ unknowns:

$$S \cdot \Delta \hat{\theta} = \Delta y, \quad (1)$$

where $\Delta y \triangleq y - y^m(\theta^i | \mathcal{K})$, $\Delta \hat{\theta} \triangleq \hat{\theta} - \theta^i$, $S \triangleq \left. \frac{\partial y^m}{\partial \theta} \right|_{\theta = \theta^i}$. S is denoted the model *sensitivity*. If no solution exists, it is common to project Δy on the column space of S , $\Delta y_{\mathcal{R}(S)}$, and solve $S \cdot \Delta \hat{\theta} = \Delta y_{\mathcal{R}(S)}$ —this equation always has a solution, which coincides with the solution of the corresponding normal equation from using the least squares method:

$$S^T S \cdot \Delta \hat{\theta} = S^T \cdot \Delta y. \quad (2)$$

The chosen method for analyzing the identifiability of parameters, is based on Brun *et al.* (2001), see also Brun *et al.* (2002), Dueñas Díez, Fjeld, Andersen & Lie (2006). This method is rooted in the linear equation of either equation 1, or 2. Essentially, these equations have a solution if the sensitivity matrix S has full rank. Clearly, if the j th column s_j of matrix S is a zero vector, then matrix S exhibits rank loss and parameter θ_j can not be found. Similarly, if $\|s_j\|$ is “small”, we may expect problems in finding θ_j . However, even when all columns of S are significantly different from zero vectors, we may have problems in finding θ if some column is linearly dependent on the other columns. This linear dependence is termed *collinearity*. Two possible measures of collinearity are the condition number κ_j , and the smallest singular value $\sigma_{\min}(S)$.

Scaling of outputs and parameters is essential, and it is recommended to use dimensionless quantities $\tilde{y}_{t,i} = y_{t,i}/y_i^\circ$ and $\tilde{\theta}_i = \theta_i/\theta_i^\circ$. In the sequel, it is assumed that such a scaling has been introduced as part of the preprocessing of the data.

In their method, Brun *et al.* (2001) consider the *sensitivity measure* δ_j^{msqr} defined as

$$\delta_j^{\text{msqr}} \triangleq \frac{1}{\sqrt{n_y T}} \|s_j\|_2. \quad (3)$$

Their strategy is then to sort the various parameters according to the value of δ_j^{msqr} . If there is a marked drop in the value of δ_j^{msqr} for some j , then they propose to assume that those parameters with the smallest sensitivity measure can not be found; hence these parameters are removed from the set of tunable parameters. It should be noted, however, that if the sensitivity measure decreases without marked jumps, and if it is within a couple of decades from the largest value, it may be unjustified to remove a parameter. In typical applications, a set of 10–20 parameters are often retained in the parameter set *after* the sensitivity screening described above.

The chosen measure of collinearity index $\gamma(\theta)$ is

$$\gamma(\theta) \triangleq \frac{1}{\sigma_{\min}(\tilde{S})} = \frac{1}{\sqrt{\sigma_{\min}(\tilde{S}^T \tilde{S})}}, \quad (4)$$

where \tilde{S} is a slightly modified sensitivity matrix where the columns have been normalized to unit norm. It should be noted that this measure is not necessarily a good measure of collinearity, see Stewart (1987) and Belsley (1991).

Collinearity is a measure of how combinations of parameters interact, thus they propose to consider all possible subsets \mathcal{P} of parameters of θ , and compute the collinearity index for all these possible subsets. Let θ_p be one of these subsets, thus $\gamma(\theta_p)$ is to be computed for all possible $p \in \mathcal{P}$. It follows from basic combinatorics that the number of possible subsets with at least 2 parameters and at most n_θ parameters is $\#\gamma(\theta_p) = \sum_{i=2}^{n_\theta} \binom{n_\theta}{i}$; with $n_\theta = 10$, $\#\gamma(\theta_p) \approx 10^3$, and with $n_\theta = 20$, $\#\gamma(\theta_p) \approx 10^6$.

In their papers, Brun *et al.* propose that when $\gamma(\theta_p) > 10$, parameter set p should be considered collinear. Thus, all parameter sets p for which $\gamma(\theta_p) \leq 10$ should be considered for further study, and parameter estimates for these sets p should be computed. The final choice of parameters is then based on a statistical analysis of the parameters for which $\gamma(\theta_p)$ is approximately less than 10.

4. The Duvbacken Plant Model

4.1. Process flow description

Wastewater first undergoes pretreatment involving removal of coarse particles and primary sedimentation of particles. Referring to Figure 1, the pretreated wastewater together with sludge return enters the anaerobe reactor step consisting of three parallel lines, each line consisting of two separate volumes in series, i.e. the anaerobe reactor step consists of a total of 6 separate volumes—each of which is supplied with mixers. The main purpose of the anaerobic reactor step, is to make conditions suitable for biological phosphorous removal.

Aerobe reactor step 1 has the same physical configuration as has the anaerobe reactor step, that is, three parallel lines, each line consisting of two separate volumes. All six volumes are aerated, either by continuous aeration or by intermittent aeration. The air is supplied from three blowers, serving a common air line distributing air to each of the six volumes—which in principle can be controlled independently—where the air enters each volume through about 140 aeration membrane dishes. The outlets from the three parallel lines converge into a common effluent line.

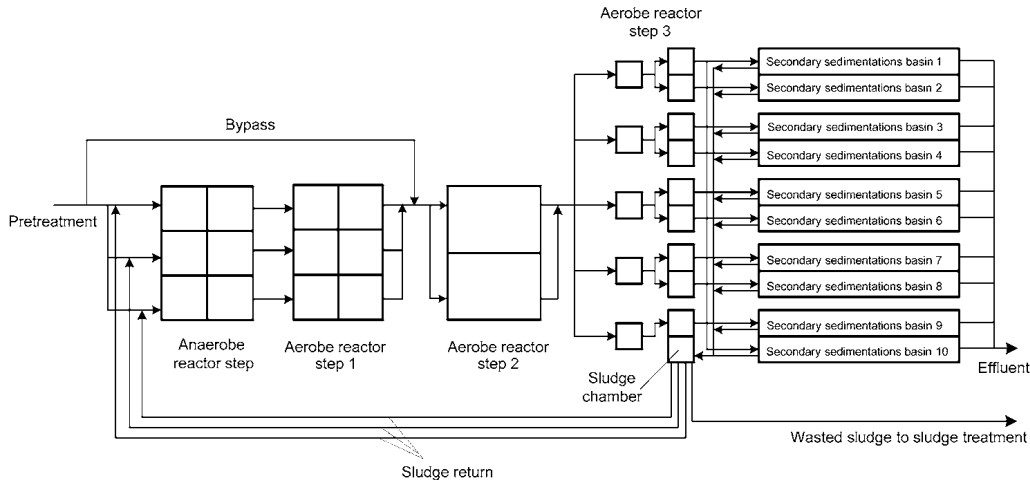


Figure 1. Flowsheet of the Duvbacken Waste Water Treatment Plant.

As can be seen from Figure 1, it is also possible to bypass a certain portion of the pretreated wastewater directly to aerobe reactor step 2. The reason for this is that effluent from the pre-treatment is rich in readily biodegradable organic matter. Hence, as to enhance denitrification in aerobe reactor step 2 (if biological nitrogen removal is the goal), bypass of the pretreated wastewater directly to reactor step 2 might be a useful strategy. Aerobe reactor step 2 consists of two rectangular and parallel basins. Each basin is aerated, either continuously or intermittently, and the air is supplied using four ejectors in each basin, distributed equally along the length of the basins.

From aerobe reactor step 2, the outlet lines converge into one effluent line which leads to aerobe reactor step 3. Aerobe reactor step 3 consists of fourteen small volumes, configured as shown in Figure 1. Each volume is aerated, either continuously or intermittently. The air is supplied from blowers, and the air is dispersed in each volume through aeration dishes.

From aerobe reactor step 3, the wastewater flows into ten parallel secondary sedimentation basins. The sludge from these basins, is pumped into a sludge chamber. The sludge chamber has two main purposes. Firstly, sludge is returned back to the inlet of the anaerobe reactor step—one sludge return line for each of the three parallel lines. Secondly, sludge is wasted and pumped to the sludge treatment which involves addition of polymer, thickening etc. The reject water from the thickening process is returned back to the plant.

4.2. Flow descriptions and sensors

The component concentrations in the influent to the plant is typically characterized as follows, Table 2. With reference to Table 2, the following comments should be made:

- The pretreatment reduces the total phosphorous content considerably due to sludge removal from the primary sedimentation basins.
- The pre-treatment reduces the unfiltered biological oxygen demand with 50. The average volumetric influent flow rate is 1400–1500 m³/h.

The plant has to obey the following governmental effluent requirements:

- $BOD_7 < 10 \text{ mg} = \text{l}$.
- $P_{\text{tot}} < 0.5 \text{ mg} = \text{l}$.

Table 2. Description of typical influent concentrations to the Duvbacken Waste Water Treatment Plant.

Parameters	Description	Units	Influent to plant		After pretreatment	
			Unfiltered	Filtered	Unfiltered	Filtered
P_{tot}	Total phosphorous	mg/l	6.0	3.2	3.4	1.2
$PO_4\text{-P}$	Phosphate phosphorous	mg/l		3.4		2.0
$NO_3\text{-N}$	Nitrate nitrogen	mg/l		1.0		1.4
$NH_4\text{-N}$	Ammonia nitrogen	mg/l		27.2		
BOD_7	Biological oxygen demand	mg/l	205	69	99	55

Table 3. On-line sensors at the Duvbacken Waste Water Treatment Plant.

Sensor type	Location	Sensors
Dissolved oxygen S_{O_2}	Aerobe reactor step 1, one sensor in each parallel line.	—
	Aerobe reactor step 2, two sensors in each basin.	y_6
	Aerobe reactor step 3, 5 sensors in 5 different volumes.	—
Nitrate nitrogen S_{NO_3}	Outlet of aerobe reactor step 2, one sensor	y_4
	Sludge chamber, one sensor.	y_8
Ammonia nitrogen S_{NH_4}	Outlet of aerobe reactor 1, one sensor.	y_1
	Effluent of the plant, one sensor.	y_9
Phosphate phosphorous	Effluent of the plant, one sensor.	—
Total phosphorous	Effluent of the plant, one sensor.	—
Dissolved phosphorous S_{PO_4}	Outlet of aerobe reactor step 3, one sensor.	y_{10}
Total solids $X_A + X_H + X_{PAO}$	Outlet of aerobe reactor step 1, one sensor.	y_3
	Outlet of aerobe reactor step 2, one sensor.	
	Sludge chamber, one sensor	y_7
	Effluent of plant, one sensor.	

As can be seen, there are currently no governmental requirements regarding the nitrogen contents in the effluent. Table 3 gives an overview of the available on-line sensors of the plant.

In addition, inorganic soluble phosphorous S_{PO_4} is measured by off-line laboratory analysis in aerobe reactors 1, 2, and 3; y_2 , y_5 , and y_{10} , respectively.

4.3. Model and model verification

A dynamic model is developed, based on the standard kinetics of the ASM2d model (Henze *et al.* 1999), (Metcalf & Eddy 2003), (Henze *et al.* 2000). In the dynamic model, each of the 4 steps in the plant (anaerobe, aerobe 1–3) is modeled as a perfectly stirred reactor. This leads to 17 states in each of 4 stirred volumes, i.e. 68 states. The model has a total of 45 parameters. The states of the model are grouped into the concentration of soluble species S_j and particulate species X_j . The model has 18 potential input variables: total volumetric flow rate q and the composition of the 17 species. In the model fitting,

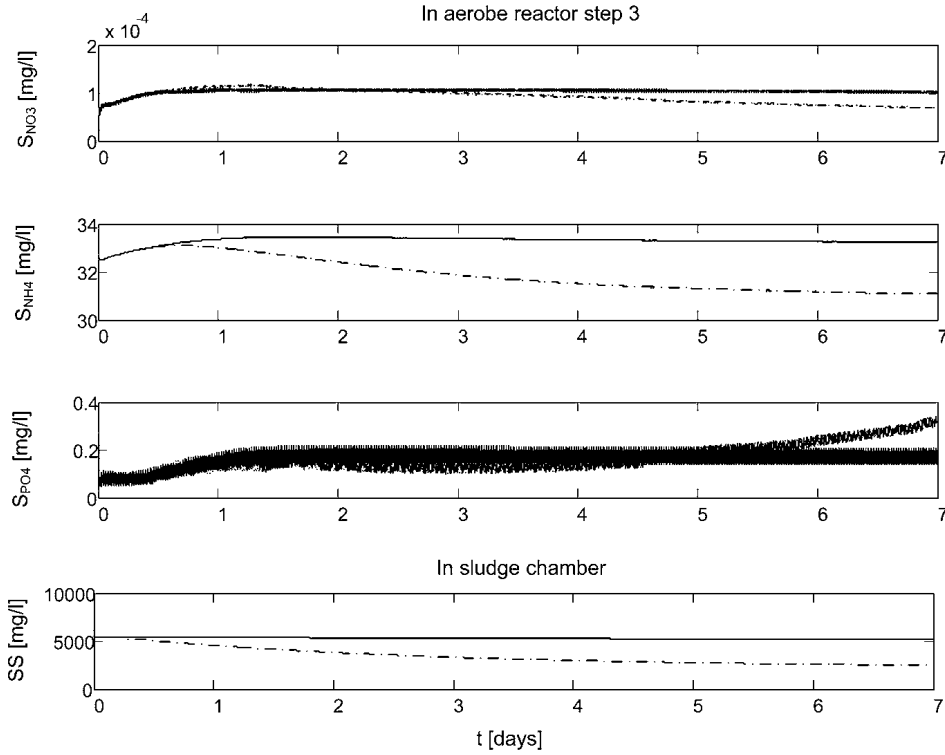


Figure 2. Simulated outputs S_{NO_3} , S_{NH_4} , S_{PO_4} , and $SS = X_A + X_H + X_{PAO}$ over 7 days, with sludge ages $T_{sludge} = 6$ d (solid, black) and $T_{sludge} = 2$ d (dashed, blue).

10 measurements are available¹². Examples of measurements are the concentration of suspended solids SS (i.e. the sum of the microorganisms), the concentration of soluble nitrates S_{NO_3} , the concentration of soluble ammonium S_{NH_4} , and the concentration of inorganic soluble phosphorous S_{PO_4} . The model has been implemented in Matlab, and the calling syntax for running the simulation code and compute the outputs for the plant is

$$[T, Y] = \text{bioout}(s0, x0, q, Sf, Xf, P0)$$

Here, $s0$ and $x0$ are the initial state vectors, q is the feed flow rate, Sf and Xf are the feed concentrations of the state variables, and $P0$ is the parameter vector. The response from the simulator is T which is a vector of time instants where the solution is available, and Y which is the matrix of output responses.

Before analyzing the identifiability and estimating parameters, it is a good idea to verify that the model gives reasonable responses (Olsson & Newell 1999). Figure 2 depicts the concentrations of SS, S_{NO_3} , S_{NH_4} , and S_{PO_4} for two different sludge ages in the plant— $T_{sludge} = 6$ d and $T_{sludge} = 2$ d,¹³ during 7 days of operation. The results are reasonable:

- Figure 2 shows that the suspended solids concentration SS is reduced when the sludge age is reduced. This makes sense: less residence time means that less solids is precipitated.

¹² In this paper, all measurements were assumed to be available every minute.

¹³ The sludge age is the residence time of the sludge in the reactor system.

- To observe bio-N operation (nitrogen removal), in general the sludge age should be no less than 6 d. To observe bio-P operation (phosphorous removal), the sludge age should be no less than 2 d. Thus, with a sludge age of 6 d in Figure 2, the system should be on the verge of dropping out of bio-N operation, and with a sludge age of 2 d, the system should be on the verge of dropping out of bio-P operation.
- Although bio-N operation may be obtained with a sludge age of 6 d, in the scenarios simulated in Figure 2, the feed contains no *autotrophic biomass* X_{AUT} , and the concentration of X_{AUT} in the effluent stays below 10^{-2} mg/l for the entire period of 7 d (not shown). To observe nitrification, X_{AUT} should have a value around, say, 300 mg/l. Nitrification would then manifest itself in e.g. a marked decrease in dissolved ammonium S_{NH_4} .
- Phosphorous removal is observed from e.g. the amount of dissolved phosphorous S_{PO_4} . In Figure 2, the concentration of S_{PO_4} starts to increase (after 4 d) in the case of the 2 d sludge age, while it stays low for the higher sludge age. This indicates bio-P operation at a sludge age of 6 d, while the bio-P operation is failing at a sludge age of 2 d.

There are some oscillations in some outputs in Figure 2—these are due to oscillating oxygen feed in the aerobic reactors.

5. Identifiability Analysis for the Duvbacken Plant Model

5.1. Experimental conditions

With our state space based model, we need a set of experimental input signals u_t in order to carry out the identifiability analysis. Identifiability analysis as discussed in this paper, does not require knowledge of the experimental response y_t from the real system. However, if we later want to estimate the parameters, then the system response is needed.

Here, we use a simulation model implemented in Matlab, with constant influent flow rate and compositions; the main transients in the system are due to initial values that are not at steady state, and an oscillating strategy for oxygen feed. Although the relatively small perturbation of the system is far from ideal for parameter identification, it will serve to illustrate the method for experimental parameter identifiability analysis.

5.2. Model sensitivity

The sensitivity was found by numerically perturbing the simulation model. First, a nominal output was computed, $y_i^m(\theta^0)$, and reshaped into $y^m(\theta^0) \in \mathbb{R}^{n_y T \times 1}$. Then the perturbed output $y_i^m(\theta^0 + e_j \cdot \delta\theta_j)$ was computed and reshaped into $y^m(\theta^0 + e_j \cdot \delta\theta_j)$; e_j is column j of the identity matrix $I \in \mathbb{R}^{n_\theta \times n_\theta}$. Then finally column s_j of the sensitivity S was computed as

$$s_j \approx \frac{y^m(\theta^0 + e_j \cdot \delta\theta_j) - y^m(\theta^0)}{\delta\theta_j}.$$

The outputs and parameters were scaled according to recommendations in Brun *et al.* (2002).

Table 4. Parameter importance ranking according to the sensitivity measures. Parameters labeled by an asterisk (*) are considered identifiable from the available data.

δ_j^{msqr}	Parameters	Parameter #	
14.812	δ_{PAO}^*	Maximum growth rate of PAO	30
3.9867	q_{PP}^*	Rate constant for storage of X_{PP}	29
3.2385	b_{PAO}	Rate for Lysis of X_{PAO}	32
3.105	K_X^*	Saturation coefficient for particulate COD	8
2.9954	K_h	Hydrolysis rate constant	4
2.9412	μ_{H}^*	Maximum growth rate on substrate	16
2.9324	q_{fe}^*	Maximum rate for fermentation	17
2.5858	$K_{\text{A,PAO}}^*$	Saturation coefficient for acetate, S_{A}	36
2.2768	b_{PHA}	Rate for Lysis of X_{PHA}	33
2.2196	q_{PHA}^*	Rate constant for storage of X_{PHA} (base X_{PP})	28
1.9906	$K_{\text{A,HET}}^*$	Saturation coefficient for growth on acetate S_{A}	23
1.9526	K_{PHA}^*	Saturation coefficient for PHA	44
1.326	$K_{\text{O}_2,\text{HYD}}$	Saturation/inhibition coefficient for oxygen	7
1.2238	b_{HET}^*	Rate for Lysis of X_{H}	19
1.1272	$K_{\text{O}_2,\text{HET}}^*$	Saturation/inhibition coefficient for oxygen	20
1.0583	η_{fe}^*	Anoxic hydrolysis reduction factor	6
0.99162	$K_{\text{O}_2,\text{PAO}}$	Saturation/inhibition coefficient for oxygen	34
0.95771	K_{F}	Saturation coefficient for growth on S_{F}	21
0.52842	$K_{\text{PS,PAO}}^*$	Saturation coefficient for phosphorus in storage of PP	38
0.46989	b_{PP}^*	Rate for Lysis of X_{PP}	45

5.3. Sensitivity index

The sensitivity measure $\delta_j^{\text{msqr}} = \|s_j\|_2 / \sqrt{n_y T}$ was computed. Out of the 45 original parameters, the 20 most sensitive parameters are shown in Table 4.

Although the decrease in δ_j^{msqr} is gradual, we thus make the deliberate choice of assuming that parameters 21–45 are not identifiable. We assign the nominal parameter value to these parameters, and then continue with collinearity analysis to see which of the 20 parameters with highest sensitivity that we can estimate.

5.4. Collinearity indices and identifiable parameters

Similarly as to in Brun *et al.* (2002), the 20 parameters of interest from the parameter sensitivity ranking can be classified in different groups according to the biological processes to which they belong. The parameters are divided into 4 different groups according to the kind of biological processes they describe: 4 parameters are related to hydrolysis of particulate substrate, 6 parameters are related to heterotrophic organisms, 10 parameters are related to phosphorus-accumulating organisms, and none are related to autotrophic organisms. The distribution of the 20 most sensitive parameters is thus:

- Hydrolysis of particulate substrate: K_X , K_h , $K_{\text{O}_2,\text{HYD}}$, η_{fe}
- Heterotrophic organisms: μ_{H} , q_{fe} , $K_{\text{A,HET}}$, b_{HET} , $K_{\text{O}_2,\text{HET}}$, K_{F}
- Phosphorus-accumulating organisms: μ_{PAO} , q_{PP} , b_{PAO} , $K_{\text{A,PAO}}$, b_{PHA} , q_{PHA} , K_{PHA} , $K_{\text{O}_2,\text{PAO}}$, $K_{\text{PS,PAO}}$, b_{PP}
- Autotrophic organisms:—

Next, the collinearity index $\gamma(\theta_p)$ is calculated for all possible subsets of the top 20 parameters. Values for $\gamma(\theta_p)$ lie in the range 1–11. This is almost within the range of

$\gamma(\theta_p) \leq 10$ as proposed by Brun *et al.* (2002), so our parameters are hardly collinear. However, to illustrate the concept of collinearity, we instead consider parameters to be collinear if $\gamma(\theta_p) > 5$. Here it is found that there are parameter subsets with up to 13 elements which fulfill $\gamma(\theta_p) \leq 5$, whereas subsets with 14 and more elements all have $\gamma(\theta_p) > 5$. Therefore, we consider a maximum of 13 parameters as potentially identifiable from the available data. The subset containing the largest number of parameters with the smallest collinearity index $\gamma(\theta_p) \leq 5$ is selected as the best identifiable; and has collinearity index $\gamma(\theta_p) = 4.27$.

