AN APPROACH TO MECHANISM RECOGNITION FOR MODEL BASED ANALYSIS OF BIOLOGICAL SYSTEMS

vorgelegt von
Master of Science in Process Engineering
Mariano Nicolás Cruz Bournazou
aus Mexiko City, Mexiko

von der Fakultät III- Prozesswissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades
Doktor der Ingenieurwissenschaften
- Dr.-Ing. -

genehmigte Dissertation

Promotionsausschuss:
Vorsitzender: Prof. Dr.-Ing. G. Tsatsaronis
Gutachter: Prof. Dr.-Ing. G. Wozny
Gutachter: Prof. Dr. P. Neubauer
Gutachter: Prof. G. Lyberatos

Tag der wissenschaftlichen Aussprache 24.01.2012

Berlin 2012

D 83

dedicada a Heberto y Helig

ACKNOWLEDGEMENTS

I want to express my gratitude to my supervisor, Professor Günter Wozny, for his constant support and useful advice on a professional and personal level and to my co-supervisor, Professor Peter Neubauer, for finding a perfect application for MR and for his intellectual input to this work.

I would also like to thank Professor Kravaris and Professor Lyberatos at the University of Patras for his collaboration and hospitality.

Special thanks go to Dr. Harvey Arellano-Garcia, Dr. Stefan Junne and Dr. Tilman Barz for interesting discussions and support during the critical phases of this project (which were quite numerous).

I must of course thank all my other friends and colleagues in the Chair of Process Dynamics and Operations and in the Chair of Bioprocesses.

I would also like to thank my second family, conformed of all my friends spread around the world, who have always motivated me to follow my goals and offered a shoulder to console my sorrows.

Finally, I would like to thank my parents, Mariano and Effi, and my family for always being at my side despite the distance and especially to Alexis Cruz, who one day might realize his great contribution to each one of the achievements in my life.

M. Nicolás Cruz B.

Ich **Mariano Nicolas Cruz Bournazou** erkläre an Eides Statt, dass die vorliegende Dissertation in allen Teilen von mir selbständig angefertigt wurde und die benutzten Hilfsmittel vollständig angegeben worden sind.

Mariano Nicolas Cruz Bournazou Berlin, 1. Februar 2012

CONTENT

Zusammenfassung v
Abstract vii
Figure content ix
Table content xi
List of Abbreviations xii
List of symbols xvii
1 Introduction 1
1.1 The gap between research and industry 1
1.2 Hierarchical modeling 3
1.3 Understanding process dynamics 4
1.4 The bridge between industry and research 5
1.5 Related work 7
1.6 Project Goal 9
1.7 Advantages of Mechanism Recognition 10
1.8 The good, the bad, and the useful model 11
2 Modeling 13
2.1 Definition 13
2.2 Model complexity 14
2.3 Engineering approach to complex systems 15
2.4 Modeling in systems biology 16
2.4.1 Systems biology 16
2.4.2 Modeling of genetic regulatory systems 17
2.5 Mathematical model for a batch biochemical reactor 19
3 Model Reduction 21
3.1 Introduction 21
3.2 Basic approaches to Model Reduction 22
3.2.1 Reaction invariants 22
3.2.2 Switching functions and the reaction invariant 24
3.2.3 Sensitivity analysis 25

3.2.4 Lumping 26
3.2.5 Perturbation theory 27
3.2.6 Time scale analysis 28
4 Optimal Experimental Design 31
4.1 The experiment 33
4.1.1 The Maximum Likelihood 34
4.1.2 Model identifiability 35
4.2 The Fisher Information Matrix 37
4.2.1 The confidence Interval 37
4.2.2 Approximation of parameter variance-covariance matrix 39
4.2.3 Limitations of the Fisher Information Matrix 40
4.3 Model discrimination 42
4.3.1 Model discrimination in Mechanism Recognition 44
5 Code generation, simulation and optimization 47
5.1 Code generation 47
5.1.1 MOSAIC 47
5.1.2 SBPD 48
5.2 Simulation 49
5.2.1 sDACL 49
5.3 Optimization 50
6 An approach to Mechanism Recognition 51
6.1 A short introduction to Mechanism Recognition 51
6.1.1 Illustrative Example 53
6.2 Methodology for Mechanism Recognition 56
6.3 Program steps 57
6.3.1 Submodels 57
6.3.2 General structure 57
6.3.3 Submodel distinguishability 58
6.3.4 Initial interval 59
6.3.5 MR initialization 59
6.3.6 Detection of switching points 60
6.3.7 Initial conditions of the interval k+1 62
6.3.8 Detection of the next switching point 62