Experience reported in the literature suggest that parameter b_{PAO} is difficult to estimate, even though the model is sensitive to this parameter (Brun *et al.* 2002): typically, an unrealistic value of b_{PAO} is found in the parameter estimation. We thus choose to take b_{PAO} out of the set of parameters to estimate. With this simplification, $\gamma(\theta_p)$ drops to 4.20. The chosen final subset of 12 parameters consists of 2 parameters related to hydrolysis of particulate substrate, 3 parameters related to heterotrophic organisms, and 7 related to phosphorus accumulating organisms; these parameters are indicated with an asterisk in Table 4.

5.5. Parameter estimation

After finding an identifiable parameter subset based on knowledge of $y_t^m(\theta^0 | \mathcal{X})$, we are ready to estimate the parameters. To do so, we need responses y_t , $t \in \{1, \dots, T\}$ from the plant. As an initial study before doing parameter estimation based on real data from the Gävle plant, we choose to check whether suitable parameter estimates can be obtained based on responses from the simulation model. Thus, the nominal parameters in the ASM2d plant are chosen as the ‘‘correct’’ parameters θ^* , and then some outputs $y_t(\mathcal{X}) = y_t^m(\theta^* | \mathcal{X})$ are computed. The initial parameter guess θ^0 was chosen as a 10% increase of θ^* . Some responses $y_t(\mathcal{X})$ are displayed in Figure 2 with a sludge age of 6 d. Clearly, the outputs indicate a relatively stiff system (compare e.g. the slow variation of S_{NH_4} to the rapid variation of S_{PO_4}), with relatively little excitation. Hence, difficulties may be expected in finding all parameters. Also, as we have seen, most of the identifiable parameters are related to the phosphorous removal, so we would expect that states related to phosphorous can be fitted better to the data.

In order to find parameter estimates, a weighted least squares criterion is applied

$$\hat{\theta} = \arg \min_{\theta} J = \arg \min_{\theta} e^T W e \quad (5)$$

where $\hat{\theta}$ contains the optimal parameter values of the parameters marked with an asterisk in Table 4 and the other parameters are chosen as θ^0 , W is block diagonal with W_t in the diagonal blocks,

$$W_t = \text{diag} (1/(y_1^{\circ})^2, \dots, 1/(y_n^{\circ})^2),$$

and e is the observation error $y - y^m$ (Dueñas Díez *et al.* 2006).

To compute the parameter estimates, the nonlinear least squares algorithm `lsqnonlin` of the Optimization Toolbox in Matlab is used. Table 5 shows the initial parameter values θ^0 , the ‘‘correct’’ values θ^* , and the estimated values $\hat{\theta}$ obtained after minimizing J .

Table 5. Comparison of initial and estimated parameter values.

Symbol	Description	θ^0	$\hat{\theta}$	θ^*	Unit
μ_{PAO}	Maximum growth rate of PAO	0.0322	0.03156	0.0293	h^{-1}
q_{PP}	Rate constant for storage of X_{PP}	0.0385	0.03876	0.0350	$\text{g } X_{PP} \text{ g}^{-1} X_{PAO} \text{ h}^{-1}$
K_X	Saturation coefficient for particulate COD	0.1100	0.11537	0.1000	$\text{g } X_S \text{ g}^{-1} X_H \text{ h}^{-1}$
μ_H	Maximum growth rate on substrate	0.1513	0.16832	0.1375	$\text{g } X_S \text{ g}^{-1} X_H \text{ h}^{-1}$
q_{fe}	Maximum rate for fermentation	0.0756	0.03970	0.0688	$\text{g } S_F \text{ g}^{-1} X_H \text{ h}^{-1}$
$K_{A,PAO}$	Saturation coefficient for acetate, S_A	4.4000	4.42080	4.0000	g COD m^{-3}
q_{PHA}	Rate constant for storage of X_{PHA} (base X_{PP})	0.0963	0.10698	0.0875	$\text{g } X_{PHA} \text{ g}^{-1} X_{PAO} \text{ h}^{-1}$
K_{PHA}	Saturation coefficient for PHA	0.0110	0.01125	0.0100	$\text{g } X_{PHA} \text{ g}^{-1} X_{PAO}$
b_{HET}	Rate for Lysis of X_H	0.0101	0.01049	0.0092	h^{-1}
η_{fe}	Anoxic hydrolysis reduction factor	0.4400	0.45286	0.4000	—
$K_{PS,PAO}$	Saturation coefficient for phosphorus in storage of PP	0.2200	0.21772	0.2000	g P m^{-3}
b_{PP}	Rate for Lysis of X_{PP}	0.0050	0.00513	0.0046	h^{-1}

Note that for several parameters, $|\theta_j^* - \theta_j^0| < |\theta_j^* - \hat{\theta}_j|$. The reason for this is that we are estimating only a subset of the parameters.

5.6. Model validation

Validation is the comparison of model output $y_t^m(\hat{\theta} | \mathcal{K}')$ with the real output $y_t(\mathcal{K}')$ based on validation data \mathcal{K}' , while parameter estimate $\hat{\theta}$ is based on training data \mathcal{K} . The main reason why the validation data \mathcal{K}' should differ from the training data \mathcal{K} is to avoid fitting the model to noise. In this paper, we constructed the training data from an assumed perfect model, $y_t = y_t^m(\theta^* | \mathcal{K})$, and we choose to use the training data for validation, too. The evolution of some key outputs are displayed in Figure 3: the “real” output $y_t = y_t^m(\theta^*)$ from the biological reactor, and the prediction outputs $\hat{y}_t = y_t^m(\hat{\theta})$. For comparison, we also include the model outputs $y_t^m(\theta^0)$, which will indicate how much the model has been improved through parameter estimation.

The concentration of S_{PO_4} after parameter estimation (dashed red curve in Figure 3) is quite close to the “real” process output (black, solid line). As argued previously, the model simulates bio-P operation, which can be observed from the time evolution of S_{PO_4} . We also argued that due to the bio-P operation, most of the identifiable parameters are related to phosphorous removal. Furthermore, more weight has been put on phosphorous removal outputs because of the governmental restrictions on total phosphorous in the effluent. Thus, the model has mainly been adjusted to give good prediction of those measurements which are related to phosphorous.

The prediction capabilities of other outputs (e.g. S_{NH_4} and S_{NO_3}) are poorer. As argued, the lack of autotrophic biomass (X_{AUT}) in the process leads to poor nitrification. Consequently, parameters related to the autotrophic (nitrifying) organism group are poorly identifiable. However, since there are no government restrictions on the nitrogen outputs, the poor prediction capability of nitrogen outputs is acceptable.

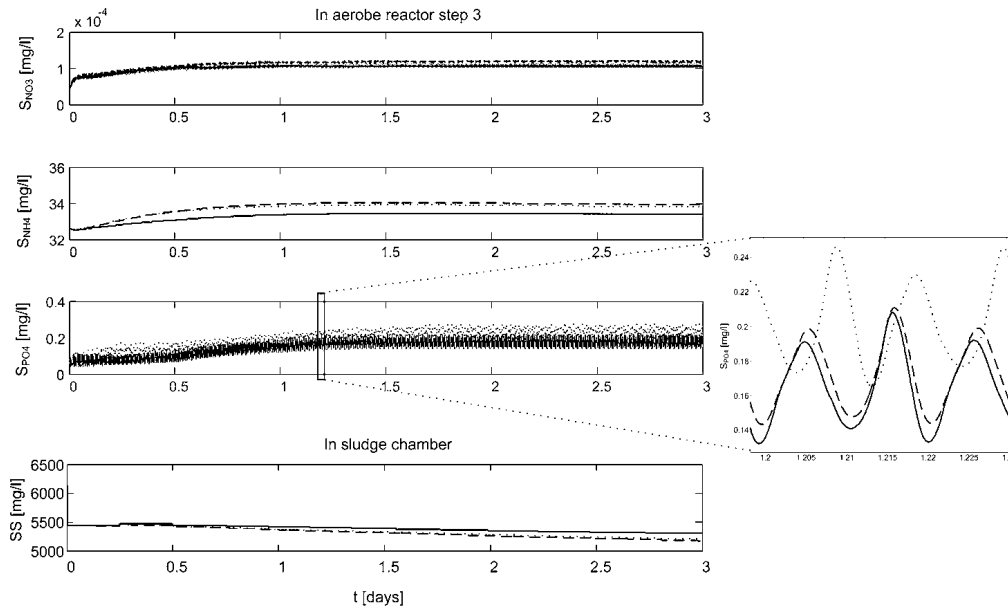


Figure 3. Comparison of model outputs: $y_t = y_t^m(\theta^*)$ (solid, black), $\hat{y}_t = y_t^m(\hat{\theta})$ (dashed, red), and $y_t^m(\theta^0)$ (dotted, blue).

6. Conclusions

In this paper, a method for assessing parameter identifiability of large scale models based on experimental data is discussed. The method is applied to a model of a biological wastewater treatment plant. First, a general description of wastewater treatment is given, with an introduction to the ASM2d dynamic model. The description is unique in the sense that it does not require much background in biological waste water modeling, apart from a background in basic process modeling. Next, an overview is given of a method for analyzing experimental identifiability which has been proposed in the literature. Then, an overview of the Duvbacken wastewater treatment plant in Gävle, Sweden, is given. Finally, the identifiability analysis theory is applied to a model of the Duvbacken plant. It should be emphasized that because real experimental data were unavailable, the study has been based on simulated data assuming a perfect model. However, the procedure can be used without modification with real data.

In the analysis, a model with 45 parameters has been studied. From the experimental identifiability analysis, 12 of the parameters were found to be identifiable. Some details in the analysis are given. Then these 12 parameters are used in a parameter estimation study, and improved parameters are found using a Matlab routine for nonlinear least squares minimization. After finding the parameters, the model is validated against the original experimental data. The model fit is not perfect. However, the results are logical from an understanding of the process operation: the model fit is quite good for those outputs that matter in the operation of the plant, and less good in outputs which are less important to the plant operation.

Some future work may include:

- The current model implementation in Matlab is relatively rigid with respect to possible input signals, and a rewriting of the model in a more flexible modeling language such as Modelica is planned.

- Verifying the model more thoroughly for other operating scenarios.
- Fitting the model to experimental data for larger initial errors in parameters, and comparing different optimization algorithms, as well as multiple shooting vs. single shooting (used here).
- Fitting the model to noisy data/real experimental data.
- Including a statistical analysis of the parameters and prediction capabilities of the model.

References

- BAILEY, J. E. & OLLIS, D. F. (1986). *Biochemical Engineering Fundamentals*, second edn, McGraw-Hill, New York.
- BARD, Y. (1974). *Nonlinear Parameter Estimation*, Academic Press, New York.
- BECK, J. V. & ARNOLD, K. J. (1977). *Parameter Estimation in Engineering and Science*, Wiley, New York.
- BELSLEY, D. A. (1991). *Conditioning Diagnostics. Collinearity and Weak Data in Regression*, John Wiley & Sons, New York.
- BRUN, R., KÜHNI, M., SIEGRIST, H., GUJER, W. & REICHERT, P. (2002). 'Practical identifiability of ASM2d parameters—systematic selection and tuning of parameter subsets', *Water Research* 36(16), 4113–4127.
- BRUN, R., REICHERT, P. & KÜNSCH, H. R. (2001). 'Practical identifiability analysis of large environmental simulation models', *Water Resour. Res.* 37(4), 1015–1030.
- DUEÑAS DÍEZ, M., FJELD, M., ANDERSEN, E. & LIE, B. (2006). 'Validation of a compartmental population balance model of an industrial leaching process: The Silgrain(R) process', *Chem. Engng. Sci.* 61, 229–245.
- HENZE, M., GUJER, W., MINO, T., MATSUO, T., WENTZEL, M. C., MARAIS, G. V. R. & VAN LOOSDRECHT, M. C. M. (1999). 'Activated sludge model no. 2d, ASM2D', *Wat. Sci. Tech.* 39(1), 165–182.
- HENZE, M., GUJER, W., VAN LOOSDRECHT, M. C. M. & MINO, T. (2000). *Activated Sludge Models ASM1, ASM2, ASM2d and ASM3.*, Vol. 9 of *Scientific and Technical Report*, IWA Publishing.
- HENZE, M., HARREMOËS, P., LA COUR JANSEN, J. & ARVIN, E. (1996). *Wastewater Treatment. Biological and Chemical Processes*, second edn, Springer, Berlin.
- HOLMBERG, A. (1982). 'On the parameter identifiability of microbial growth models incorporating michaelis-menten type nonlinearities', *Math. Biosci.* 62, 23–43.
- JEPPSON, U. (1993). On the verifiability of the activated sludge system dynamics, Tekn. lic. dissertation, Lund Institute of Technology.
- LJUNG, L. (1999). *System Identification. Theory for the User*, second edn, Prentice Hall, Upper Saddle River, New Jersey.
- LJUNG, L. & GLAD, S. T. (1994). 'On global identifiability for arbitrary model parameterizations', *Automatica* 30(2), 265–276.
- METCALF & EDDY, I. (2003). *Wastewater Engineering. Treatment and Reuse*, fourth edn, McGraw Hill, New York.
- OLSSON, G. & NEWELL, B. (1999). *Wastewater Treatment Systems. Modelling, Diagnosis and Control*, IWA Publishing, London.
- POHJANPALO, H. (1978). 'System Identifiability Based on the Power Series Expansion of the Solution', *Mathematical Biosciences* 41, 21–33.
- RAWLINGS, J. B. & EKERDT, J. G. (2002). *Chemical Reactor Analysis and Design Fundamentals*, Nob Hill Publishing, Madison, Wisconsin.
- SMITH, W. R. & MISSEN, R. W. (1991). *Chemical Reaction Equilibrium Analysis: Theory and Algorithms*, Krieger Publishing Company, Malabar, Florida.
- SÖDERSTRÖM, T. & STOICA, P. (1989). *System Identification*, Prentice Hall International.
- STEWART, G. (1987). 'Collinearity and least squares regression', *Statistical Science* 2(1), 68–84.
- WALTER, E. & PRONZATO, L. (1997). *Identification of Parametric Models from Experimental Data*, Springer.

Paper B: Parameter Identifiability Analysis and Model Fitting of a Biological Wastewater Model

This paper was presented at the conference and published in the Proceedings of 16th European Symposium on Computer Aided Process Engineering and 9th International Symposium on Process Systems Engineering held in Garmisch-Partenkirchen, Germany, 9 – 13 July 2006, pp. 409 – 414.

Parameter Identifiability Analysis and Model Fitting of a Biological Wastewater Model

Qian Chai^a, Sverre H. Amrani^b, and Bernt Lie^{a*}

^a*Telemark University College, P.O. Box 203, N- 3901 Porsgrunn, Norway*

^b*BioTek AS, N- 3918 Porsgrunn, Norway*

Abstract

A biological wastewater treatment plant situated at Duvbacken in Gävle, Sweden is described. A dynamic model based on the Activated Sludge Model No. 2d (ASM2d) is developed in the simulation language Modelica, and is exported from Dymola (a Modelica environment) into Matlab. It is of interest to adjust the parameters in the model such that the developed model fits better to experimental data from the Duvbacken plant. In this paper, a systematic approach for parameter identifiability analysis is described. A standard least squares criterion is used in combination with a direct optimization method to calculate the improved parameter estimates, and the adjusted model is validated against experimental data from the industrial plant.

Keywords: ASM2d; Identifiability analysis; Parameter estimation; Model fitting

1. INTRODUCTION

The Activated Sludge Models (ASMs) defined by the International Water Association (IWA) have been widely applied to modeling and simulation of wastewater treatment plants (WWTPs) for nitrogen (N) and phosphorus (P) removal, (Gernaey et al. 2004). The ASMs provide a standardized set of basis models with nominal parameter values suggested in publications. In practice, however, plant operation, influent, and sludge characteristics vary from plant to plant. The model parameters are not universal in the sense that every plant could be modeled using the same parameter values. It is thus necessary to fit the model parameters to experimental data from a specific wastewater stream in order to ensure good predictive properties of a model. However, the ASMs hold a large number of parameters, and few measurements are available. This results in the parameters being poorly identifiable from experimental data.

In this paper, we consider the Duvbacken WWTP in Gävle, Sweden, which is designed to treat municipal wastewater from approximately 100.000 person equivalents. The plant is used to remove biological phosphorous primarily using bacteria, and is described by an ASM2d model (Henze et al. 2000). A previous work on analysis of parameter identifiability for this plant was reported in Sarmiento Ferrero et al. (2006). In their work, the model implementation in Matlab was relatively rigid with respect to possible input signals. To give an open and flexible model representation, a more general-purpose modeling language such as Modelica is chosen here. Modelica is an object-oriented modeling language with advantages such as suitability for multi-domain modeling, the usage of general equations of physical phenomena, the reusability

*Corresponding author: Bernt.Lie@hit.no

of model components, and a hierarchical model structure (Fritzson 2004). The model is thus developed in Modelica, based on modifications of the free Modelica library `WasteWater` (Reichl 2003), and the simulation is carried out in the environments Dymola and Matlab.

The paper is organized as follows. In section 2, the implementation of the Duvbacken plant model in Modelica is described. In section 3, parameter identifiability analysis for the Duvbacken model is carried out, and some results of model fitting are given. Finally, some conclusions are drawn.

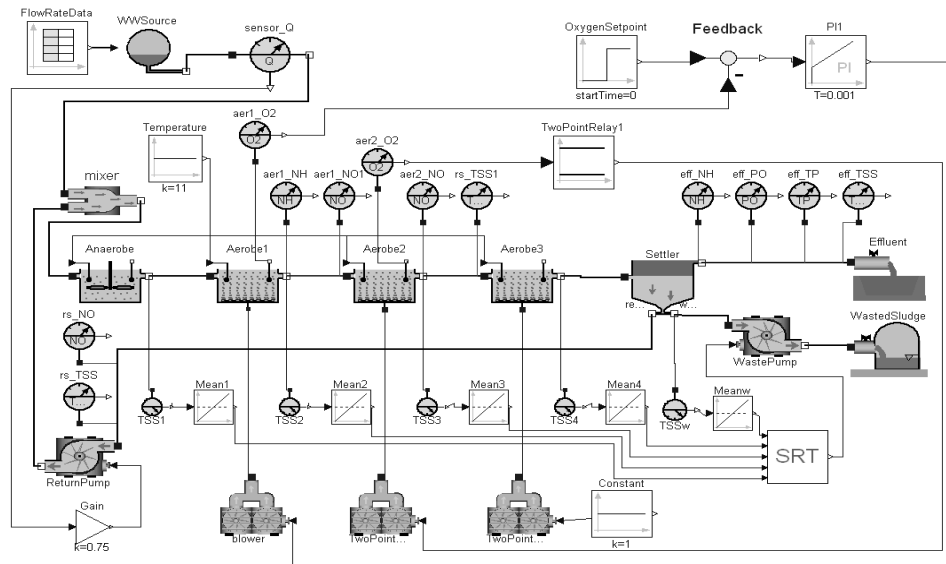


Fig. 1: Model diagram of the Duvbacken plant implemented in Modelica/Dymola.

2. IMPLEMENTATION OF THE DUVBACKEN PLANT MODEL IN MODELICA

The Duvbacken plant consists of 1 anaerobe step, 3 aerobe steps, and 1 secondary sedimentation step (see Fig. 1). More details of the plant process and flow description are given in Sarmiento Ferrero et al. (2006). A dynamic model is developed using a slight modification of ASM2d, where two processes for chemical precipitation of phosphates are neglected. In the dynamic model, each of the 4 steps in the plant (anaerobe, aerobe 1–3) is modeled as a perfectly stirred reactor. With 17 states in each of the reactors, this gives a model with 68 states. The states of the model are grouped into the concentration of dissolved species S_j and particulate species X_j . The model has potentially 18 input variables (feed composition of 17 varying states, and volumetric influent flow Q_{in}), and has 42 (kinetic) parameters considered for parameter identifiability analysis. In the model fitting, 10 measurements are available.

The dynamic model is implemented in Modelica using the Dymola environment (Dynasim 2004). Within Dymola, the model is easily constructed by dragging, dropping, and connecting components chosen from the modified `WasteWater` library. During the simulation, a system of differential algebraic equations (DAE) is established by Dymola. The DASSL integration procedure of Dymola is used to solve the DAE system. Moreover, Dymola provides a convenient interface to Matlab such that Modelica models can be easily executed within Matlab.

The Modelica application library `WasteWater` contains relatively complete components needed for modeling and simulation using ASMs. In addition, with the flexibility of the Modelica language, the user can easily extend the specific components. To apply the `WasteWater`

library to the Duvbacken plant, the following modifications are made:

- A dynamic influent data file including time and influent flowrate Q_{in} is provided in the component `FlowRateData`. The information in the data file is read by the modified component `WWSource`, where influent compositions are defined as functions of Q_{in} .
- Two kinds of biological reactors are developed in the `WasteWater` library: anaerobe/denitrification reactor (`Deni`) and aerobe/nitrification reactor (`Nitri`). All the aerobic reactors are aerated, either by continuous aeration or by intermittent aeration. The air is supplied from blowers. In this plant, three aeration control strategies are applied: (1) Aerobe step 1: setpoint control. A PI controller is used to keep the oxygen concentration around 2 mg/l. (2) Aerobe step 2: zone control. The reactor is aerated by intermittent aeration to restrict the oxygen concentration between 1.4 mg/l and 0.2 mg/l. (3) Aerobe step 3: continuous constant aeration. The oxygen transfer coefficient K_La is made to be 19.2 d^{-1} .
- A part of the sludge from the settler is returned back to the inlet of the anaerobe step, while the remaining sludge is wasted and pumped to the sludge treatment. The return sludge flowrate Q_{rs} is proportional to the influent flowrate, i.e. $Q_{rs} = 0.85 \times Q_{in}$. The waste sludge flowrate Q_w is used to manipulate the sludge age², which is the most critical parameter for activated sludge design. In the plant model, the typical value of the sludge age is 5 d, and Q_w is updated every 6 hours based on the value of the sludge age.
- The initial values of the model are defined in a Modelica script file with file extension `.mos`.

3. PARAMETER IDENTIFIABILITY ANALYSIS FOR THE DUVBACKEN PLANT MODEL

3.1. Execution of the Modelica model in Matlab

Since Matlab has available toolboxes, it is chosen as the simulation platform for identifiability analysis and parameter estimation. The implementation of the parameter estimation is based on the least squares criterion of the model error. The model error/uncertainty $e_t \in \mathbb{R}^{n_y}$ represents the difference between the experimental output y_t and the simulated model output y_t^m with the input signals u_t for $t \in \{1, \dots, T\}$, and we introduce the notation $y, y^m, e \in \mathbb{R}^{n_y \cdot T \times 1}$, while $u \in \mathbb{R}^{n_u \cdot T \times 1}$.

The Modelica model is compiled to an executable Dymosim file (`dymosim.exe`), which can be directly called as an m-function from Matlab:

```
d = dymosim(experiment, x0, param);
```

Here, the three input arguments correspond to the three matrices specified in a Dymosim input file `dsin.txt`. `experiment` is the experiment setup vector, `x0` is the initial state vector, and `param` is the parameter vector. The response from the simulator is `d` which is a Matlab structure containing the simulation results.

3.2. Identifiability analysis

We need a set of experimental input signals u_t in order to carry out the experimental identifiability analysis. Here, we use the same volumetric influent flowrate Q_{in} as in the real plant during the period of October 24 – 29, 2005. Since there are no composition measurements in the plant influent, the influent compositions are calculated from the value of Q_{in} , based on process experience.

²The sludge age or solids retention time (SRT) is the residence time of the sludge in the reactor system.

With the given model and experimental conditions, an experimental parameter identifiability study is carried out based on the systematic approach described in Brun et al. (2001); see also Sarmiento Ferrero et al. (2006) who applied this approach to the Duvbacken plant. Next, the identifiability analysis is discussed based on measures of sensitivity and collinearity.

Model sensitivity

The sensitivity was found by numerically perturbing the simulation model. Column s_j of the sensitivity S was computed as $s_j \approx \frac{y^m(\theta^0 + e_j \cdot \delta\theta_j) - y^m(\theta^0)}{\delta\theta_j}$, where e_j is column j of the identity matrix $I \in \mathbb{R}^{n_\theta \times n_\theta}$. Here, scaling of outputs and parameters is essential.

Sensitivity index

In the method of Brun et al. (2001), the sensitivity measure δ_j^{msqr} is defined as $\delta_j^{\text{msqr}} \triangleq \frac{1}{\sqrt{n_y T}} \|s_j\|_2$. Out of the 42 original parameters, the 20 most sensitive parameters are

$$\theta_S = (\mu_{\text{PAO}}, \mu_{\text{AUT}}, K_{\text{O}_2, \text{HET}}, q_{\text{PP}}, K_h, b_{\text{HET}}, K_{\text{PS, PAO}}, K_{\text{O}_2, \text{PAO}}, q_{\text{fe}}, K_{\text{P, AUT}}, K_X, K_{\text{PHA}}, q_{\text{PHA}}, b_{\text{PP}}, \mu_{\text{H}}, K_{\text{F}}, K_{\text{PP}}, b_{\text{PAO}}, K_{\text{A, PAO}}, b_{\text{AUT}}),$$

where $\delta_j^{\text{msqr}} \in [0.1071, 0.5145]$.

Collinearity indices and identifiable parameters

Collinearity is a measure of how combinations of parameters interact, thus we consider all possible subsets of the 20 parameters of interest from the parameter sensitivity ranking. Brun et al. (2001) defined the collinearity index as $\gamma(\theta_p) \triangleq \frac{1}{\sigma_{\min}(\tilde{S}_{\theta_p})}$, where \tilde{S}_{θ_p} is a $n_y \cdot T \times p$ submatrix of \tilde{S} containing those columns that correspond to the p parameter subset of θ ; \tilde{S} is a slightly modified sensitivity matrix where the columns have been normalized, and $\sigma_{\min}(\tilde{S}_{\theta_p})$ denotes the smallest singular value of \tilde{S}_{θ_p} . In Sarmiento Ferrero et al. (2006), the parameters are considered to be collinear if $\gamma(\theta_p) > 5$. Here it is found that there are parameter subsets with up to 14 elements which fulfill $\gamma(\theta_p) \leq 5$, whereas subsets with 15 and more elements all have $\gamma(\theta_p) > 5$. Therefore, the subset containing the largest number of parameters with the smallest collinearity index $\gamma(\theta_p) \leq 5$ is selected as the best identifiable; the collinearity index is $\gamma(\theta_p) = 4.59$. However, experience reported in Brun et al. (2002) suggest that parameter b_{PAO} is difficult to estimate. We thus choose to take b_{PAO} out of the set of parameters to estimate. In order to ensure quick convergence in the computation of the estimated parameters, we additionally fix the 3 least sensitive parameters b_{PP} , μ_{H} , and b_{AUT} . The chosen final subset contains 10 parameters with $\gamma(\theta_p) = 4.20$.

3.3. Parameter estimation

After finding an identifiable parameter subset based on knowledge of $y_t^m(\theta^0)$, we are ready to estimate the parameters. To do so, we need responses y_t from the plant. The available on-line and off-line measurements of the Duvbacken plant have been indicated in Sarmiento Ferrero et al. (2006). Out of these measurements, the 10 outputs of interest for model fitting are downloaded from the plant data server and are converted to Matlab data files.

To find parameter estimates, a weighted least squares criterion is applied:

$$\hat{\theta} = \arg \min_{\theta} J = \arg \min_{\theta} e^T W e,$$

where $\hat{\theta}$ is the optimal parameter value of the estimated parameters listed in Table 1 and the other parameters are chosen as θ^0 , W is a block diagonal matrix with the appropriate scaling factors, e.g. one of the data points seemed to be an outlier, so the corresponding weighting

factor was chosen as 0. Our criterion J describes a multi-step prediction error, and it is well known that such a criterion may have multiple minima even for linear models (Damslara 1998). A gradient-based method such as Matlab's `lsqnonlin` may thus easily be captured in a local minimum, and it is safer to use a global method such as the Finkel-Kelley `Direct` optimization algorithm (Finkel 2004).

Table 1
Comparison of initial and estimated parameter values.

Parameter	K_h	K_X	q_{fe}	b_{HET}	q_{PP}	μ_{PAO}	$K_{PS,PAO}$	K_{PHA}	μ_{AUT}	$K_{P,AUT}$
θ^0	3.00	0.10	3.00	0.40	1.50	1.00	0.20	0.01	1.00	0.01
$\hat{\theta}$	3.000	0.122	2.889	0.477	1.000	1.000	0.193	0.010	0.852	0.014

3.4. Model validation

The evolution of some key outputs are displayed in Fig. 2: the real output y_t from the plant, and the prediction outputs $\hat{y}_t = y_t^m(\hat{\theta})$ over 6 days. For comparison, we also include the model outputs $y_t^m(\theta^0)$, which will indicate how much the model has been improved through parameter estimation.

The concentrations of soluble phosphorous S_{PO_4} and total phosphorous P_{tot} after parameter estimation (dashed) are quite close to the real data. Although some data points seem to be outliers, the simulation results seem to match the data reasonably well. According to requirements of the plant, the model primarily simulates biological phosphorous removal, which can be observed from the time evolution of S_{PO_4} and P_{tot} , see also Sarmiento Ferrero et al. (2006). Thus, the model has mainly been adjusted to give good prediction of the measurements related to phosphorous.

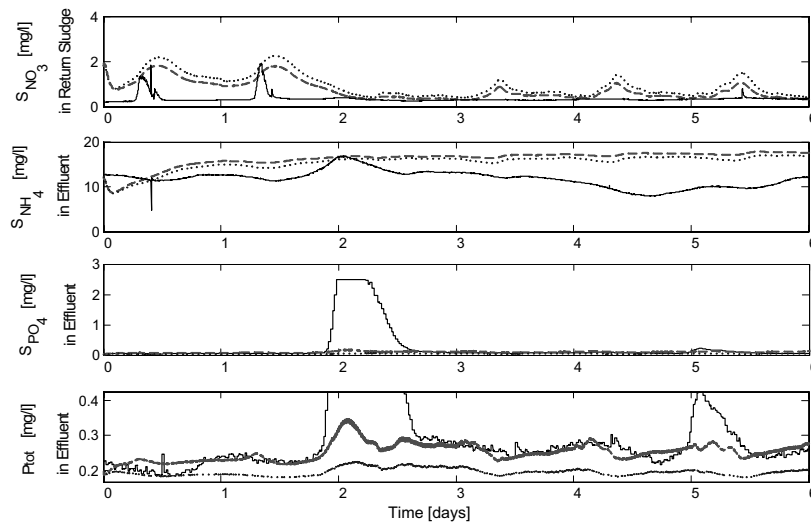


Fig. 2: Comparison of model outputs after parameter estimation \hat{y} (dashed) with the real data y (solid), including initial model output $y^m(\theta^0)$ (dotted).

The prediction of other outputs (e.g. the concentrations of soluble nitrates S_{NO_3} and soluble ammonium S_{NH_4}) are less good. This is probably due to the lack of measurement of the influent compositions from the real plant, such that the input of the model cannot exactly represent the

characteristic of the experimental influent. Otherwise, in the ASM2d, average species from populations are considered in order to simplify the models. The strong simplifications of the ASM2d may also cause the poor prediction. In any case, the model seems to capture the main dynamics of the system. Since we mainly concern ourselves with phosphorous outputs, the relatively poor prediction capability of nitrogen outputs is acceptable.

4. CONCLUSIONS

Parameter estimation and validation of ASM2d is of importance if the models are to be used for prediction. However, the parameters of the ASM2d model are usually poorly identifiable from the available data. In this paper, a method for assessing parameter identifiability of large scale models based on experimental data is discussed. The method is applied to a model of a biological wastewater treatment plant. The improved parameter estimates are found using the Finkel-Kelley `Direct` optimization algorithm for least squares minimization. Finally, the model is validated against the original experimental data. The model fit is quite good for those outputs that matter in the operation of the plant. However, the fit of other outputs is less good. The reason for this is that the influent compositions have not been measured, instead some compositions based on process experience have been used.

Future work will include statistical analysis of the parameters and prediction capabilities of the model. Also, other optimization algorithms may be combined with the `Direct` method to improve the computation speed and accuracy of parameter estimation. Although the influent composition is an important input to the model, it is difficult to measure in practice. It may thus be of interest to fix model parameters and estimate influent compositions instead.

REFERENCES

- Brun, R., Kühni, M., Siegrist, H., Gujer, W. & Reichert, P. (2002), 'Practical identifiability of ASM2d parameters - systematic selection and tuning of parameter subsets', *Water Research* **36**(16), 4113–4127.
- Brun, R., Reichert, P. & Künsch, H. R. (2001), 'Practical identifiability analysis of large environmental simulation models', *Water Resour. Res.* **37**(4), 1015–1030.
- Damslora, A. J. (1998), Optimal Control of Batch Emulsion Polymerisation of Vinyl Chloride, PhD thesis, Norwegian University of Science and Technology.
- Dynasim (2004), *Dymola Multi-Engineering Modeling and Simulation*, Dynasim AB, Lund, Sweden.
- Finkel, D. E. (2004), *DIRECT - A Global Optimization Algorithm*, <http://www4.ncsu.edu/~definkel/research/index.html>.
- Fritzson, P. (2004), *Principles of Object-Oriented Modeling and Simulation with Modelica 2.1*, IEEE Press, Piscataway, NJ. ISBN 0-471-47163-1.
- Gernaey, K. V., Loosdrecht, M. C. V., Henze, M., Lind, M. & Jørgensen, S. B. (2004), 'Activated sludge wastewater treatment plant modelling and simulation: State of the art', *Environmental Modelling and Software* **19**, 763–783.
- Henze, M., Gujer, W., van Loosdrecht, M. C. M. & Mino, T. (2000), *Activated Sludge Models ASM1, ASM2, ASM2d and ASM3.*, Vol. 9 of *Scientific and Technical Report*, IWA Publishing.
- Reichl, G. (2003), 'WasteWater a library for modelling and simulation of wastewater treatment plants in modelica', *Proceedings of the 3rd International Modelica Conference* pp. 171–176.
- Sarmiento Ferrero, C., Chai, Q., Dueñas Díez, M., Amrani, S. H. & Lie, B. (2006), 'Systematic analysis of parameter identifiability for improved fitting of a biological wastewater model to experimental data', *Modeling, Identification and Control*. In press.

Paper C: Comparison of State Estimation Techniques, Applied to a Biological Wastewater Treatment process

This paper was presented at the conference and published in the Proceedings of the 10th International IFAC Symposium on Computer Applications in Biotechnology (CAB - 2007) held in Cancun, Mexico, 4 – 6 June 2007, pp. 353 – 358.

COMPARISON OF STATE ESTIMATION TECHNIQUES, APPLIED TO A BIOLOGICAL WASTEWATER TREATMENT PROCESS

Qian Chai *, Beathe Furenes *, Bernt Lie *,¹

* *Telemark University College, P.O. Box 203, N-3901
Porsgrunn, Norway*

Abstract: In this paper the problem of optimal state estimation in a biological wastewater treatment process (WWTP) is considered. The standard Kalman filter (KF) and its extensions: the extended Kalman filter (EKF) and the unscented Kalman filter (UKF), are used to estimate the unmeasured state. The prediction of the state with the standard KF is poor due to the high nonlinearity of the biological WWTP. Thus, the nonlinear estimation approaches are focused, with a comparison between the EKF and the UKF. The simulation results show that the UKF provides slightly better state estimate than the EKF for both an observable process and an unobservable process. *Copyright © 2007 IFAC*

Keywords: State estimation, Kalman filter, Wastewater treatment process

1. INTRODUCTION

State estimation of dynamic systems is an important prerequisite for safe and economical process operation. There is a wide range of techniques in the literature designed for state estimation in both linear and nonlinear dynamic systems. The *Kalman filter* (KF) (Kalman 1960) is one of the most widely used methods for state estimation and tracking due to its simplicity and robustness. The *standard Kalman filter* is the algorithm of choice in state estimation for linear systems. The approach can be used to reconstruct variables that are not measured and to reduce the effect of noise on the available measurements. However, since most practical systems involve nonlinearity of one kind or another, the problem of nonlinear estimation is extremely important. Therefore, the Kalman filter was early on extended in a somewhat ad hoc way to include nonlinear dynamic models in the *extended Kalman filter* (EKF). For the computation of the filter gain, the EKF simply linearizes the nonlinear models about the current estimate so that the standard Kalman fil-

ter gain computation algorithm can be applied. Although the EKF maintains the elegant and computationally efficient recursive update form of the standard KF, the EKF typically works well only in the region where the first-order Taylor series linearization adequately approximates the nonlinear probability distribution (Crassidis & Junkins 2004). To overcome this limitation, Julier et al. (1995) proposed the *unscented Kalman filter* (UKF) as an alternative to the EKF in nonlinear estimation. The UKF generates a population of so-called *sigma points* based on the current mean and covariance of the state, and permits the direct propagation of the mean and covariance through the actual nonlinear system. The applications and comparisons of the EKF and the UKF have been discussed in a number of studies, e.g. Wan & van der Merwe (2000) and Romanenko & Castro (2004).

In the last decades, due to the increasingly strict requirements on treated wastewater, the need to monitor systems and automatically control biological wastewater treatment processes (WWTPs) is rapidly increasing. To reconstruct the state of the system from the few available measure-

¹ Corresponding author: Bernt.Lie@hit.no

ments and to control important biological variables, appropriate algorithms have to be developed. Some state estimation algorithms for WWTPs have been described in Dochain & Vanrolleghem (2001), and Olsson & Newell (1999). Also, in Keesman (2002), Boaventura et al. (2001), Lukasse et al. (1999), and Zhao & Kümmel (1995), the EKF has been applied to simplified WWTP models, but the application of the UKF to WWTPs is not studied much in the literature.

The objective of this study is to demonstrate (i) the application of the standard KF, the EKF, and the UKF to state estimation of a biological WWTP and (ii) the differences in performance of these estimation approaches. As an example, a nitrogen removal process described by the Activated Sludge Model No. 3 (ASM3) (Gujer et al. 1999) is considered. The state estimation is carried out for the complete ASM3 model.

The paper is organized as follows. In Section 2, the Kalman filter based state estimation approaches are summarized. In Section 3, the approaches are illustrated on a wastewater nitrogen removal process. The comparison between the EKF and the UKF is shown. Finally, some conclusions are drawn in Section 4.

2. STATE ESTIMATION APPROACHES

2.1 Standard Kalman filter (KF)

To better understand how nonlinear estimators work, it is useful to focus on the linear case first. A linear system, approximated in discrete time is given as

$$\begin{aligned}x_{k+1} &= Ax_k + Bu_k + Gw_k \\ y_k &= Cx_k + v_k\end{aligned}$$

where $x_k \in \mathbb{R}^{n_x}$ is the state, $u_k \in \mathbb{R}^{n_u}$ is the control input, and $y_k \in \mathbb{R}^{n_y}$ is the controlled output. $w_k \in \mathbb{R}^{n_w}$ and $v_k \in \mathbb{R}^{n_v}$ present the process noise and measurement noise, respectively. They are assumed to be zero-mean Gaussian noise processes $w_k \sim \mathcal{N}(0, Q_k)$ and $v_k \sim \mathcal{N}(0, R_k)$.

A standard KF formulation for estimating the process state x , where x is assumed to be initially normally distributed with expectation $\hat{x}_{1|1}$ and covariance $P_{1|1}$, is given by:

$$\begin{aligned}\hat{x}_{k+1|k} &= A\hat{x}_{k|k} + Bu_k \\ P_{k+1|k} &= AP_{k|k}A^T + GQ_kG^T \\ K_{k+1} &= P_{k+1|k}C^T (CP_{k+1|k}C^T + R_{k+1})^{-1} \\ \hat{x}_{k+1|k+1} &= \hat{x}_{k+1|k} + K_{k+1}(y_{k+1} - C\hat{x}_{k+1|k}) \\ P_{k+1|k+1} &= (I - K_{k+1}C)P_{k+1|k}(I - K_{k+1}C)^T \\ &\quad + K_{k+1}R_{k+1}K_{k+1}^T,\end{aligned}$$

where y_{k+1} denotes the actual measurement at step $k+1$. The Kalman gain K_k is chosen to minimize the *a posteriori* estimate error covariance $P_{k|k}$. Note that the equation for the *a posteriori* state covariance $P_{k|k}$ is called the stabilized implementation, because it has better numerical properties than some other frequently used equations for $P_{k|k}$, e.g. $P_{k|k} = (I - K_kC)P_{k|k-1}$.

2.2 Extended Kalman filter (EKF)

For several reasons, state estimation for nonlinear systems is considerably more difficult and admits a wider variety of solutions than the linear problem (Crassidis & Junkins 2004). In this paper, we discuss two different approaches to achieve estimation for nonlinear systems: the method of EKF uses the recursive update form of standard KF by linearizing the nonlinear model at each time step, and the method of UKF uses a set of appropriately chosen weighted points to parameterize the means and covariances of probability distributions.

The nonlinear system is described by the stochastic model

$$\begin{aligned}x_{k+1} &= f(x_k, u_k, w_k) \\ y_k &= g(x_k, u_k, v_k)\end{aligned}$$

where $x_1 \sim \mathcal{N}(\hat{x}_{1|1}, P_{1|1})$, $w_k \sim \mathcal{N}(0, Q_k)$, and $v_k \sim \mathcal{N}(0, R_k)$. $\hat{x}_{1|1}$, $P_{1|1}$, Q_k , and R_k are assumed to be known.

The EKF algorithm can be summarized as the following set of equations:

$$\begin{aligned}\hat{x}_{k+1|k} &= f(\hat{x}_{k|k}, u_k, 0) \\ P_{k+1|k} &= A_kP_{k|k}A_k^T + G_kQ_kG_k^T \\ K_{k+1} &= P_{k+1|k}C_{k+1}^T (C_{k+1}P_{k+1|k}C_{k+1}^T \\ &\quad + H_{k+1}R_{k+1}H_{k+1}^T)^{-1} \\ \hat{x}_{k+1|k+1} &= \hat{x}_{k+1|k} + K_{k+1}(y_{k+1} - g(\hat{x}_{k+1|k}, u_{k+1}, 0)) \\ P_{k+1|k+1} &= (I - K_{k+1}C_{k+1})P_{k+1|k}(I - K_{k+1}C_{k+1})^T \\ &\quad + K_{k+1}H_{k+1}R_{k+1}H_{k+1}^TK_{k+1}^T,\end{aligned}$$

where matrices A_k , G_k , C_k , and H_k are defined as the following Jacobians:

$$\begin{aligned}A_k &= \left. \frac{\partial f}{\partial x} \right|_{\hat{x}_{k|k}, u_k, 0}, \quad G_k = \left. \frac{\partial f}{\partial w} \right|_{\hat{x}_{k|k}, u_k, 0} \\ C_k &= \left. \frac{\partial g}{\partial x} \right|_{\hat{x}_{k|k-1}, u_k, 0}, \quad H_k = \left. \frac{\partial g}{\partial v} \right|_{\hat{x}_{k|k-1}, u_k, 0}.\end{aligned}$$

2.3 Unscented Kalman filter (UKF)

We initiate the UKF by

$$x_1 \sim \mathcal{N}(\hat{x}_{1|1}, P_{1|1}).$$

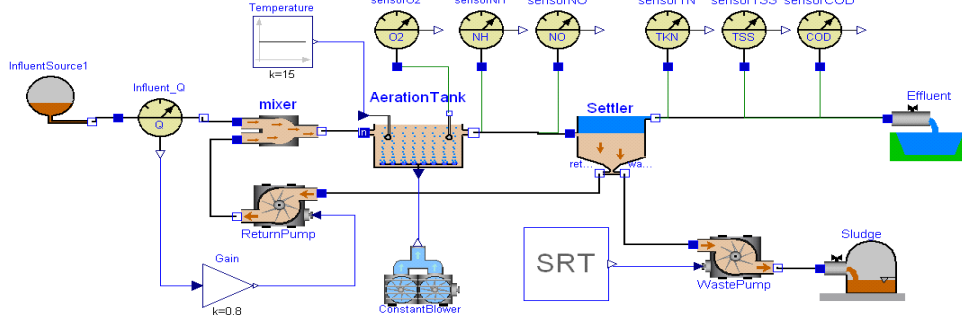


Fig. 1. Model diagram of the nitrogen removal plant implemented in Modelica/Dymola.

First, an augmented state is defined as

$$x_k^a = \begin{bmatrix} x_k \\ w_k \\ v_k \end{bmatrix}.$$

The mean and covariance of the augmented state vector are

$$\tilde{x}_{k|k} = \begin{bmatrix} \hat{x}_{k|k} \\ 0 \\ 0 \end{bmatrix}, \quad \tilde{P}_{k|k} = \begin{bmatrix} P_{k|k} & 0 & 0 \\ 0 & Q_k & 0 \\ 0 & 0 & R_k \end{bmatrix}.$$

The size of $\tilde{x}_{k|k}$ is $n_{\tilde{x}} = n_x + n_w + n_v$. Introduce matrix V as the Cholesky root of $(n_{\tilde{x}} + \lambda) \tilde{P}_{k|k}$:

$$V^T V = (n_{\tilde{x}} + \lambda) \tilde{P}_{k|k}$$

where $\lambda = \alpha^2 (n_{\tilde{x}} + \kappa) - n_{\tilde{x}}$ is a scaling parameter (typically $\alpha \in [10^{-4}, 1]$ and $\kappa = 0$). Let V_i denote column i of matrix V .

Next, we define $2n_{\tilde{x}} + 1$ sigma points $\tilde{\mathcal{X}}_{k|k}^i$ with corresponding weights W_i as follows:

$$\begin{aligned} \tilde{\mathcal{X}}_{k|k}^0 &= \tilde{x}_{k|k} \\ \tilde{\mathcal{X}}_{k|k}^i &= \tilde{x}_{k|k} + V_i \quad i = 1, \dots, n_{\tilde{x}} \\ \tilde{\mathcal{X}}_{k|k}^i &= \tilde{x}_{k|k} - V_{i-n_{\tilde{x}}} \quad i = n_{\tilde{x}} + 1, \dots, 2n_{\tilde{x}} \\ W_m^0 &= \lambda / (n_{\tilde{x}} + \lambda) \\ W_c^0 &= \lambda / (n_{\tilde{x}} + \lambda) + (1 - \alpha^2 + \beta) \\ W_m^i &= W_c^i = 1 / [2(n_{\tilde{x}} + \lambda)] \quad i = 1, \dots, 2n_{\tilde{x}} \end{aligned}$$

where β is used to incorporate prior knowledge of the distribution (with Gaussian distribution $\beta = 2$).

Define $\mathcal{X}_{k|k}^i$ to be the vector of the first n_x elements of $\tilde{\mathcal{X}}_{k|k}^i$, $\mathcal{W}_{k|k}^i$ to be the vector of the next n_w elements of $\tilde{\mathcal{X}}_{k|k}^i$, and $\mathcal{V}_{k|k}^i$ to be the vector of the last n_v elements of $\tilde{\mathcal{X}}_{k|k}^i$. The predicted sigma points $\mathcal{X}_{k+1|k}^i$ and $\mathcal{Y}_{k+1|k}^i$ are given by:

$$\begin{aligned} \mathcal{X}_{k+1|k}^i &= f(\mathcal{X}_{k|k}^i, u_k, \mathcal{W}_{k|k}^i), \quad i = 0, \dots, 2n_{\tilde{x}} \\ \mathcal{Y}_{k+1|k}^i &= g(\mathcal{X}_{k+1|k}^i, u_k, \mathcal{V}_{k|k}^i), \quad i = 0, \dots, 2n_{\tilde{x}}. \end{aligned}$$

We then compute the predicted means as

$$\begin{aligned} \hat{x}_{k+1|k} &= \sum_{i=0}^{2n_{\tilde{x}}} W_m^i \mathcal{X}_{k+1|k}^i \\ \hat{y}_{k+1|k} &= \sum_{i=0}^{2n_{\tilde{x}}} W_m^i \mathcal{Y}_{k+1|k}^i, \end{aligned}$$

and the predicted covariances as

$$\begin{aligned} P_{k+1|k} &= \sum_{i=0}^{2n_{\tilde{x}}} W_c^i [\mathcal{X}_{k+1|k}^i - \hat{x}_{k+1|k}] [\mathcal{X}_{k+1|k}^i - \hat{x}_{k+1|k}]^T \\ E_{k+1|k} &= \sum_{i=0}^{2n_{\tilde{x}}} W_c^i [\mathcal{Y}_{k+1|k}^i - \hat{y}_{k+1|k}] [\mathcal{Y}_{k+1|k}^i - \hat{y}_{k+1|k}]^T \\ S_{k+1|k} &= \sum_{i=0}^{2n_{\tilde{x}}} W_c^i [\mathcal{X}_{k+1|k}^i - \hat{x}_{k+1|k}] [\mathcal{Y}_{k+1|k}^i - \hat{y}_{k+1|k}]^T, \end{aligned}$$

where $E_{k+1|k}$ represents the innovation covariance, and $S_{k+1|k}$ represents the cross correlation matrix between $\hat{x}_{k+1|k}$ and $\hat{y}_{k+1|k}$.

We can now compute the UKF gain K_{k+1} as

$$K_{k+1} = S_{k+1|k} E_{k+1|k}^{-1},$$

and finally the updated estimates of state and covariance as

$$\begin{aligned} \hat{x}_{k+1|k+1} &= \hat{x}_{k+1|k} + K_{k+1} (y_{k+1} - \hat{y}_{k+1|k}) \\ P_{k+1|k+1} &= P_{k+1|k} - K_{k+1} E_{k+1|k} K_{k+1}^T. \end{aligned}$$

Compared with the EKF, the UKF features two main advantages: (i) the UKF avoids the computation of the Jacobian matrices, and (ii) the UKF achieves a third order accuracy in the Taylor series expansion of the statistics of the state, whereas the EKF is at most first order accurate.

3. APPLICATION TO A NITROGEN REMOVAL PROCESS

3.1 Process model

In this study, a laboratory-scale nitrogen removal plant is considered. More details of the process are given in Chai et al. (2006). The process model consists of an aeration tank model in which oxygen

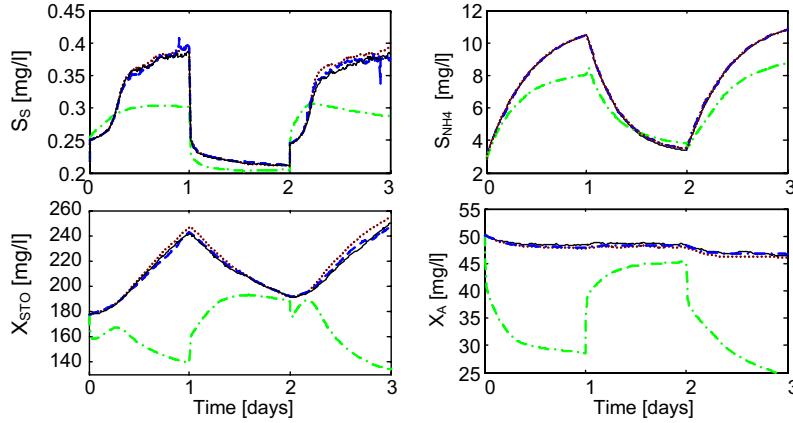


Fig. 2. Simulation results of the standard KF (dash-dot), the EKF (dotted), and the UKF (dashed). The solid line represents the true states.

is supplied, and a settler model where the sludge is separated from the liquid being treated. ASM3 describes the biological processes involved in the aeration tank. The states of the model are grouped into the concentration of soluble components S_j and particulate components X_j . The kinetic and stoichiometric parameter values are taken from the typical parameter values suggested in ASM3 publications. Assuming perfect mixing in the reactor, the mass balance in the aeration tank results in:

$$\frac{dx_a}{dt} = \frac{Q_{in}x_{in} + Q_{rs}x_{rs} - (Q_{in} + Q_{rs})x_a}{V_a} + r + e_1 \mathcal{A}_{O_2},$$

where x_{in} , x_{rs} , $x_a \in \mathbb{R}^{13}$ contain the concentrations in the influent, in the recycled sludge, and in the reactor, respectively; their components are

$$x_l = [S_{O_2,l} \ S_{I,l} \ S_{S,l} \ S_{NH_4,l} \ S_{N_2,l} \ S_{NOX,l} \ S_{ALK,l} \ X_{I,l} \ X_{S,l} \ X_{H,l} \ X_{STO,l} \ X_{A,l} \ X_{SS,l}]^T,$$

$l \in \{in, rs, a\}$; $r \in \mathbb{R}^{13}$ is the reaction rates of each component (defined in ASM3); e_1 is the standard basis for the first coordinate in \mathbb{R}^{13} . \mathcal{A}_{O_2} describes the oxygen transfer: $\mathcal{A}_{O_2} = K_L a (S_{O_2}^{sat} - S_{O_2,a})$, where $K_L a$ is the oxygen transfer coefficient which is the control input variable in this study, and $S_{O_2}^{sat}$ is the saturated dissolved oxygen concentration.

Assuming that the settler is a perfect splitter, the resulting mass balance equations are as follows:

Effluent concentration:

$$S_{j,eff} = S_{j,a}, \quad X_{j,eff} = 0$$

Recycled sludge concentration:

$$S_{j,rs} = S_{j,a}, \quad X_{j,rs} = \frac{Q_{in} + Q_{rs}}{Q_{rs} + Q_w} X_{j,a}.$$

The dynamic model is implemented in the object-oriented modeling language Modelica (Fritzson

2004) using the Dymola simulation environment (Dynasim 2004), see Figure 1, based on modifications of the free Modelica library `WasteWater` (Reichl 2003). Dymola generates a convenient interface to Matlab such that Modelica models can be executed within MATLAB.

3.2 Results and discussion

MATLAB is used for implementation of the estimation approaches. To estimate the state, we need measurements from the plant. As an initial study before doing state estimation based on real data from the plant, we choose to check whether reasonable state estimates can be obtained based on response from the simulation model.

Observability is an important concept in dynamic systems in general. Observability depends mainly on the choice of measurement variables. In this case, to ensure that the system is fully observable, the following measurement variables are chosen: the concentration of total nitrogen TN in the effluent, as well as the concentrations of dissolved oxygen S_{O_2} , dinitrogen S_{N_2} , inert particulate organics X_I , and total suspended solids X_{SS} in the aeration tank². The state and measurement vectors are subject to zero mean additive white noise with covariances

$$Q_k = Q = \text{diag} \left((x_o \cdot 10^{-3})^2 \right) \quad \text{and} \\ R_k = R = \text{diag} \left((y_o \cdot 10^{-4})^2 \right),$$

respectively, where (x_o, y_o) is the operating point. The chosen noise levels are relatively low in this introductory phase of the study. Larger noise covariances will be considered in future work.

² In practice, some of these measurement variables may be difficult to measure; in this study, we assume that they are available.

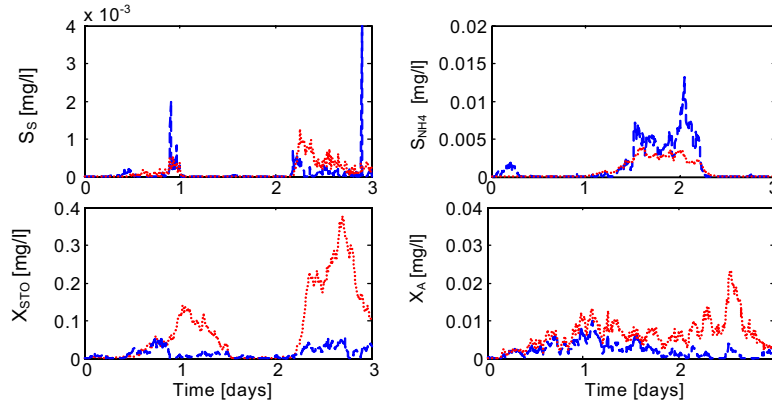


Fig. 3. Comparison of the estimation errors of the EKF (dotted) and the UKF (dashed) for the *observable* process.

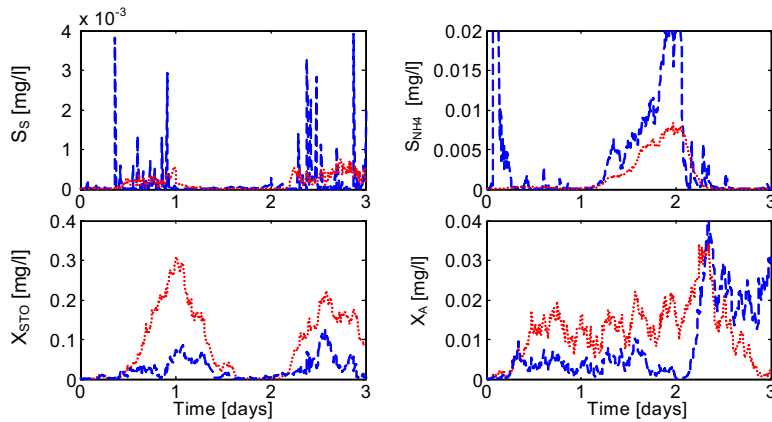


Fig. 4. Comparison of the estimation errors of the EKF (dotted) and the UKF (dashed) for the *unobservable* process.

Figure 2 illustrates the performances of the standard KF, the EKF, and the UKF. Four unmeasured states, which most clearly exhibit the difference of the estimation results between the EKF and the UKF, are shown in the figure: the concentrations of readily biodegradable substrates S_S , ammonium S_{NH_4} , organics stored by heterotrophs X_{STO} , and autotrophic biomass X_A . The prediction of the states with the standard KF is poor because of the nonlinear characteristics of the biological WWTP dynamics. Nonlinear estimation is therefore imperative for biological WWTPs. In Figure 2, both the EKF and the UKF track all the states of the process; indeed, the performance of the EKF and the UKF are quite similar. To compare the estimation qualities of the EKF and the UKF, Figure 3 plots the relative squared estimation errors of the filters which are relative to the true states. The result shows that the EKF and the UKF give the similar estimation error for the concentrations of the soluble components (S_S , S_{NH_4}). However, for the concentrations of the particulate components (X_{STO} , X_A), the superior performance of the UKF is clear.

In WWTPs, measuring most state variables is either difficult or expensive. This time, we assume that only the concentration of total nitrogen TN in the effluent and the concentration of dissolved oxygen S_{O_2} in the aeration tank are the available measurement variables for the nitrogen removal process. By performing observability analysis of a linearized model, we find that this system is not observable. However, the system is detectable since the dynamics of the filter error is asymptotically stable; detectability guarantees convergence of the filter. Figure 4 depicts the estimation errors of the unobservable process with the EKF and the UKF. By comparing Figure 4 with Figure 3, we find that the estimation errors of the unobservable process are slightly larger than the observable process for both the EKF and the UKF. With the UKF, the state estimates for the unobservable process are more noisy than for the observable process. However, for the unobservable (but detectable) process, the EKF and the UKF still give reasonable estimates for the unmeasured states. Hence, if the available measurements are adequate for the dynamic system's requirements,

Table 1. Mean estimation errors over 3 days.

Method	S_S	S_{NH_4}	X_{STO}	X_A
Observable process				
EKF	1.3×10^{-4}	8.5×10^{-4}	7.6×10^{-2}	6.4×10^{-3}
UKF	7.6×10^{-5}	1.5×10^{-3}	1.4×10^{-2}	2.2×10^{-3}
Unobservable process				
EKF	1.5×10^{-4}	1.5×10^{-3}	8.9×10^{-2}	1.2×10^{-2}
UKF	1.3×10^{-4}	4.4×10^{-3}	2.5×10^{-2}	8.4×10^{-3}

a fully observable system may not be necessary (Crassidis & Junkins 2004).

Finally, to give an overview of the estimation capabilities of the EKF and the UKF, the mean estimation errors for the aforementioned states over 3 days are summarized in Table 1. For the most states, UKF provides better state estimation than the EKF for both the observable process and the unobservable process (also see Figures 3 and 4).

4. CONCLUSIONS

In this paper, we investigate the use of the standard KF, the EKF, and the UKF in state estimation of a typical biological WWTP and compare the differences in performance of these estimation approaches. First, the general formulations of the Kalman filter based algorithms are introduced. Then, the performances of the estimation algorithms are evaluated by simulation studies. The estimation results show that state estimation with the linear approach (the standard KF) is poor because of the high nonlinearity of the bioprocess. It is therefore important to exploit nonlinear estimation algorithms for biological WWTPs. For the nitrogen removal process, the UKF provides more accurate estimates than the EKF for both the observable process and the unobservable process (see Table 1), which attests to the power of using the UKF for nonlinear systems. In the computation, the EKF takes 8 min, while the UKF takes 23 min, i.e., the computation time of the UKF is about 3 times longer than that of the EKF. The reason for this is that the calculation of the predicted sigma points in the UKF is time-consuming for the nitrogen removal model. Nevertheless, since the UKF avoids the computation of the Jacobian matrices, the UKF is simpler to implement than the EKF.

REFERENCES

Boaventura, K., Roqueiro, N., Coelho, M. & Araújo, O. (2001), ‘State observers for a biological wastewater nitrogen removal process in a sequential batch reactor’, *Bioresource Technology* **79**, 1–14.

Chai, Q., Bakke, R. & Lie, B. (2006), ‘Object-oriented modeling and optimal control of a biological wastewater treatment process’, *Proceedings of the Eighth IASTED International Conference on Control and Applications* pp. 218–223. ISBN: 0-88986-553-1.

Crassidis, J. & Junkins, J. (2004), *Optimal estimation of dynamic systems*, ISBN 1-58488-391-X, Chapman & Hall/CRC.

Dochain, D. & Vanrolleghem, P. (2001), *Dynamical modelling and estimation in wastewater treatment processes*, ISBN 1900222 50 7, IWA Publishing, London, UK.

Dynasim (2004), *Dymola Multi-Engineering Modeling and Simulation*, Dynasim AB, Lund, Sweden.

Fritzson, P. (2004), *Principles of Object-Oriented Modeling and Simulation with Modelica 2.1*, IEEE Press, Piscataway, NJ. ISBN 0-471-47163-1.

Gujer, W., Henze, M., Mino, T. & van Loosdrecht, M. (1999), ‘Activated Sludge Model No. 3’, *Wat. Sci. Tech.* **39**(1), 183–193.

Julier, S., Uhlmann, J. & Durrant-Whyte, H. (1995), ‘A new approach for filtering nonlinear systems’, *American control conference* pp. 1628–1632.

Kalman, R. (1960), ‘A new approach to linear filtering and prediction problems’, *the ASME-Journal of Basic Engineering* **82 (Series D)**, 35–45.

Keesman, K. (2002), ‘State and parameter estimation in biotechnical batch reactors’, *Control Engineering Practice* **10**, 219–225.

Lukasse, L., Keesman, K. & van Straten, G. (1999), ‘A recursively identified model for short-term predictions of nh4/no3-concentrations in alternating activated sludge processes’, *Journal of Process Control* **9**, 87–100.

Olsson, G. & Newell, B. (1999), *Wastewater Treatment Systems. Modelling, Diagnosis and Control*, IWA Publishing, London.

Reichl, G. (2003), ‘WasteWater a library for modelling and simulation of wastewater treatment plants in modelica’, *Proceedings of the 3rd International Modelica Conference* pp. 171–176.

Romanenko, A. & Castro, J. (2004), ‘The unscented filter as an alternative to the EKF for nonlinear state estimation: a simulation case study’, *Computers and Chemical Engineering* **28**, 347–355.

Wan, E. & van der Merwe, R. (2000), The unscented Kalman filter for nonlinear estimation, in ‘Proc. of IEEE Symposium 2000 (AS-SPCC)’, Lake Louise, Alberta, Canada.

Zhao, H. & Kümmel, M. (1995), ‘State and parameter estimation for phosphorus removal in an alternating activated sludge process’, *Journal of Process Control* **5**(5), 341–351.

Paper D: Object-oriented Modeling and Optimal Control of a Biological Wastewater Treatment Process

This paper was presented at the conference and published in the Proceedings of the Eighth IASTED International Conference on Control and Applications (CA 2006) held in Montreal, Quebec, Canada, 24 – 26 May 2006, pp. 218 – 223.

OBJECT-ORIENTED MODELING AND OPTIMAL CONTROL OF A BIOLOGICAL WASTEWATER TREATMENT PROCESS

Qian Chai
Telemark University College
P.O. Box 203, Kjoelnes ring 56
N-3901 Porsgrunn, Norway
email: Qian.Chai@hit.no

Rune Bakke, Bernt Lie
Telemark University College
P.O. Box 203, Kjoelnes ring 56
N-3901 Porsgrunn, Norway
email: Rune.Bakke@hit.no, Bernt.Lie@hit.no

ABSTRACT

This paper deals with optimal control of a biological wastewater treatment process. A pilot plant is designed to remove nitrogen with simultaneous nitrification/denitrification in a single reactor. The objective of the control in this paper is to force the total nitrogen concentration to track given set points. The dynamic modeling and start-up simulation of the process are first presented based on the ASM3 model and the object-oriented modeling language Modelica. Next, the optimization problem is specified and formulated using a quadratic performance index. Dynamic optimization is used in model predictive control (MPC) based on a linearized model. The control strategy is successfully applied to the simulation model of the pilot plant.

KEY WORDS

Wastewater treatment; ASM3 model; Optimal control; Predictive control

1 Introduction

Water is one of our most precious resources, hence wastewater treatment is a large and growing industry throughout the world. Wastewater contains large quantities of organic matter that tends to deplete oxygen in the water. Furthermore, nitrogen and phosphorus content lead to algae growth. It is thus of importance to reduce the content of organic matter, as well as nitrogen and phosphorus. As one of the means of biological treatment, the activated sludge process is commonly used in wastewater treatment [1].

For small-sized wastewater treatment plants which are appropriate for communities of about 10,000 person equivalents (p.e.) [2], the process generally consists of a single aeration tank in which oxygen is either supplied by diffusers or surface turbines, and a settler where the microbial culture is separated from the liquid being treated, see Figure 1. In the aeration tank, the micro organisms are used to convert the colloidal and dissolved carbonaceous organic matter into various gases and into cell tissue. Because cell tissue has a specific gravity slightly greater than that of water, the resulting cells can be removed from the treated liquid by gravity settling in the settler. Most of the cells are recycled and mixed with incoming wastewater to

maintain convenient sludge age characteristics and high reaction rates. Specifically, for nitrogen removal, nitrification and denitrification occur simultaneously in the unique aeration tank. Nitrogen is removed in these two steps: ammonium is first oxidized to nitrate under nitrification/the aerobic step; the produced nitrate is then transformed into nitrogen gas under denitrification/the anoxic step [3].

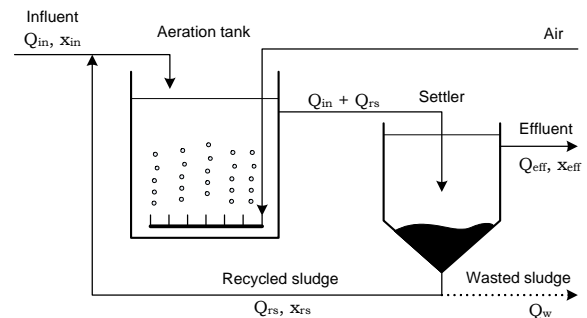


Figure 1: Schematic diagram of a typical activated sludge process.

The simultaneous nitrification-denitrification (SNdN) applications require oxygen control or other types of control methods to assure that both nitrification and denitrification occur in a single tank. Most of small size wastewater treatment plants use very simple control strategies (e.g. time control, manual control) or implement more "advanced" controllers such as proportional controllers [4]. Although biological removal of organic matter from wastewater is correctly handled in most cases by means of the aforementioned controls, the nitrogen concentration in the treated wastewater may significantly exceed the allowed levels.

In this study, a laboratory-scale nitrogen removal plant is considered. The objective is to develop aeration strategies for optimal nitrogen removal by using dynamic optimization methods. The application of dynamic optimization methods requires a sufficiently accurate mathematical model describing the wastewater treatment process. The present work uses the recently developed Activated Sludge Model No. 3 (ASM3) [5]. ASM3 is for the time being considered the best choice in the modeling of the biological processes involved in the activated sludge tank, especially in dealing with the aerobic and anoxic processes in

the same tank [6]. To give an open and flexible model representation, the general-purpose modeling language Modelica is chosen. Modelica is an object-oriented modeling language with advantages such as suitability for multi-domain modeling, the usage of general equations of physical phenomena, the reusability of model components, and a hierarchical model structure [7].

The paper is organized as follows. In section 2, the dynamic model of the nitrogen removal plant is described and implemented in Modelica. In section 3, the optimization problem is specified; the optimization method and results are also discussed. In section 4, the optimization method is used in a closed loop by applying model predictive control (MPC). Finally, some conclusions are drawn in section 5.

2 Process Model

2.1 Nitrogen removal plant model

The pilot nitrogen removal process consists of a unique aeration tank ($V_a = 40\text{ l}$) equipped with an air blower with diffuser which provides oxygen, and a mixer which makes the tank behave as continuous stirred tank reactor (CSTR). The settler is a cylindrical pipe located in the aeration tank, where the sludge is either recycled to the aeration tank or extracted from the system. The recycled sludge is proportional to the influent flowrate, i.e. $Q_{rs} = 0.8 \times Q_{in}$. The waste sludge flowrate Q_w is used to manipulate the solids retention time SRT^1 , which is the most critical parameter for activated sludge design. The process is operated with constant influent flow rate ($Q_{in} = 96\text{ l/d}$) and compositions.

Hence, the plant model consists of an aeration tank model in which the microbiological processes take place, and a settler model where the liquid and solid phases are separated.

Aeration tank model

ASM3 describes the biological processes involved in the aeration tank. The states of the model are grouped into the concentration of soluble components S_j and particulate components X_j . ASM3 consists of 13 state variables and 36 parameters, see [5] and [8]. The kinetic and stoichiometric parameter values are taken from typical parameter values suggested in ASM3 publications.

Assuming perfect mixing in the reactor, the mass balance in the aeration tank results in:

$$\frac{dx_a}{dt} = \frac{Q_{in}x_{in} + Q_{rs}x_{rs} - (Q_{in} + Q_{rs})x_a}{V_a} + r + e\mathcal{A}_O \quad (1)$$

¹The solids retention time (SRT) is the residence time of the sludge in the reactor system. The simplified definition of the SRT is $\text{SRT} = V_a/Q_w$.

where $x_{in}, x_{rs}, x_a \in \mathbb{R}^{13}$ contain the concentrations in the influent, in the recycled sludge, and in the reactor, respectively; their components are

$$x_l = [S_{O_2,l} S_{I,l} S_{S,l} S_{NH_4,l} S_{N_2,l} S_{NOX,l} S_{ALK,l} X_{I,l} X_{S,l} X_{H,l} X_{STO,l} X_{A,l} X_{SS,l}]^T,$$

$l \in \{in, rs, a\}$; $r \in \mathbb{R}^{13}$ is the vector formed by the reaction rates of each component (defined in ASM3); e is a 13-dimensional auxiliary vector with 1 in the first element and 0 in the other elements of the vector. The mass balance equation contains the additional term \mathcal{A}_O which describes the oxygen transfer from the diffusers:

$$\mathcal{A}_O = K_L a (S_O^{sat} - S_{O,a})$$

where $K_L a$ is the oxygen transfer coefficient which is a function of the supplied air flowrate, and S_O^{sat} is the saturated dissolved oxygen concentration ($S_O^{sat} \approx 10\text{ g/m}^3$).

Settler model

In this work, the settler is considered a perfect splitter: (i) the sum of the three outflows (effluent, wasted sludge, and recycled sludge flow rates) equals the settler influent flow rate; (ii) the effluent is exclusively constituted of soluble components. The resulting mass balance equations are as follows:

- Effluent concentration

$$S_{j,\text{eff}} = S_{j,a}, \quad X_{j,\text{eff}} = 0 \quad (2)$$

- Recycled sludge concentration

$$S_{j,rs} = S_{j,a}, \quad X_{j,rs} = \frac{Q_{in} + Q_{rs}}{Q_{rs} + Q_w} X_{j,a}. \quad (3)$$

General model

Eliminating x_{rs} from Eq. 1 using equations 2 and 3, one can reformulate the dynamic system as a set of ODEs:

$$\frac{dx_a}{dt} = f(x_a, t).$$

2.2 Model implementation in Modelica

The dynamic model is implemented in Modelica using the Dymola simulation environment [9], based on modifications of the free Modelica library `WasteWater` [10]. Compared with most modeling languages available today (e.g. Matlab), Modelica offers important advantages for the implementation of the model. The object-oriented architecture allows the user to move and change a component without modifying the rest of the model, and the components can be reused in various models. Multi-domain capability gives the possibility to easily combine electrical, mechanical etc. models, as well as control systems etc. within

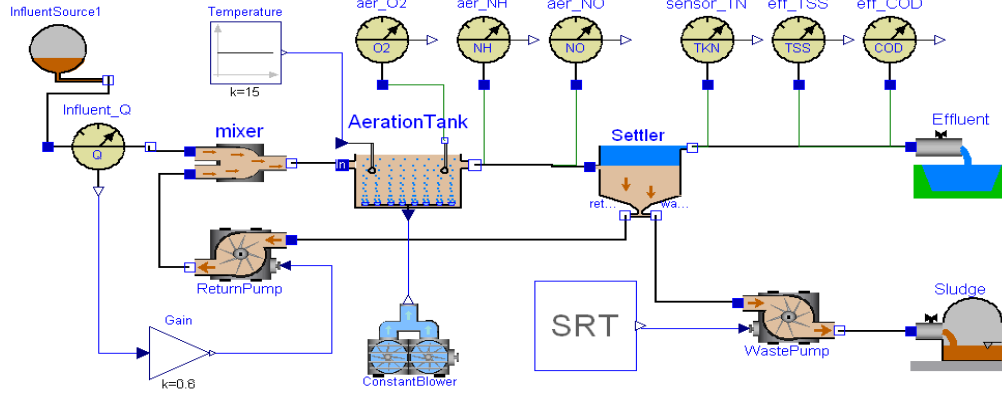


Figure 2: Model diagram of the nitrogen removal plant implemented in Modelica/Dymola.

in the same model. The model diagram of the wastewater treatment process is shown in Figure 2.

Within Dymola, the model is easily constructed by dragging, dropping, and connecting components chosen from the modified `WasteWater` library. During the simulation, a system of differential algebraic equations (DAE) is established by Dymola. The DASSL integration procedure of Dymola is used to solve the DAE system. Moreover, Dymola generates a convenient interface to Matlab such that Modelica models can be executed within Matlab. An application executing a Modelica model in Matlab is described in [11].

To identify the continuous operation dynamics, simulation of the start-up period of the plant is provided by starting with the aeration tank filled by the incoming wastewater. The influent wastewater composition is

$$x_{in} = [0, 30, 60, 16, 0, 0, 5, 25, 115, 30, 0, 1, 125]^T.$$

A continuous air supply is employed for 30 days with $K_L a = 50 \text{ d}^{-1}$ and $\text{SRT} = 30 \text{ d}$. The simulation results of the start-up operation are shown in Figures 3–5, where the system approaches steady-state after about 20 days.

3 Optimal Control

3.1 Optimization problem statement

Tight requirements have been imposed on the effluent of wastewater treatment plants by the European Union during the last decade [2]. In particular, nitrogen discharge is strictly regulated in order to reduce the environmental impact of the treated effluent and to ensure the reliability of the treatment plant. The limit value of the concentration of total nitrogen (TN) is 15 g m^{-3} for agglomerations of p.e. between 10,000 and 100,000. However, an objective of 10 g m^{-3} is often aimed at to make the plant operation more robust. Furthermore, the minimal value of TN which can be reached for the process is found to be 9.5 g m^{-3} by

means of simulation. A set point of 10 g m^{-3} is thus chosen in this study.

In this paper, we are concerned with the problem of controlling the total nitrogen (TN) around a specified set point ($\text{TN}_{\text{sp}} = 10 \text{ g m}^{-3}$) under a set of constraints. We thus consider the oxygen transfer coefficient $K_L a$ as control input $u \in \mathbb{R}^{n_u}$ ($n_u = 1$). The output is $y = \text{TN} \in \mathbb{R}^{n_y}$ ($n_y = 1$) and the state variable is $x = x_a \in \mathbb{R}^{n_x}$ ($n_x = 13$). The nonlinear model is linearized about the steady-state conditions indicated in section 2.2, and is discretized with the sampling time $T_s = 10 \text{ min}$. The linear, discrete-time model is in the form:

$$x_{k+1} = Ax_k + Bu_k \quad (4)$$

$$y_k = Cx_k. \quad (5)$$

The chosen objective function for the optimization problem is:

$$J = \frac{1}{2} \sum_{i=1}^N \left(e_i^2 + R \cdot (K_L a_i)^2 + S \cdot (\Delta K_L a_i)^2 \right) \quad (6)$$

where R and S are constant weighting factors. Here, the output error e_i is

$$e_i = \text{TN}_{\text{sp},i} - \text{TN}_i, \quad (7)$$

and the control increment $\Delta K_L a_i$ is

$$\Delta K_L a_i = K_L a_i - K_L a_{i-1}. \quad (8)$$

In addition, the control input is constrained to $K_L a \in [0, 240] \text{ d}^{-1}$.

3.2 Optimization method and results

The optimization problem described above can be posed as a quadratic programming (QP) problem of the form

$$\begin{aligned} \min_z F(z) &= \min_z \frac{1}{2} z^T H z + c^T z \quad (9) \\ \text{s.t. } \mathcal{A}z &= a \\ \mathcal{B}z &\leq b \\ z^\ell &\leq z \leq z^u. \end{aligned}$$

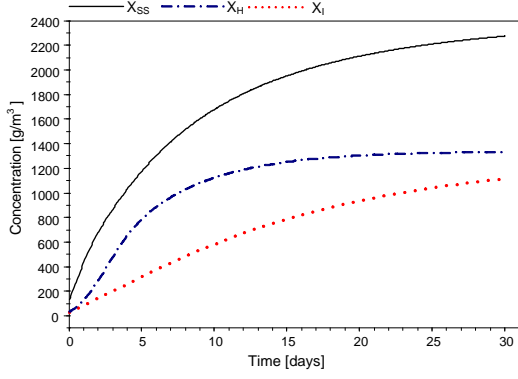


Figure 3: Simulation profiles for suspended solids (X_{SS}), heterotrophic bacteria (X_H), and inert particulate (X_I).

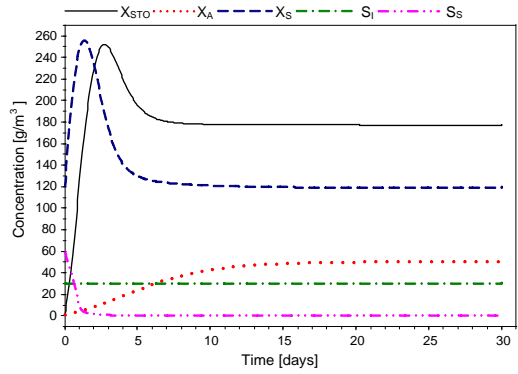


Figure 4: Simulation profiles for stored organics (X_{STO}), autotrophic bacteria (X_A), slowly biodegradables (X_S), inert soluble (S_I), and readily biodegradables (S_S).

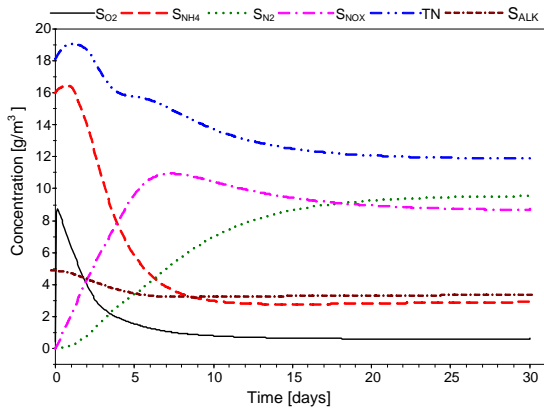


Figure 5: Simulation profiles for dissolved oxygen (S_{O_2}), ammonium (S_{NH_4}), dinitrogen (S_{N_2}), nitrite plus nitrate (S_{NOX}), alkalinity (S_{ALK}), and total nitrogen (TN).

A general introduction to QP formulations and solvers for MPC has been given in [12]. Based on their presentation, we first define the vector of unknowns z as:

$$z^T = (K_L a^T, \Delta K_L a^T, e^T, TN^T, x_a^T),$$

Table 1. Matrices for complete variable set QP-formulation.

$$\begin{aligned} \mathcal{A}_{11} &= -I_{N,-1} \otimes B \\ \mathcal{A}_{15} &= I_{N \cdot n_x} - I_{N,-1} \otimes A \\ \mathcal{A}_{25} &= -I_N \otimes C \\ \mathcal{A}_{41} &= -I_{N \cdot n_u} + I_{N,-1} \otimes I_{n_u} \\ TN_{sp} &= (TN_{sp,1} \quad \dots \quad TN_{sp,N})^T \end{aligned}$$

where

$$\begin{aligned} K_L a^T &= (K_L a_1, \dots, K_L a_N), \\ \Delta K_L a^T &= (\Delta K_L a_1, \dots, \Delta K_L a_N), \\ e^T &= (e_1, \dots, e_N), \\ TN^T &= (TN_1, \dots, TN_N), \\ x_a^T &= (x_{a,1}^T, \dots, x_{a,N}^T), \end{aligned}$$

$K_L a, \Delta K_L a \in \mathbb{R}^{N \cdot n_u \times 1}$, $e, TN \in \mathbb{R}^{N \cdot n_y \times 1}$, and $x_a \in \mathbb{R}^{N \cdot n_x \times 1}$. Matrix H and vector c of Eq. 9 are determined from the requirement that J of Eq. 6 should equal $F(z)$ in Eq. 9. The equality condition $\mathcal{A}z = a$ contains the model in Eqs. 4-5 and the definitions in Eqs. 7-8. For the optimization problem defined here, inequality $\mathcal{B}z \leq b$ is empty, while constraints are contained in z^ℓ and z^u . Thus, the following matrices result (\otimes denotes Kronecker product):

$$\begin{aligned} H &= \text{diag}(I_N \otimes R, I_N \otimes S, I_{N \cdot n_y}, 0_{N \cdot n_y}, 0_{N \cdot n_x}) \\ c &= 0_{N \cdot (2n_u + 2n_y + n_x) \times 1} \\ \mathcal{A} &= \begin{pmatrix} \mathcal{A}_{11} & 0 & 0 & 0 & \mathcal{A}_{15} \\ 0 & 0 & 0 & I_{N \cdot n_y} & \mathcal{A}_{25} \\ 0 & 0 & I_{N \cdot n_y} & I_{N \cdot n_y} & 0 \\ \mathcal{A}_{41} & I_{N \cdot n_u} & 0 & 0 & 0 \end{pmatrix} \\ a &= \left(\begin{pmatrix} A \cdot x_{a,0} + B \cdot K_L a_0 \\ 0_{(N-1) \cdot n_x \times 1} \end{pmatrix}^T, 0_{N \cdot n_y \times 1}^T, \right. \\ &\quad \left. TN_{sp}^T, \begin{pmatrix} -K_L a_0 \\ 0_{(N-1) \cdot n_u \times 1} \end{pmatrix}^T \right)^T, \end{aligned}$$

where matrices \mathcal{A}_{ij} and TN_{sp} are defined in Table 1, and $x_{a,0}$ and $K_L a_0$ are known variables. Next,

$$\begin{aligned} z^\ell &= (I_{N \cdot n_u \times 1} \cdot K_L a^\ell, -\infty \cdot I_{N \cdot (n_u + 2n_y + n_x) \times 1})^T \\ z^u &= (I_{N \cdot n_u \times 1} \cdot K_L a^u, \infty \cdot I_{N \cdot (n_u + 2n_y + n_x) \times 1})^T. \end{aligned}$$

To solve the QP problem, QP solver quadprog of the Optimization Toolbox in Matlab is used. The weighting factors R, S are tuning parameters; in this study, we set $R = 0$ and $S = 0.1$. To minimize the deviation between the predictor output and the reference, R should be given a small value or 0. Optimal control is considered over a 24 h period, i.e. the horizon is $N = 144$ when $T_s = 10$ min. The optimal aeration trajectory for the linearized model is

depicted in Figure 6. The corresponding effluent concentrations of total nitrogen TN, chemical oxygen demand COD, and dissolved oxygen S_{O_2} are shown in Figure 7. For comparison, we also include the simulation results by using the optimal input sequence K_{La} to the nonlinear Modelica model, which will indicate whether the optimal control is valid for the nonlinear process model.

The concentration of TN is controlled to track the set point. Moreover, as another requirement on the effluent of wastewater treatment plants, COD should be less than 125 g m^{-3} . This effluent criterion is satisfied in Figure 7. Hence, the optimal input sequence gives satisfactory control results for both the linearized model used in Matlab and the nonlinear Modelica model.

However, open loop operation as described above is very sensitive to model error and unknown disturbance.

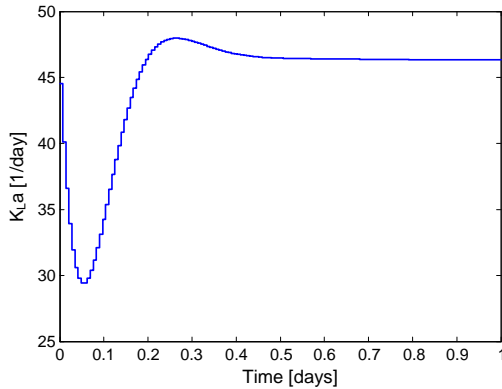


Figure 6: Evolution of K_{La} for the optimal control over 1 day, based on a linearized model.

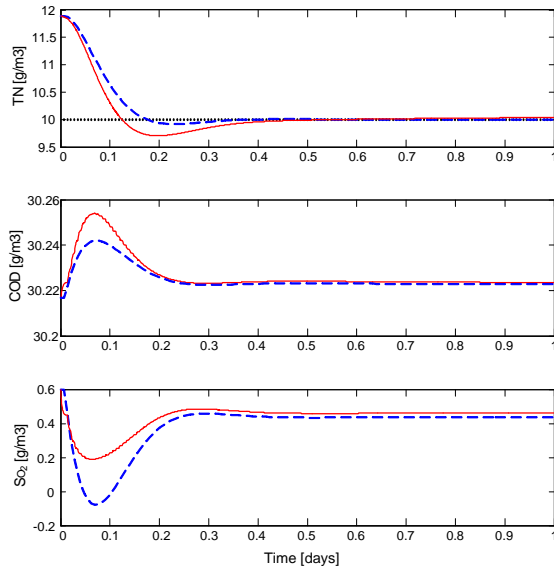


Figure 7: Comparison of TN, COD, and S_{O_2} for the QP optimal control in Matlab (dashed), with the simulation results by injecting the optimal input into the nonlinear Modelica model (solid).

4 Model Predictive Control (MPC)

Model based predictive control (MPC) changes the open loop optimal control problem of the previous section into a closed loop solution. In order to apply MPC to the model of the pilot plant, we need state estimation to get the current states from measurements. However, in this initial study, it is assumed that a perfect model of the pilot plant is available without any disturbances, and that all state variables are available. Based on these assumptions, the state estimation is not considered in this work. An approach to nonlinear moving horizon state estimation using an activated sludge model, is described in [13]. More introductions of theoretical and practical issues associated with MPC technology can also be found in [12], [14], and [15].

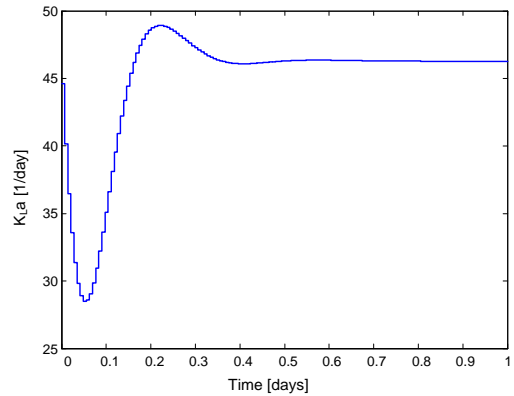


Figure 8: Evolution of K_{La} for the MPC over 1 day.

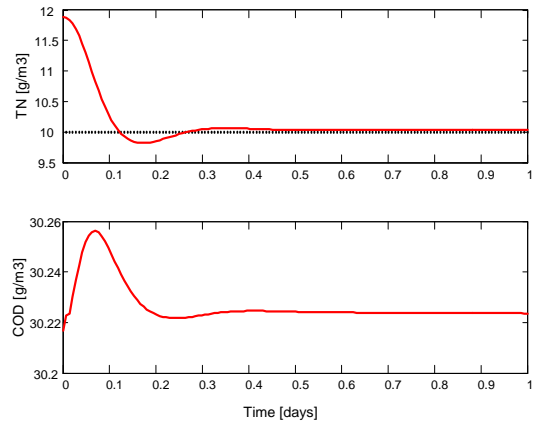


Figure 9: MPC simulation results with prediction horizon $N = 15$.

The prediction horizon N is an important tuning parameter for optimization results. Since the larger N , the closer the MPC solution approaches the full horizon optimal solution, a large N is desirable. On the other hand, N should be as small as possible because 1) the computational burden increases dramatically with N , and 2) larger N requires inclusion of more-steps-ahead predictions and hence more uncertainty. To investigate the effect of N in this application, the controlled process is simulated for N ranging from 5 to 30 sampling intervals. The simulation results show that increasing N increases the MPC accuracy

and the computational burden. Specifically, Figures 8 and 9 indicate the results with the prediction horizon $N = 15$. The MPC controller performance is satisfactory: the concentration of TN is controlled at its set point, while the concentration of COD is in conformity with the EU effluent standard.

5 Conclusion

In this paper, the use of a dynamic optimization method to design an optimal aeration strategy for small-size wastewater treatment plants is discussed. The method is applied to a simulated pilot nitrogen removal plant including simultaneous nitrification-denitrification.

A dynamic model describing the pilot plant is first developed based on the ASM3 model and is implemented in Modelica with the modified WasteWater library. The dynamic start-up simulation of the plant is pursued in order to evaluate the normal operating conditions. Then, the determination of the optimal aeration profile for the system is accomplished by both the open loop optimal control and the closed loop operation (MPC). The optimal controllers are designed to maintain the total nitrogen concentration TN at prespecified levels by acting on the oxygen transfer coefficient K_{La} . We formulate the optimization problem using the state space model of the process, and pose it as a QP problem with a complete set of variables [12]. The performances of the control strategies are evaluated by simulation studies. The simulation results show that the concentration of TN is controlled to track the set point. Thus, the controller performance is satisfactory.

In this work, the optimal control strategy has been applied to a simulation model. In the future, the strategy will be tested on the laboratory model. To get the current state based on the measurements from the real process, application of a state estimation approach (e.g. Kalman filter) in MPC is planned. Also, the convergence capabilities of the optimization method may be studied. In addition, current operating modes are given over short time horizons (24 h period). However, since the influent flow rate and compositions are likely to exhibit large variations from one day to the other, additional studies are required to analyze long term effects of optimization.

References

- [1] K. V. Gernaey, M. C. V. Loosdrecht, M. Henze, M. Lind, and S. B. Jørgensen, "Activated sludge wastewater treatment plant modelling and simulation: State of the art," *Environmental Modelling and Software*, vol. 19, pp. 763–783, 2004.
- [2] B. Chachuat, N. Roche, and M. Latifi, "Long-term optimal aeration strategies for small-size alternating activated sludge treatment plants," *Chemical Engineering and Processing*, vol. 44, pp. 593–606, 2005.
- [3] I. Metcalf & Eddy, *Wastewater Engineering. Treatment and Reuse*, 4th ed. New York: McGraw Hill, 2003.
- [4] B. Chachuat, N. Roche, and M. Latifi, "Dynamic optimisation of small size wastewater treatment plants including nitrification and denitrification processes," *Computers & Chemical Engineering*, vol. 25, pp. 585–593, 2001.
- [5] W. Gujer, M. Henze, T. Mino, and M. van Loosdrecht, "Activated Sludge Model No. 3," *Wat. Sci. Tech.*, vol. 39, no. 1, pp. 183–193, 1999.
- [6] S. Balku and R. Berber, "Dynamics of an activated sludge process with nitrification and denitrification: Start-up simulation and optimization using evolutionary algorithm," *Computers & Chemical Engineering*, 2005, in Press.
- [7] P. Fritzson, *Principles of Object-Oriented Modeling and Simulation with Modelica 2.1*. Piscataway, NJ: IEEE Press, 2004, ISBN 0-471-47163-1.
- [8] M. Henze, P. Harremoës, J. la Cour Jansen, and E. Arvin, *Wastewater Treatment. Biological and Chemical Processes*, 2nd ed. Berlin: Springer, 1996.
- [9] Dynasim, *Dymola Multi-Engineering Modeling and Simulation*. Lund, Sweden: Dynasim AB, 2004.
- [10] G. Reichl, "WasteWater a library for modelling and simulation of wastewater treatment plants in modelica," *Proceedings of the 3rd International Modelica Conference*, pp. 171–176, 2003.
- [11] Q. Chai, S. Amrani, and B. Lie, "Parameter identifiability analysis and model fitting of a biological wastewater model," *Accepted by ESCAPE-16 and PSE'2006*, 2006.
- [12] B. Lie, M. Dueñas Díez, and T. A. Hauge, "A comparison of implementation strategies for MPC," *Modeling, Identification and Control*, vol. 26, pp. 39–50, 2005.
- [13] E. Arnold and S. Dietze, "Nonlinear moving horizon state estimation of an activated sludge model," *Large Scale System: Theory and Applications. 9th IFAC/IFORS/IMACS/IFIP symposium*, pp. 554–559, 2001.
- [14] S. J. Qin and T. A. Badgwell, "A survey of industrial model predictive control technology," *Control Engineering Practice*, vol. 11, pp. 733–764, 2003.
- [15] T. Ziehn, G. Reichl, and E. Arnold, "Application of the Modelica Library WasteWater for Optimization Purpose," *Proceedings of the 4th International Modelica Conference*, pp. 351–356, 2005.

Paper E:
**Optimal Control of a Biological
Wastewater Treatment Process with
Intermittent Aeration**

This paper was presented at the conference and published in the Proceedings of the Scandinavian Conference on Simulation and Modeling 2006 (SIMS 2006) held in Helsinki, Finland, 27 – 29 September 2006, pp. 44 – 49.

OPTIMAL CONTROL OF A BIOLOGICAL WASTEWATER TREATMENT PROCESS WITH INTERMITTENT AERATION

Qian Chai ^{*}, Rune Bakke ^{*}, Bernt Lie ^{*,1}

^{} Telemark University College, P.O. Box 203, N-3901
Porsgrunn, Norway*

Abstract: In this paper, the optimal control of a biological wastewater treatment process with intermittent aeration is discussed in order to design the aeration strategy (air-on and air-off periods) which minimizes the energy dissipated by the aeration system, with adherence to the limits of the effluent requirements and the operating constraints. The dynamic model is first presented based on the Activated Sludge Model No. 3 (ASM3) and the object-oriented modeling language Modelica. Next, the optimization problem is formulated and used with a receding horizon in model predictive control (MPC). The simulation results show that significant improvements can be achieved with respect to the initial aeration strategy, with reductions in the energy consumption by 33% for the optimal control over one day and by 16% for MPC over 7 days.

Keywords: optimal control, model predictive control (MPC), Modelica

1. INTRODUCTION

Wastewater treatment is getting more and more attention in a highly populated, industrialized world. To improve water quality, it is important to remove nutrients such as nitrogen and phosphorus from wastewater. However, due to the increasingly strict requirements on treated wastewater, the operation of the process is challenging for many small wastewater treatment plants for economical and technical reasons. It is therefore of interest to apply advanced control techniques to conform to the tight effluent requirements and to save operation costs.

The activated sludge process (ASP) is the most generally applied biological wastewater treatment method. For small-sized wastewater treatment plants which are appropriate for communities of about 10,000 person equivalents, intermittent aeration ASPs are widely used. The process generally consists of a single aeration tank with al-

ternating aerobic-anoxic conditions by switching the supplied air on and off, and a settler where the microbial culture is separated from the liquid being treated, see Figure 1. Most of the culture is recycled and mixed with incoming wastewater to maintain convenient sludge age characteristics and high reaction rates. Specifically, nitrogen is removed in these two steps: ammonium is first oxidized to nitrate under nitrification/the aerobic step; the produced nitrate is then transformed into nitrogen gas under denitrification/the anoxic step. As a fact, aeration control is of great importance, since an insufficient amount of dissolved oxygen would impair the oxidation process and eventually lead to biomass death, whereas too much dissolved oxygen may cause the sludge to settle poorly. In addition, since the aeration induces 60%–80% of the global energy consumption and subsequently operating costs of a treatment plant (Chachuat et al. 2005), excessive aeration is undesirable from an economic point of view.

In this paper, optimal control of the aeration system is considered to improve the efficiency and

¹ Corresponding author: Bernt.Lie@hit.no

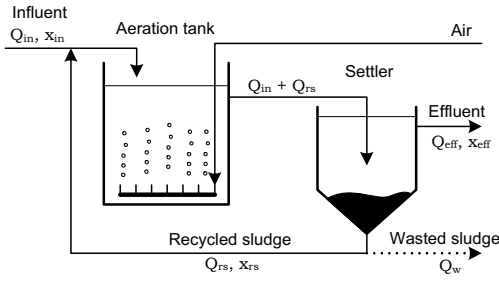


Fig. 1. Schematic diagram of a typical activated sludge process.

reliability of an ASP with intermittent aeration, used for removal of nitrogen from domestic wastewater. The objective of the control is to design an aeration strategy (air-on and air-off periods) which minimizes the energy dissipated by the aeration system, with adherence to the limits of the effluent requirements and the operating constraints. The approach used is based on dynamic optimization methods. The application of dynamic optimization methods requires a sufficiently accurate mathematical model describing the wastewater treatment process. The present work uses the Activated Sludge Model No. 3 (ASM3) (Gujer et al. 1999). ASM3 is for the time being considered the best choice in the modeling of the biological processes involved in the activated sludge tank, especially in dealing with the aerobic and anoxic processes in the same tank (Balku & Berber 2005). To give an open and flexible model representation, the general-purpose modeling language Modelica is chosen. Modelica is an object-oriented modeling language with advantages such as suitability for multi-domain modeling, the usage of general equations of physical phenomena, the reusability of model components, and a hierarchical model structure (Fritzson 2004).

2. PROCESS MODEL

In this study, a laboratory-scaled nitrogen removal plant is considered. The process consists of a unique aeration tank with a working volume $V_a = 40\text{l}$. The tank is equipped with an air blower with diffusers which provide oxygen, and a mixer making the tank behave as continuous stirred tank reactor (CSTR). The settler is a cylindrical pipe located in the aeration tank with a volume of 2.5l , where the sludge is either recycled to the aeration tank or extracted from the system. The recycled sludge is proportional to the influent flowrate, i.e. $Q_{rs} = 0.8 \times Q_{in}$. The waste sludge flowrate Q_w is used to manipulate the sludge retention time SRT², which is the most critical parameter for activated sludge design. The SRT

for the overall process is maintained at 15 d. As an initial study, the process is assumed to be operated with constant influent flow rate and compositions. Hydraulic retention time (HRT) of the process is chosen to be 10 h. Thus, the influent flow rate is $Q_{in} = V_a / \text{HRT} = 96\text{l/d}$.

The plant model consists of an aeration tank model and a settler model. The models has been discussed in detail in our previous work Chai, Bakke & Lie (2006). The ASM3 model describes the biological processes involved in the aeration tank. The states of the model are grouped into the concentration of soluble components S_j and particulate components X_j . Assuming perfect mixing in the reactor, the mass balance in the aeration tank results in:

$$\frac{dx_a}{dt} = \frac{Q_{in}x_{in} + Q_{rs}x_{rs} - (Q_{in} + Q_{rs})x_a}{V_a} + r + e_1\mathcal{A}_O \quad (1)$$

where $x_{in}, x_{rs}, x_a \in \mathbb{R}^{13}$ contain the concentrations in the influent, in the recycled sludge, and in the reactor, respectively; their components are

$$x_l = [S_{O_2,l} \ S_{I,l} \ S_{S,l} \ S_{NH_4,l} \ S_{N_2,l} \ S_{NOX,l} \ S_{ALK,l} \\ X_{I,l} \ X_{S,l} \ X_{H,l} \ X_{STO,l} \ X_{A,l} \ X_{SS,l}]^T,$$

$l \in \{in, rs, a\}$; $r \in \mathbb{R}^{13}$ is the vector formed by the reaction rates of each component (defined in ASM3); e_1 is a 13-dimensional auxiliary vector with 1 in the first element and 0 in the other elements of the vector. The mass balance equation contains the additional term \mathcal{A}_O which describes the oxygen transfer from the diffusers:

$$\mathcal{A}_O = u_a \cdot K_L a (S_O^{sat} - S_{O,a})$$

where $K_L a$ is the oxygen transfer coefficient which is a function of the supplied air flow rate, and S_O^{sat} is the saturated dissolved oxygen concentration ($S_O^{sat} \approx 10\text{g m}^{-3}$). The variable u_a is a binary sequence switching between 0 and 1 and represents the state of the blower (on/off) aerating the plant. It is assumed that the blower is on at time $t = 0$.

In this work, the settler is considered a perfect splitter, see Chai, Bakke & Lie (2006). The resulting mass balance equations are as follows:

- Effluent concentration

$$S_{j,\text{eff}} = S_{j,a}, \quad X_{j,\text{eff}} = 0 \quad (2)$$

- Recycled sludge concentration

$$S_{j,rs} = S_{j,a}, \quad X_{j,rs} = \frac{Q_{in} + Q_{rs}}{Q_{rs} + Q_w} X_{j,a}. \quad (3)$$

In summary, eliminating x_{rs} from Equation 1 using Equations 2 and 3, one can reformulate the dynamic system as a set of ODEs:

$$\frac{dx_a}{dt} = \mathcal{F}(x_a, u_a).$$

² The simplified definition of the SRT is $\text{SRT} = V_a / Q_w$.

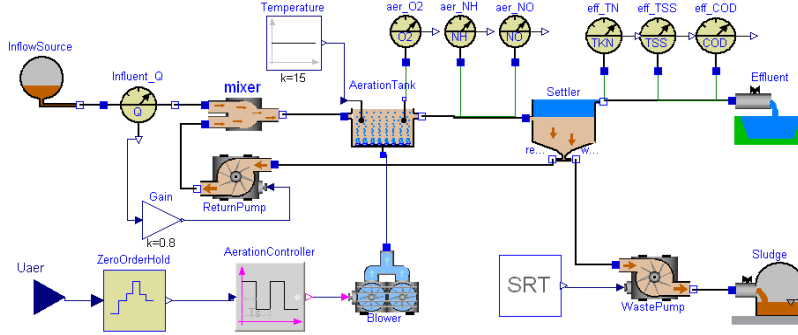


Fig. 2. Model diagram of the nitrogen removal plant implemented in Modelica/Dymola.

The dynamic model is implemented in Modelica using the Dymola simulation environment (Dynasim 2004), see Figure 2, based on modifications of the free Modelica library `WasteWater` (Reichl 2003). Dymola generates a convenient interface to Matlab such that Modelica models can be executed within Matlab. An application executing a Modelica model in Matlab is described in Chai, Amrani & Lie (2006).

Nitrogen removal is achieved by switching the blower on and off, resulting in alternating aerobic-anoxic conditions in the aeration tank. Hence, the control variable that influences the operation of the plant is the sequence of switching times, i.e. times when the blower switch on/off. In the Modelica model, the control variable is defined as an input variable U_{aer} which is read from the default Dymola file `dsu.txt` saving the trajectories of input signals; and the component `AerationController` is used to generate a boolean signal based on the sequence of switching times.

3. OPTIMAL CONTROL

3.1 Optimization problem statement

The aeration process can be seen as a succession of cycles where each cycle consists of an air-on period followed by an air-off period, i.e. the period between two consecutive starts of the blower, see Figure 3. For a given optimization horizon T_h , we assume that there are N_c aeration cycles. In the k th aeration cycle, a_k is the duration of the cycle and d_k is the length of the duty cycle, i.e. the air-on period. The aeration fraction f_k , which is defined as $f_k = d_k/a_k$, is often introduced instead of d_k . Hence, the aeration time can be optimized by manipulating a_k and f_k for $k = 1, \dots, N_c$. In order to avoid a *mixed integer programming* problem that may complicate the solution, it is supposed that the number of cycles N_c is known. For a specified N_c , in addition, the length of aeration cycles are made constant $a = T_h/N_c$.

The set of optimized parameters is then reduced from $2 \times N_c$ to N_c parameters. Also, to reduce the computational effort, the optimization is carried out over short-term horizons, e.g. 24 h.

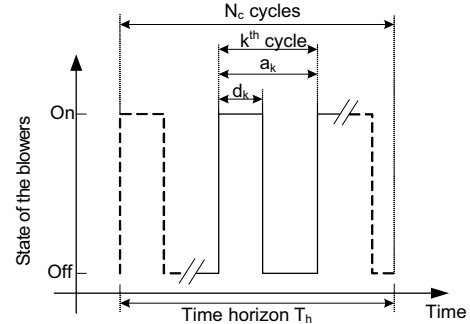


Fig. 3. Parameters of the aeration cycles.

Objective function

The objective of this study is to determine the aeration profile that minimizes the energy consumption. Since the energy consumption is mainly decided by the aeration time of the process, energy optimization is therefore achieved by minimizing the aeration time. Thus, the objective function J is defined as the total aeration time divided by the time horizon:

$$J = \frac{\sum_{k=1}^{N_c} d_k}{T_h} = \frac{\sum_{k=1}^{N_c} a \cdot f_k}{a \cdot N_c} = \frac{1}{N_c} \sum_{k=1}^{N_c} f_k.$$

J is a dimensionless variable and can be seen as the global aeration fraction over the optimization horizon: the smaller the global aeration fraction, the less the energy consumption.

Inequality constraints

Maximum residual concentrations are imposed in order to cope with the EU effluent standards on chemical oxygen demand COD, suspended solids SS, and total nitrogen TN:

$$\begin{aligned} \text{COD} &\leq \text{COD}_{\max} = 125 \text{ g m}^{-3} \\ \text{SS} &\leq \text{SS}_{\max} = 35 \text{ g m}^{-3} \\ \text{TN} &\leq \text{TN}_{\max} = 10 \text{ g m}^{-3}. \end{aligned}$$

In this work, we assume that there are no particulate components in the effluent ($SS \equiv 0$). Also, the limit of COD is generally easily satisfied as a large part of the biodegradable organic carbon is consumed during denitrification stages. Based on these considerations, only the constraint on TN is therefore considered in the optimization problem.

In practice, several constraints are defined to ensure the feasibility of the computed aeration strategies and to prevent the blower from damaging. The constraints consist of imposing minimum ($t_{on,min}$) and maximum ($t_{on,max}$) air-on periods, and minimum ($t_{off,min}$) and maximum ($t_{off,max}$) air-off periods within an aeration cycle. With reference to Chachuat et al. (2005), we define $t_{on,min} = t_{off,min} = 5$ min and $t_{on,max} = t_{off,max} = 2$ h. Hence, the boundary conditions for the aeration and non-aeration sequences are:

$$t_{on,min} \leq a \cdot f_k \leq t_{on,max} \quad (4)$$

$$t_{off,min} \leq a \cdot (1 - f_k) \leq t_{off,max}. \quad (5)$$

Adding Inequalities 4 and 5, we get the boundaries for the aeration cycle duration a :

$$t_{on,min} + t_{off,min} \leq a \leq t_{on,max} + t_{off,max}. \quad (6)$$

Then, the number of cycles N_c must be constrained to

$$\frac{T_h}{a^u} \leq N_c \leq \frac{T_h}{a^l} \quad (7)$$

where a^u and a^l are the upper and lower boundaries of a in Inequality 6, respectively.

From Inequalities 4 and 5, the boundaries for the aeration fraction f_k can be formulated using two inequalities:

$$\frac{t_{on,min}}{a} \leq f_k \leq \frac{t_{on,max}}{a}$$

$$1 - \frac{t_{off,max}}{a} \leq f_k \leq 1 - \frac{t_{off,min}}{a}.$$

Optimization problem

In summary, the dynamic optimization problem on a given time horizon T_h can be stated as:

$$\min_{f_1, \dots, f_{N_c}} J = \frac{1}{N_c} \sum_{k=1}^{N_c} f_k$$

$$\text{s.t. } \frac{dx_a}{dt} = \mathcal{F}(x_a, u_a)$$

$$0 \leq TN_{max} - TN$$

$$\frac{t_{on,min}}{a} \leq f_k \leq \frac{t_{on,max}}{a}$$

$$1 - \frac{t_{off,max}}{a} \leq f_k \leq 1 - \frac{t_{off,min}}{a}.$$

In this case, the problem is a constrained nonlinear optimization problem, which contains the nonlinear model implemented in Modelica.

3.2 Optimization method and results

The resulting optimization problem is solved by the `conSolve` solver in the TOMLAB³ optimization toolbox for Matlab. `conSolve` solves general constrained nonlinear optimization problems with the Schittkowski sequential quadratic programming (SQP) method or the Han-Powell SQP method. In this work, the Schittkowski SQP is used because of its faster convergence.

To identify the continuous operation dynamics, simulation of the start-up period of the plant is provided by starting with the aeration tank filled by the incoming wastewater. The influent wastewater composition is

$$x_{in} = [0, 30, 60, 16, 0, 0, 5, 25, 115, 30, 0, 1, 125]^T.$$

An intermittent air supply is employed with $K_L a = 72 \text{ d}^{-1}$, $N_c = 15 \text{ d}^{-1}$ (i.e. $a = 1.6$ h), and

$$f_k = 0.6 \text{ for } k = 1, \dots, N_c. \quad (8)$$

In the sequel, the aeration strategy of Equation 8 will be referred to as the *initial aeration strategy*. The system approaches steady state after about 30 days. In this work, the steady state concentrations are considered initial conditions for the optimization.

First, the optimal number of aeration cycles N_c is sought by computing optimal aeration profiles for the values of $N_c \in [6, 144] \text{ d}^{-1}$ given in Inequality 7 over a 24 h optimization horizon. The corresponding values of objective function J are shown in Figure 4. The globally optimal value of N_c is found to be 32 with the lowest value of $J = 39.92\%$. However, one of the major difficulties encountered in solving the optimization problem is the large amount of CPU time required. Since the computation time increases with the number of unknown variables involved in the optimization problem, a small N_c is desirable. In this study, N_c is chosen to be 16 with the second lowest value of $J = 40.08\%$, i.e. the length for each aeration cycle is 1.5 h. For some values of N_c , for example $N_c \in [116, 120] \text{ d}^{-1}$, the objective function increases dramatically. The reason for this is that the optimization problem has no solution when $N_c \in [116, 120] \text{ d}^{-1}$. It should also be noticed that the SQP algorithm is a local optimization method. Convergence to the global solution of the problems can not be guaranteed. Since the dynamic model is nonlinear, the problems may indeed exhibit several local minima.

The optimal aeration profile and the corresponding state of the blower are depicted in Figure 5.

³ TOMLAB is a powerful optimization tool for Matlab, offering about 100 numerically robust algorithms for linear, discrete, nonlinear, global optimization and constrained nonlinear parameter estimation. For more information, see the TOMLAB home page <http://www.tomlab.biz>.

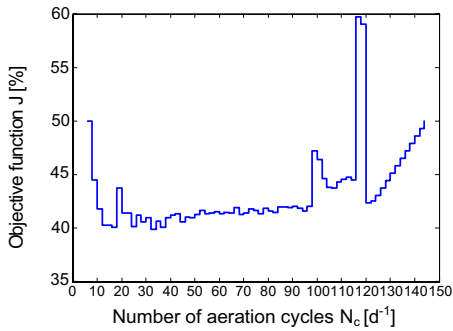


Fig. 4. Objective function J vs. number of cycles N_c .

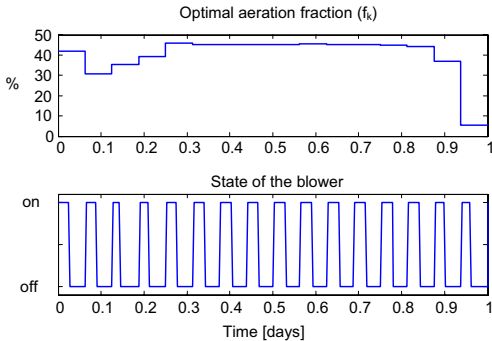


Fig. 5. Optimal aeration profile with $N_c = 16 \text{ d}^{-1}$.

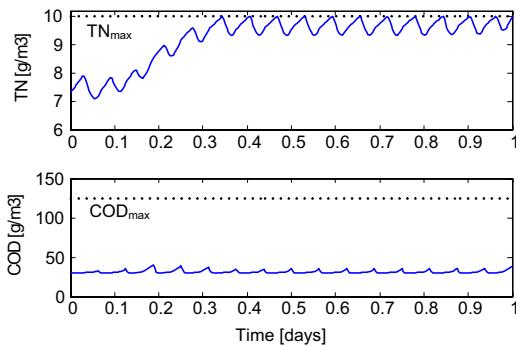


Fig. 6. Results of the optimization with $N_c = 16 \text{ d}^{-1}$: the concentrations of TN and COD.

The corresponding effluent concentrations of TN and COD are shown in Figure 6. Effluent constraint of COD always remains inactive, so this constraint has been removed from the optimization problem. The optimization results demonstrate that the global aeration fraction of 40.08%, i.e. 9 h 37 min of aeration time a day, is sufficient to ensure the effluent constraints over 24 h. Compared with the initial aeration strategy given in Equation 8, the energy consumption is reduced by $\frac{60-40.08}{60} \approx 33\%$.

One of the characteristics of wastewater treatment processes is the wide variations in the flowrate and composition of the influent wastewater. These variations are the most important source of process disturbances. However, the open-loop operation as described above is very sensitive

to model error and unknown disturbances. The ideal solution is to apply the dynamic optimization with a receding horizon in model predictive control (MPC), which changes the open-loop optimal control into a closed-loop solution.

4. MODEL PREDICTIVE CONTROL (MPC)

For the nonlinear process model, nonlinear MPC⁴ is required for optimum performance (Sunan et al. 2002). In order to apply MPC to the model of the pilot plant, we need state estimation to get the current states from measurements. A prominent problem with nonlinear MPC is that state estimation becomes more complicated when the model is nonlinear. Thus, in this initial study, it is assumed that a perfect model of the pilot plant is available without any disturbances, and that all state variables are available. Based on these assumptions, the state estimation is not considered in this work. An approach to state estimation for a large-scale wastewater treatment system is described in Tenno & Uronen (1996). A modified extended Kalman filter is also used in state estimation for a wastewater treatment model in Keesman (2002).

The dynamic optimization described in Section 3 is used in MPC with a receding horizon. The predictive horizon is chosen to be the same horizon as in optimal control, i.e. 16 aeration samples per day. The MPC simulation results are shown in Figures 7 and 8. The computation of f_k over 7 days (for 16 cycles a day) with MPC yields an average value of the aeration fraction of 50.31%. This value is larger than the one obtained over a 24 h period ($J = 40.08\%$) in optimal control. The higher aeration fraction results in that the effluent concentrations of TN is held at lower values ($8.5\text{--}9 \text{ gm}^{-3}$) than the effluent constraint 10 gm^{-3} . This seemingly disappointing result for MPC can be explained as follows: due to the periodic operation, the stationary optimal future sequence of f_k will look similar to Figure 5, and will never reach some “steady state” — there will always be some initial “transient” in f_k .

The aeration strategy for MPC provides a reduction in energy consumption by 16% compared to the initial aeration strategy in Equation 8, while more than fulfilling the strict effluent requirements at any time.

⁴ Nonlinear MPC is an optimization-based multivariable constrained control technique using a nonlinear dynamic process model for the prediction of the process outputs.

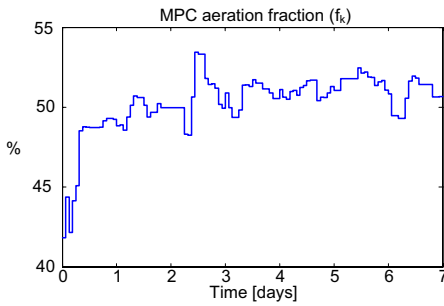


Fig. 7. Evolution of aeration fraction f_k for the MPC over 7 days.

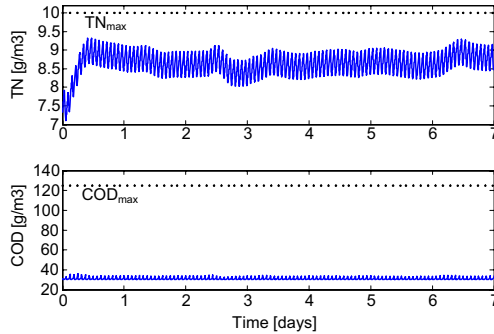


Fig. 8. MPC simulation results over 7 days.

5. CONCLUSIONS AND FUTURE WORK

An aeration control strategy for a small-sized nitrogen removal ASP is developed in this paper. A dynamic model describing the pilot plant is first developed based on the ASM3 model and is implemented in Modelica using a modified `WasteWater` library. Then, optimization over a 24 h horizon is discussed. The optimization problem is formulated as a constrained nonlinear optimization problem, which is solved by the `conSolve` solver of the TOMLAB optimization toolbox. Next, the optimization method is used with a receding horizon in MPC based on the nonlinear process model. The simulation results show that significant improvements can be achieved with respect to the initial aeration strategy (Equation 8), with reductions in the energy consumption by 33% for the optimal control over one day and by 16% for the MPC over 7 days.

In this work, an optimal control strategy has been applied to a simulation model. In the future, the strategy will be tested on the laboratory plant. The optimal strategy may also be applied to large wastewater treatment processes with more complex models. For a complex model, the computation time of optimization problem may increase, which is not expected for online implementation of the control strategy. Nevertheless, the computation time can be reduced by developing a simplified model (Antoulas 2005) or by simplifying the optimization problem. In MPC, the current state must be known. To get the current state from measurement data of the real process, application

of a state estimation approach (e.g. Kalman filter) in MPC is planned. The problem of overfulfilling the constraints with MPC will be further studied. In addition, since the influent flow rate and composition are likely to exhibit large variations, it is of interest to study the performance of the controller with model error and disturbances.

REFERENCES

- Antoulas, A. (2005), *Approximation of large-scale dynamical systems*, ISBN 0-89871-529-6, siam, Philadelphia.
- Balku, S. & Berber, R. (2005), ‘Dynamics of an activated sludge process with nitrification and denitrification: Start-up simulation and optimization using evolutionary algorithm’, *Computers & Chemical Engineering*. In Press.
- Chachuat, B., Roche, N. & Latifi, M. (2005), ‘Long-term optimal aeration strategies for small-size alternating activated sludge treatment plants’, *Chemical Engineering and Processing* **44**, 593–606.
- Chai, Q., Amrani, S. & Lie, B. (2006), Parameter identifiability analysis and model fitting of a biological wastewater model, in ‘Proceedings of 16th European Symposium on Computer Aided Process Engineering and 9th International Symposium on Process Systems Engineering’, Elsevier, Garmisch-Partenkirchen, Germany, pp. 409–414. ISSN: 1570-7946.
- Chai, Q., Bakke, R. & Lie, B. (2006), ‘Object-oriented modeling and optimal control of a biological wastewater treatment process’, *Proceedings of the Eighth IASTED International Conference on Control and Applications* pp. 218–223. ISBN: 0-88986-553-1.
- Dynasim (2004), *Dymola Multi-Engineering Modeling and Simulation*, Dynasim AB, Lund, Sweden.
- Fritzson, P. (2004), *Principles of Object-Oriented Modeling and Simulation with Modelica 2.1*, IEEE Press, Piscataway, NJ. ISBN 0-471-47163-1.
- Gujer, W., Henze, M., Mino, T. & van Loosdrecht, M. (1999), ‘Activated Sludge Model No. 3’, *Wat. Sci. Tech.* **39**(1), 183–193.
- Keesman, K. (2002), ‘State and parameter estimation in biotechnical batch reactors’, *Control Engineering Practice* **10**, 219–225.
- Reichl, G. (2003), ‘WasteWater a library for modelling and simulation of wastewater treatment plants in modelica’, *Proceedings of the 3rd International Modelica Conference* pp. 171–176.
- Sunan, H., Kiong, T. & Heng, L. (2002), *Applied predictive control*, ISBN 1-85233-338-3, Springer, London.
- Tenno, R. & Uronen, P. (1996), ‘State estimation for a large-scale wastewater treatment system’, *Automatica* **32**, 305–317.

Paper F: Predictive Control of an Intermittently Aerated Activated Sludge Process

This paper has been accepted for publishing in the proceedings of the 2008 American Control Conference held in Seattle, Washington, USA, 11 – 13 June 2008.

Predictive Control of an Intermittently Aerated Activated Sludge Process

Qian Chai and Bernt Lie

Abstract—This paper presents model-based optimal control and predictive control of a biological wastewater treatment process with intermittent aeration. The objective of the control is to design an aeration strategy which minimizes the energy consumption induced by the aeration system, with adherence to the EU effluent standards and the operating constraints. The developed optimization problem is used with a receding horizon in nonlinear MPC based on the complete ASM3 model. The MPC aeration profile guarantees that the plant fulfills the effluent requirements at any time over long time periods. Significant energy saving is also obtained when comparing MPC to three traditional rule-based control strategies.

I. INTRODUCTION

Due to legislation on wastewater treatment, especially the strict EU Guideline *Urban Wastewater Directive 91/271/EEC*, there are strong incentives to upgrade existing wastewater treatment plants in order to comply with the effluent standards and to reduce operation costs. The activated sludge process (ASP) is the most generally applied biological wastewater treatment method. In the classical biological treatment systems, aerobic nitrification and anoxic denitrification are maintained in separated zones. In the last two decades, intermittently aerated ASPs have been developed in a way through which the aerobic and anoxic zones are periodically exchanged [1], [2].

The intermittently aerated ASP generally consists of a single aeration tank with alternating aerobic-anoxic conditions by switching the supplied air on and off, and a settler where the microbial culture is separated from the liquid being treated. Most of the culture is recycled and mixed with incoming wastewater to maintain convenient sludge age characteristics and high reaction rates. The alternating aerobic-anoxic technique can easily be applied to the existing nitrogen removal plants with nitrogen removal efficiencies of 70–90% [3]. Compared to the classical biological treatment processes, the intermittently aerated ASP offers significant energy savings and easy plant retrofitting. An important feature of the intermittently aerated ASP is its flexible control ability, which makes the process better accommodate variable influent loading conditions and makes it suitable for optimization of operating costs [2]. Generally, control of the aeration system is of great importance since (i) the concentration of dissolved oxygen is directly related to nitrogen removal and (ii) the energy consumption of the aeration system is the main operating cost.

Process control of ASPs is a challenging task since the processes are characterized by large disturbances, significant nonlinearities, and stiff dynamics. To date, in many municipal treatment plants, the lengths of aerobic and anoxic phases are typically fixed or scheduled daily using a plant's supervisory control and data acquisition system. A more advanced control approach is rule-based feedback control, which uses varying phase length by establishing a switch point for each nitrification/denitrification phase. Since the on-line monitoring of ammonia and nitrate is difficult, some indirect measurements of ammonia and nitrate (e.g. oxidation-reduction potential (ORP), pH, dissolved oxygen (DO)) are commonly used to control ASPs [4]. Recently, with the development of mathematical models, particularly the Activated Sludge Models (ASMs) [5], application of optimal control has been the subject of a number of studies [6], [7], [8], [9]. Moreover, the application of MPC in the intermittently aerated ASPs has been studied in [10], [11], and [12] using simplified models, but the application of MPC with the complete ASM models is not studied much in the literature.

In this paper, optimal control of the aeration system is considered for improving the efficiency and reliability of an intermittently aerated ASP, used for removal of nitrogen from domestic wastewater. The objective of the control is to design an aeration strategy (air-on and air-off periods) which minimizes the energy dissipated by the aeration system, with adherence to the limits of the effluent requirements and the operating constraints. In Section II, the configuration and dynamic model of the nitrogen removal plant are described. In Section III, the optimization problem is formulated, and the optimization method and results are discussed. In Section IV, special emphasis is placed on using dynamic optimization in MPC with a receding horizon for a usable online implementation and to show the long-term effects of the optimal aeration strategy. In Section V, the simulation results of MPC are compared to three traditional control strategies for nitrogen removal in the intermittently aerated ASP. Finally, some conclusions are drawn in Section VI.

II. PROCESS CONFIGURATION AND MODELING

A. Process configuration

In this study, a model of a laboratory-scale nitrogen removal plant is considered. The process consists of a unique aeration tank ($V_a = 40\text{l}$) and a cylindrical settler ($V_{\text{set}} = 2.5\text{l}$). More details of the process are given in [13]. The sludge retention time (SRT) for the overall process is maintained at 15 d. The influent wastewater to the plant is

Corresponding author: Bernt.Lie@hit.no; tel: +47 35575169.
Address: Telemark University College, P.O. Box 203, N-3901 Porsgrunn, Norway.

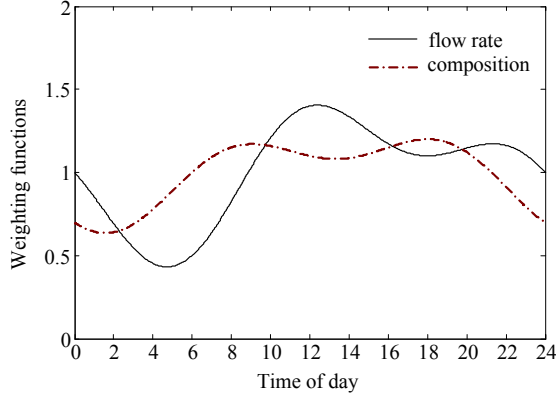


Fig. 1. Typical daily variations of inlet wastewater flow rate and composition, taken from [14].

primarily domestic wastewater. The average influent flow rate (Q_{in}) is 961/d which gives a hydraulic retention time (HRT) of 10 h, and the average total chemical oxygen demand (COD_{in}) and total nitrogen (TN_{in}) concentrations are 260 g m^{-3} and 25 g m^{-3} , respectively. Typical daily variations of dry weather conditions are simulated using weighting functions defined in [14] for both inlet wastewater flow rate and composition, see Fig. 1.

B. Process model

The plant model consists of an aeration tank model and a settler model. The complete Activated Sludge Model No. 3 (ASM3) describes the biological processes involved in the aeration tank. The states of the model are grouped into the concentration of soluble components S_j and particulate components X_j . Assuming perfect mixing in the reactor, the mass balance in the aeration tank results in:

$$\frac{dx_a}{dt} = \frac{Q_{in}x_{in} + Q_{rs}x_{rs} - (Q_{in} + Q_{rs})x_a}{V_a} + r + e_1 \mathcal{A}_{O_2}, \quad (1)$$

where x_{in} , x_{rs} , $x_a \in \mathbb{R}^{13}$ contain the concentrations in the influent, in the recycled sludge, and in the reactor, respectively; their components are

$$x_l = [S_{O_2,l} \ S_{I,l} \ S_{S,l} \ S_{NH_4,l} \ S_{N_2,l} \ S_{NOX,l} \ S_{ALK,l} \ X_{I,l} \ X_{S,l} \ X_{H,l} \ X_{STO,l} \ X_{A,l} \ X_{SS,l}]^T,$$

$l \in \{in, rs, a\}$; $r \in \mathbb{R}^{13}$ is the vector formed by the reaction rates of each component (defined in ASM3); e_1 is the standard basis for the first coordinate in \mathbb{R}^{13} ; and \mathcal{A}_{O_2} describes the oxygen transfer:

$$\mathcal{A}_{O_2} = u_a \cdot K_L a (S_{O_2}^{sat} - S_{O_2,a}), \quad (2)$$

where $K_L a$ is the oxygen transfer coefficient which is a function of the supplied air flow rate, and $S_{O_2}^{sat}$ is the saturated dissolved oxygen concentration ($S_{O_2}^{sat} \approx 10 \text{ g m}^{-3}$). Variable u_a is a binary sequence switching between 0 and 1 and represents the state of the blower (on/off) aerating the plant. It is assumed that the blower is on at time $t = 0$.

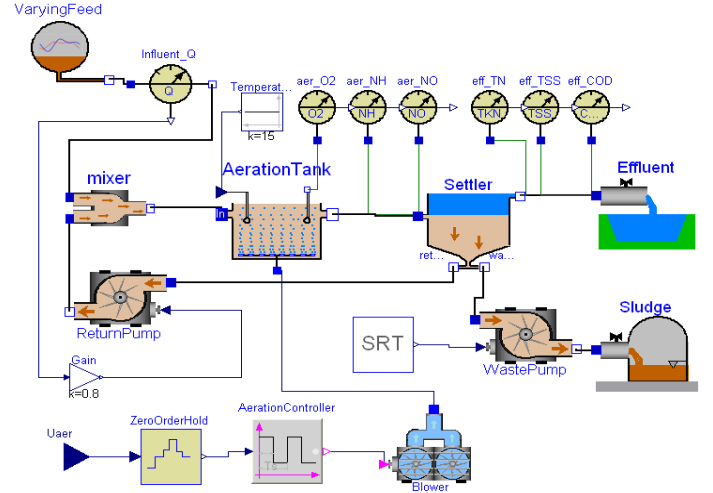


Fig. 2. Model diagram of the intermittently aerated ASP implemented in Modelica/Dymola.

Here, the control variable that influences the operation of the plant is the sequence of switching times, i.e., times when the blower switches on/off.

For simplicity, the settler is often considered a perfect splitter, thus the settler model can be expressed as follows:

Effluent concentrations are

$$S_{j,eff} = S_{j,a}, \quad X_{j,eff} = 0 \quad (3)$$

Recycled sludge concentrations are

$$S_{j,rs} = S_{j,a}, \quad X_{j,rs} = \frac{Q_{in} + Q_{rs}}{Q_{rs} + Q_w} X_{j,a}. \quad (4)$$

In summary, the general model of the dynamic system can be simply described as a set of ODEs:

$$\frac{dx_a}{dt} = \mathcal{F}(x_a, u_a, t). \quad (5)$$

The dynamic model is implemented in the object-oriented modeling language Modelica [15] using the Dymola simulation environment [16], see Fig. 2, based on the modified Modelica library WasteWater [17]. Dymola generates a convenient interface to Matlab such that Modelica models can be executed within Matlab. An application executing a Modelica model in Matlab is described in [18].

III. OPTIMIZATION PROBLEM

A. Definition of optimization problem

The aeration process can be seen as a succession of cycles where each cycle consists of an air-on period followed by an air-off period, i.e. the period between two consecutive starts of the blower, see Fig. 3. For a given optimization horizon T_h , let there be N_c aeration cycles. In the k th aeration cycle, a_k is the duration of the cycle and d_k is the length of the duty cycle, i.e. the air-on period. The aeration fraction f_k , which is defined as $f_k = d_k/a_k$, is often introduced instead of d_k . Hence, the aeration time can be optimized

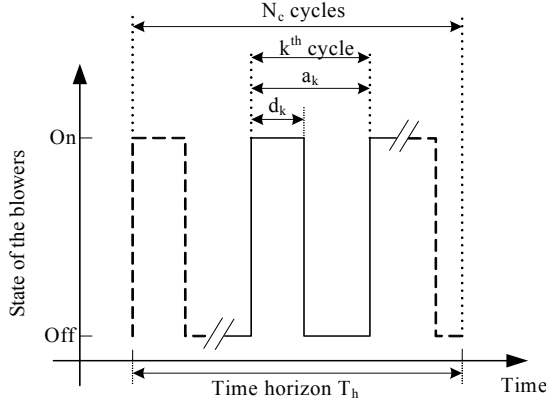


Fig. 3. Parameters of the aeration cycles, based on [19].

by manipulating a_k and f_k for $k = 1, \dots, N_c$. In order to avoid a *mixed integer programming* problem that may complicate the solution, let the number of cycles N_c be fixed. In addition, the length of the aeration cycles are made constant, $a = T_h/N_c$. The set of optimized parameters is then reduced from $2 \times N_c$ to N_c parameters. Also, the optimization horizon is chosen to be 24 h.

1) *Objective function*: The objective of this study is to determine the aeration profile that minimizes the energy consumption. The energy consumption is mainly decided by the aeration time of the process, and the extra power consumption induced by starting the aeration system is neglected. Energy optimization is therefore achieved by minimizing the aeration time. Thus, the objective function J is defined as the total aeration time divided by the total optimization time:

$$J = \frac{\sum_{k=1}^{N_c} a \cdot f_k}{T_h}. \quad (6)$$

2) *Inequality constraints*: The inequality constraints are defined in order to cope with the EU effluent standards on chemical oxygen demand COD, suspended solids SS, and total nitrogen TN:

$$\text{COD}_{\text{eff}} \leq \text{COD}_{\text{max}} = 125 \text{ g m}^{-3} \quad (7)$$

$$\text{SS}_{\text{eff}} \leq \text{SS}_{\text{max}} = 35 \text{ g m}^{-3} \quad (8)$$

$$\text{TN}_{\text{eff}} \leq \text{TN}_{\text{max}} = 10 \text{ g m}^{-3}. \quad (9)$$

In this work, we assume that there are no particulate components in the effluent ($\text{SS}_{\text{eff}} \equiv 0$). Also, the COD constraint is generally easily satisfied as a large part of the biodegradable organic carbon is consumed during denitrification stages. Based on these considerations, only the constraint on TN_{eff} is therefore considered in the optimization problem.

To ensure the feasibility of the computed aeration strategies and to prevent the blower from damage, constraints on the aeration and non-aeration sequences are introduced:

$$t_{\text{on},\text{min}} \leq a \cdot f_k \leq t_{\text{on},\text{max}} \quad (10)$$

$$t_{\text{off},\text{min}} \leq a \cdot (1 - f_k) \leq t_{\text{off},\text{max}}, \quad (11)$$

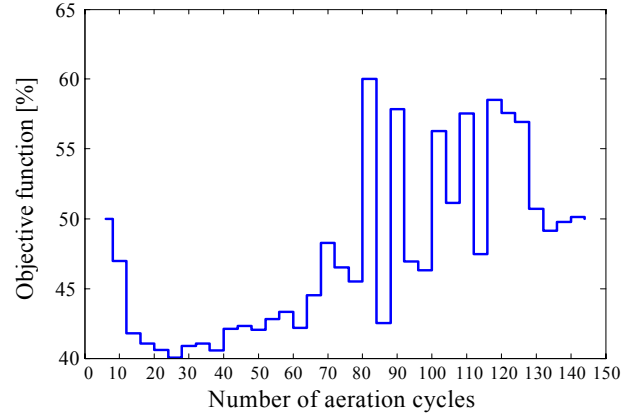


Fig. 4. Objective function J vs. number of cycles N_c .

where we define $t_{\text{on},\text{min}} = t_{\text{off},\text{min}} = 5$ min and $t_{\text{on},\text{max}} = t_{\text{off},\text{max}} = 2$ h with reference to [19].

To decide the constraints for the number of cycles N_c , we first get the constraints of the aeration cycle duration a by adding inequalities 10 and 11:

$$t_{\text{on},\text{min}} + t_{\text{off},\text{min}} \leq a \leq t_{\text{on},\text{max}} + t_{\text{off},\text{max}}. \quad (12)$$

Then, N_c must be constrained to

$$\frac{T_h}{a^u} \leq N_c \leq \frac{T_h}{a^l} \quad (13)$$

where a^u and a^l are the upper and lower boundaries of a in inequality 12, respectively.

3) *Optimization problem*: Finally, the dynamic optimization problem on a given optimization horizon T_h can be stated as:

$$\min_{f_1, \dots, f_{N_c}} J = \frac{\sum_{k=1}^{N_c} a \cdot f_k}{T_h} = \frac{1}{N_c} \sum_{k=1}^{N_c} f_k \quad (14)$$

$$\text{s.t. } \frac{dx_a}{dt} = \mathcal{F}(x_a, u_a, t)$$

$$0 \leq \text{TN}_{\text{max}} - \text{TN}_{\text{eff}}$$

$$\frac{t_{\text{on},\text{min}}}{a} \leq f_k \leq \frac{t_{\text{on},\text{max}}}{a}$$

$$1 - \frac{t_{\text{off},\text{max}}}{a} \leq f_k \leq 1 - \frac{t_{\text{off},\text{min}}}{a}.$$

B. Optimization methods and results

The resulting optimization problem is solved by the SNOPT solver in the TOMLAB¹ optimization toolbox for Matlab. SNOPT is a fast and robust solver for solving large-scale nonlinear optimization problems with a sequential quadratic programming (SQP) algorithm [20].

First, simulation of the start-up period of the plant is done using an intermittent air supply with $K_L a = 72 \text{ d}^{-1}$, $N_c = 15 \text{ d}^{-1}$ (i.e. $a = 1.6$ h), and

$$f_k = 0.6 \text{ for } k = 1, \dots, N_c. \quad (15)$$

¹For more information, see the TOMLAB home page <http://www.tomlab.biz>.

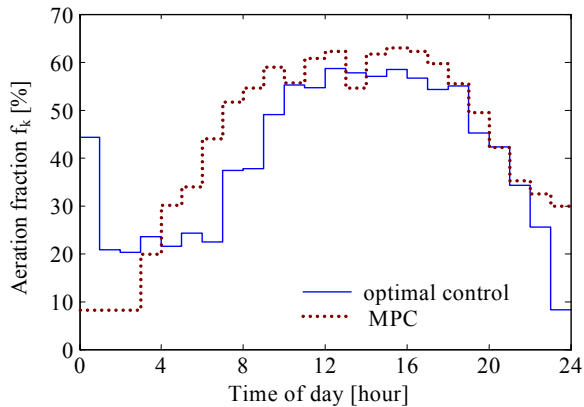


Fig. 5. Evolution of aeration fraction f_k for open-loop optimal control (solid) and for MPC (dotted) over 24 h.

The system approaches steady state after about 30 days according to simulation, see also [21]. In this work, the steady state concentrations are used as the initial conditions, and are used for all control strategies here so that the obtained results can be compared.

The optimal number of aeration cycles N_c is sought by computing optimal aeration profiles for the values of $N_c \in [6, 144] \text{ d}^{-1}$ given in inequality 13 over a 24 h optimization horizon. The corresponding values of the objective function J are given in Fig. 4. The optimal value of N_c is found to be 24 with the lowest value of $J = 40.05\%$, i.e. the length for each aeration cycle is 1 h. The values of objective function have large variations when $N_c > 60 \text{ d}^{-1}$, because the SQP algorithm may easily be captured in local minima with the increasing number of unknown optimization variables.

The optimal aeration strategy for 24 cycles a day is represented in Fig. 5 in solid line. The corresponding effluent concentrations of TN and COD are shown in Fig. 6 (solid lines). The effluent constraint on COD always remains inactive, so this constraint has been removed from the optimization problem. The optimization results demonstrate that the total aeration fraction of 40.05%, i.e. 9 h 37 min of aeration time a day, is sufficient to ensure the effluent constraints over 24 h.

IV. MODEL PREDICTIVE CONTROL

The nonlinear optimal control strategy described above is an open-loop operation which is very sensitive to model errors and unknown disturbances. For a usable online implementation, a nonlinear MPC is applied to change the open-loop optimal control into a closed-loop solution. In order to apply MPC to the model of the pilot plant, we need state estimation to get the current unmeasured states from measurements. In this simulation study, it is assumed that all states are available, therefore the state estimation is not considered here. The study of state estimation of a similar nitrogen removal process is described in [22].

The constrained nonlinear optimization defined in Section III is used in MPC with a receding horizon. The MPC

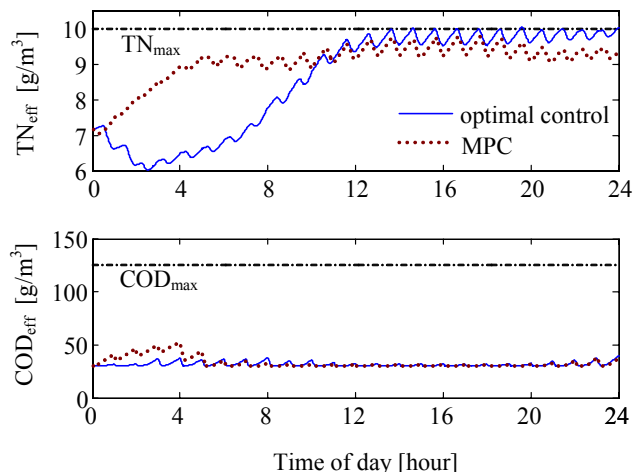


Fig. 6. Results of optimal control (solid) and MPC (dotted) over 24 hours: the effluent concentrations of TN and COD.

horizon is an important tuning parameter for optimization results. In this study, the MPC horizon is tuned to be 4 aeration cycles. The MPC aeration fraction and simulation results over 24 h are shown in Figs. 5 and 6 with dotted lines. We note that the optimal aeration strategy is closely related to the variation of the feed concentration: the high TN load (see Fig. 1) induces the high aeration fraction (e.g. between 8:00 and 20:00), whereas the low TN load results in the low aeration fraction. Compared with optimal control, the TN concentration with MPC increases faster at the beginning due to the shorter MPC horizon. The application of MPC leads to a total aeration fraction of 43.46% over a 24 h period. This value is larger than the one obtained in optimal control ($J = 40.05\%$). The higher aeration fraction results in that the effluent concentration of TN is held at lower values ($8.5\text{--}9.5 \text{ g m}^{-3}$) than the effluent constraint 10 g m^{-3} , see Figs. 6 and 7. This seemingly disappointing result for MPC can be explained as follows: due to the periodic operation, the stationary optimal future sequence of f_k will look similar to the solid line in Fig. 5, and will never reach some “steady state” — there will always be some initial “transient” in f_k .

Also, MPC is applied over a long simulation horizon to guarantee that the effluent requirements are fulfilled and the energy saving is lasting, see Fig. 7. The computation of the optimal f_k over 10 days (for 24 cycles a day) yields a total aeration fraction of 46.42% and more than fulfills the strict effluent requirements at any time.

V. COMPARISON BETWEEN MPC AND RULE-BASED CONTROL

In this section, the MPC controller developed above is compared with the traditional rule-based control strategies for intermittently aerated ASPs. The rule-based control strategies used in practice and in the literature can be categorized as (1) open-loop control using fixed lengths of aerobic and anoxic phases, or (2) feedback control using varying phase lengths by establishing a switch point for each phase. The switch point is typically the concentration of dissolved

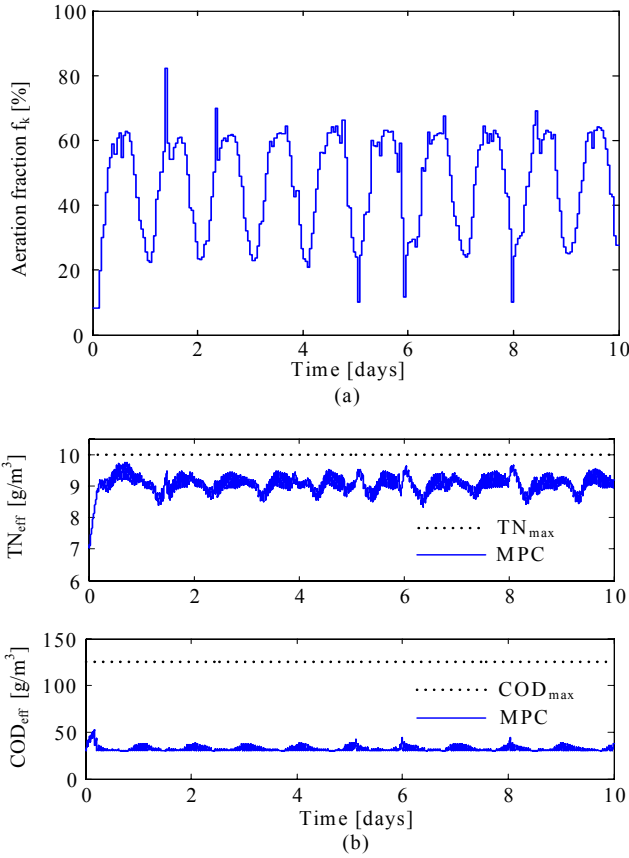


Fig. 7. MPC simulation results over 10 days: (a) MPC aeration fraction and (b) the corresponding effluent concentrations of TN and COD.

oxygen, nitrate, or ammonia [23], [24]. In this work, three rule-based control strategies are compared with MPC: fixed phase length, oxygen based control, and ammonia based control. The comparison is carried out by simulation using the sequence of aeration/non-aeration times as the control input and the TN concentration as the primary output. The objective of control strategies is to minimize the aeration time of the process while satisfying the effluent constraints. Evaluation criterion J is described in Equation 6. The applied rule-based controllers are described as follows:

- *Fixed phase length control*: In this controller, the lengths of aerobic and anoxic phases are fixed. We assume that there are N_c aeration cycles for a given optimization horizon 1 d; then the length of aeration cycles a is $\frac{1\text{d}}{N_c}$. The controller thus has 2 tuning parameters: the number of cycles N_c each day and the aeration fraction f , which are found to be $N_c = 60\text{ d}^{-1}$ and $f = 48.61\%$ by minimizing J over 1 d.
- *Oxygen (DO) based control*: In practice, a well-accepted technique is to turn the aeration system on and off based on the online measurement of DO concentration. In [19], an oxygen based controller is designed as: each aeration cycle has a constant duration of $a = \frac{1\text{d}}{N_c}$ and the aeration system is stopped when the DO concentration reaches DO_{\max} . In this study, to

TABLE I

COMPARISON BETWEEN RULE BASED CONTROL STRATEGIES AND MPC.

Controller	Criterion	Tuning parameters	CPU time for simulating 24 h
Fixed phase	48.61%	$N_c = 60\text{ d}^{-1}$ $f = 48.61\%$	39 s
DO based	53.12%	$N_c = 50\text{ d}^{-1}$ $\text{DO}_{\max} = 1.64\text{ g m}^{-3}$	20 s
S_{NH_4} based	47.59%	$S_{\text{NH}_4,\max} = 7.9\text{ g m}^{-3}$ $S_{\text{NH}_4,\min} = 7.4\text{ g m}^{-3}$	30 s
MPC	43.46%	$N_c = 24\text{ d}^{-1}$ $f_k, k = 1, \dots, N_c$	22 min

fulfill the effluent constraints, N_c and DO_{\max} are tuned to be 50 d^{-1} and 1.64 g m^{-3} , respectively.

- *Ammonia (S_{NH_4}) based control*: The aeration system can also be controlled by the upper and lower bounds of S_{NH_4} based on the measurement of S_{NH_4} (see [25]): the aeration is turned on when $S_{\text{NH}_4} \geq S_{\text{NH}_4,\max}$ and is turned off when $S_{\text{NH}_4} \leq S_{\text{NH}_4,\min}$. In fact, large values of $S_{\text{NH}_4,\max}$ and $S_{\text{NH}_4,\min}$ will result in lower aeration consumption, since ammonia is oxidized to nitrate under the aerobic step. To make the control results comparable, $S_{\text{NH}_4,\max}$ and $S_{\text{NH}_4,\min}$ are tuned to be 7.9 g m^{-3} and 7.4 g m^{-3} because S_{NH_4} varies within this region when using the MPC solution.

A comparison of MPC with the rule-based controllers is given in Table I. All computations are performed on a 1.7GHz Pentium M computer with 1Gbyte RAM. The values of criterion J show that MPC provides the shortest aeration time, whereas the DO based controller is least satisfactory. We find that MPC achieves 18.2% (i.e., 2.3h) aeration reduction each day in comparison to the DO based control. Compared to the fixed phase length and S_{NH_4} based control, which have similar J values, MPC reduces the average aeration time by up to 10.6% and 8.7%. Although the computation time of MPC is much longer than the simpler controllers, it is still feasible with on-line implementation of an MPC controller. Fig. 8 presents the TN concentrations obtained with all these control strategies. It is seen that the performances of the optimally tuned fixed phase length controller and DO based controller are similar, e.g., the aeration switching frequencies are high for both controllers, and the S_{NH_4} based controller is designed to mimic the performance of MPC. It should be noticed that the fixed phase length strategy is limited in that predetermined fixed aerobic and anoxic phases cannot compensate for unanticipated loading variations (e.g. rain/storm events). Although feedback is utilized for the DO and S_{NH_4} based control strategies, the tuning of the parameters are rather sensitive and must be performed carefully. In summary, the MPC presents the following advantages:

- 1) MPC provides the most reduction of the aeration time.
- 2) It is easier to deal with the inequality constraints in both effluent requirements and operating conditions.
- 3) MPC handles disturbances (e.g. load changes) in a more natural way.

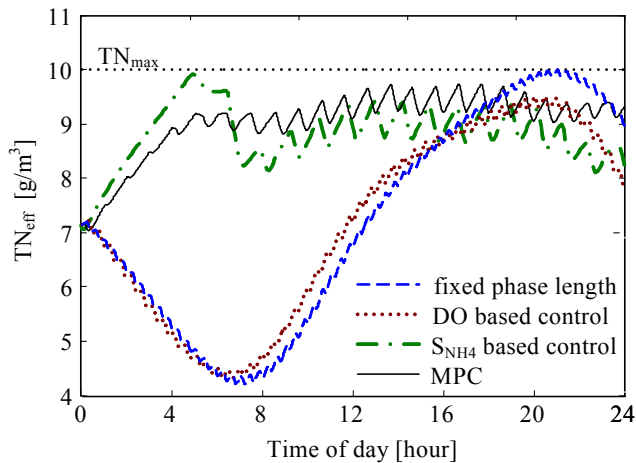


Fig. 8. Comparison of effluent TN concentrations obtained with rule based control strategies and with MPC.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, an optimal control strategy for a small-sized nitrogen removal ASP is developed to reduce the energy consumption induced by the aeration system, with adherence to the limits of the effluent standards. The optimization problem is formulated as a constrained nonlinear optimization problem, which is solved by the SNOPT solver of the TOMLAB optimization toolbox. Then, the optimization method is used with a receding horizon in nonlinear MPC based on the complete ASM3 model. The application of MPC over a long simulation horizon (10 d) is discussed, and the MPC aeration profile guarantees that the plant fulfills the tight EU effluent requirements at any time. Based on the same control objective as MPC, traditional rule-based control strategies, such as fixed phase length, DO based, and S_{NH_4} based control are discussed using well-tuned controller parameters. The comparison between the rule-based controllers and MPC shows that better aeration profiles, with reductions of energy consumption of up to 18.2%, could be obtained by applying MPC.

In the present study, the predictive control strategy has been applied to a simulation model. The plan is to test the strategy on the laboratory plant in the future. It is also of interest to study the applications of MPC to full-scale industrial plants.

REFERENCES

- [1] S. Ip, J. Bridger, and N. Mills, "Effect of alternating aerobic and anaerobic conditions on the economics of the activated sludge system," *Wat. Sci. Tech.*, vol. 19, no. Rio, pp. 911–918, 1987.
- [2] O. Hao and J. Huang, "Alternating aerobic-anoxic process for nitrogen removal: process evaluation," *Water Environ. Res.*, vol. 68, pp. 83–93, 1996.
- [3] J. Nakajima and M. Kaneko, "Practical performance of nitrogen removal in small-scale sewage treatment plants operated in intermittent aeration mode," *Wat. Sci. Tech.*, vol. 23, no. Kyoto, pp. 709–718, 1991.
- [4] M. Fuerhacker, H. Bauer, R. Ellinger, U. Sree, H. Schmid, F. Zibuschka, and H. Puxbaum, "Approach for a novel control strategy for simultaneous nitrification/denitrification in activated sludge reactors," *Wat. Res.*, vol. 34, no. 9, pp. 2499–2506, 2000.

- [5] M. Henze, W. Gujer, M. C. M. van Loosdrecht, and T. Mino, *Activated Sludge Models ASM1, ASM2, ASM2d and ASM3.*, vol. 9 of *Scientific and Technical Report*. IWA Publishing, 2000.
- [6] H. Kim, T. McAvoy, J. Anderson, and O. Hao, "Control of an alternating aerobic-anoxic activated sludge system — part 2: optimization using a linearized model," *Control Engineering Practice*, vol. 8, pp. 279–289, 2000.
- [7] H. Zhao, S. Isaacs, H. Sørensen, and M. Kümmel, "Nonlinear optimal control of an alternating activated sludge process in a pilot plant," *Journal of Process Control*, vol. 4, no. 1, pp. 33–43, 1994.
- [8] B. Chachuat, N. Roche, and M. Latifi, "Optimal aeration control of industrial alternating activated sludge plants," *Biochemical Engineering Journal*, vol. 23, pp. 277–289, 2005.
- [9] S. Balku and R. Berber, "Dynamics of an activated sludge process with nitrification and denitrification: Start-up simulation and optimization using evolutionary algorithm," *Computers and Chemical Engineering*, vol. 30, pp. 490–499, 2006.
- [10] S. Weijers, *Modelling, identification and control of activated sludge plants for nitrogen removal*. PhD thesis, Technische Universiteit Eindhoven, 2000.
- [11] H. Zhao, S. Isaacs, H. Sørensen, and M. Kümmel, "A Novel Control Strategy for Improved Nitrogen Removal in an Alternating Activated Sludge Process — Part II. Control Development," *Wat. Res.*, vol. 28, pp. 535–542, 1994.
- [12] L. Lukasse, K. Keesman, and G. van Straten, "11-norm optimal control of n-removal in an activated sludge process," *Control Engineering Practice*, vol. 7, pp. 63–70, 1999.
- [13] Q. Chai and B. Lie, "Optimal control of a biological wastewater treatment process with intermittent aeration," in *Proceedings of SIMS 2006 conference*, pp. 44–49, September 2006. Helsinki, Finland.
- [14] S. Isaacs and D. Thornberg, "A comparison between model and rule based control of a periodic activated sludge process," *Wat. Sci. Tech.*, vol. 37, no. 12, pp. 343–351, 1998.
- [15] P. Fritzson, *Principles of Object-Oriented Modeling and Simulation with Modelica 2.1*. Piscataway, NJ: IEEE Press, 2004. ISBN 0-471-47163-1.
- [16] Dynasim, *Dymola Multi-Engineering Modeling and Simulation: Dymola user manual*. Lund, Sweden: Dynasim AB, 2004.
- [17] G. Reichl, "WasteWater a library for modelling and simulation of wastewater treatment plants in modelica," *Proceedings of the 3rd International Modelica Conference*, pp. 171–176, 2003.
- [18] Q. Chai, S. Amrani, and B. Lie, "Parameter identifiability analysis and model fitting of a biological wastewater model," in *Proceedings of 16th European Symposium on Computer Aided Process Engineering and 9th International Symposium on Process Systems Engineering*, (Garmisch-Partenkirchen, Germany), pp. 409–414, Elsevier, July 2006. ISSN: 1570-7946.
- [19] B. Chachuat, N. Roche, and M. Latifi, "Long-term optimal aeration strategies for small-size alternating activated sludge treatment plants," *Chemical Engineering and Processing*, vol. 44, pp. 593–606, 2005.
- [20] P. Gill, W. Murray, and M. Saunders, *User's Guide for SNOPT 5.3: A Fortran Package for Large-scale Nonlinear Programming*. Stanford Systems Optimization Laboratory, December 1998.
- [21] Q. Chai, R. Bakke, and B. Lie, "Object-oriented modeling and optimal control of a biological wastewater treatment process," *Proceedings of the Eighth IASTED International Conference on Control and Applications*, pp. 218–223, 2006. ISBN: 0-88986-553-1.
- [22] Q. Chai, B. Furenes, and B. Lie, "Comparison of state estimation techniques, applied to a biological wastewater treatment process," in *Proceedings of 10th International IFAC Symposium on Computer Applications in Biotechnology*, (Cancun, Mexico), pp. 353–358, June 4–6 2007.
- [23] T. Potter, B. Koopman, and S. Svoronos, "Optimization of a periodic biological process for nitrogen removal from wastewater," *Water Research*, vol. 30, no. 1, pp. 142–152, 1996.
- [24] S. Isaacs and D. Thornberg, "Rule based control of a periodic activated sludge process," *Water Science & Technology*, vol. 38, no. 3, pp. 281–289, 1998.
- [25] L. Lukasse, K. Keesman, A. Klapwijk, and G. van Straten, "A comparison of NH₄/NO₃ control strategies for alternating activated sludge processes," *Wat. Sci. Tech.*, vol. 39, pp. 93–102, 1999.