6.3.9 Flow diagram ... 63
7 Mechanism Recognition applied on Sequencing Batch Reactors ... 65
7.1 Introduction ... 65
7.1.1 Activated Sludge ... 65
7.1.2 Sequencing Batch Reactor ... 66
7.1.3 Nitrate Bypass Generation ... 67
7.1.4 Monitoring of wastewater processes ... 68
7.2 Submodel building ... 68
7.3 A proposed 9state model ... 69
7.3.1 Storage ... 69
7.3.2 Reduction of the extended ASM3 model to a 9state model ... 70
7.3.3 Mathematical representation of the 9state model ... 71
7.3.4 Stoichiometric matrix ... 73
7.3.5 Limitations of the reduced models ... 74
7.4 A proposed 6state model ... 74
7.5 A proposed 5state model ... 75
7.6 Results ... 75
7.6.1 Simulations Results ... 77
7.7 Mechanism Recognition in SBR processes ... 78
7.8 Recognition of organic matter depletion ... 79
7.8.1 Conditions for proper process description with Mechanism Recognition 79
7.8.2 Conditions for accurate switching point detection ... 80
7.8.3 MR initialization ... 82
7.8.4 Detection of switching points ... 82
7.9 Conclusions ... 83
8 Mechanism Recognition in *Escherichia coli* cultivations ... 85
8.1 *Escherichia coli* cultivations ... 85
8.2 Models for the description of *Escherichia coli* cultivations ... 86
8.2.1 Division of physiological states ... 87
8.3 Modeling *Escherichia coli* batch fermentations with Mechanism Recognition . 90
8.3.1 General model ... 90
8.3.2 Submodels for dividing metabolic states ... 92
8.4 Material and methods ... 96

8.4.1 Strain and culture conditions 96
8.4.2 Online analysis 97
8.4.3 Offline analysis 99
8.4.4 Data treatment 103
8.5 Experimental validation 104
8.5.1 Conditions for proper process description with MR 104
8.5.2 Conditions for accurate switching point detection 105
8.5.3 Data set 106
8.5.4 Recognition of overflow and substrate limitation regimes 109
8.5.5 Simulations vs. experimental data 110
8.5.6 Results 111
8.6 Conclusions 113
8.7 Future work 114
9 Conclusions and outlook 115
9.1 Conclusions 115
9.2 Outlook 116
9.2.1 General theory for submodel generation 116
9.2.2 Switching point identification 117
9.2.3 Global optimization 117
9.2.4 Online monitoring 118
10 Appendix 119

ZUSAMMENFASSUNG

Ziel dieser Arbeit ist die Entwicklung innovativer Ansätze zur Beschreibung komplexer Prozesse mit Hilfe von reduzierten Modellen. Die resultierenden Beschränkungen für die Vorhersage des Prozessverhaltens auf Basis von reduzierten Modellen werden durch den Einsatz von Methoden zur Mechanismenerkennung genutzt, um Indikatoren für relevante Änderungen im Prozessgeschehen zu erzeugen.

Empirische Kenntnisse, Analogien zu anderen Modellen aus der Literatur, Methoden zur Bewertung des Zustand eines Systems und Ansätze zur Modellreduktion werden kombiniert, in einem Versuch ein Set exakter Teilmodelle mit einer großen Robustheit und Identifizierbarkeit zu generieren. Der Ansatz zur Mechanismenerkennung ist ein Werkzeug zur effizienten Nutzung von Kenntnissen aus der Grundlagenforschung und der Modellierung und ermöglicht ein tieferes Verständnis für den gesamten Prozess.

Biologische Prozesse stellen ein wichtiges Anwendungsgebiet für die Mechanismenerkennung dar. Im Rahmen dieser Arbeit werden zwei Fallstudien vorgestellt, für die sowohl die Anwendbarkeit als auch die Vorteile dieser Methode nachgewiesen werden. Es wird gezeigt, dass die systematische Analyse des Prozesses und seiner gemessenen sowie auf Basis von Modellen vorausberechneten Zustände, die Beschreibung und Überwachung des Prozesses mit einer höheren Effizienz erlaubt.

Die erste Fallstudie beschreibt die Überwachung des Belebtschlammverfahrens in Sequencing Batch Reaktoren. Dazu wird das dem aktuellen Forschungsstand entsprechende Modell (ASM3 erweitert für die zweistufige Nitrifikation und Denitrifikation) auf ein einfaches Teilmodell reduziert. Das resultierende Modell ist effizient anzuwenden, liefert eine exakte Beschreibung des Prozesses in einem wohldefinierten Bereich und erlaubt die Erkennung des Abbaus organischer Stoffe.

Die zweite Fallstudie ist die Kultivierung von Escherichia coli im Batch-Prozess. Ein erfolgreich validiertes Modell wird analysiert und reduziert. Die Methodik der Mechanismenerkennung ermöglicht die Erzeugung von drei Teilmodellen, die in der Lage sind, Batch-Kultivierungen mit einfachen ODE-Systemen zu beschreiben.

Abschließend wird die Fähigkeit der Mechanismen Erkennung als Unterstützungswerkzeug für die Zusammenarbeit zwischen Grundlagenforschung und Industrie analysiert.

ABSTRACT

This work aims at finding new manners to accurately describe complex processes based on simple models. Furthermore, the approach to Mechanism Recognition proposes to exploit the description limitations of these submodels and to use them as indicators of non-measurable variables.

Empirical knowledge, analogies to other models from literature, methods to analyze the state of information of the system and model reduction techniques are brought together in an effort to create an adequate set of accurate models with a significantly larger tractability. It is worth stressing the approach to Mechanism Recognition does not intend to substitute human reasoning or make up for lack of process knowledge. On the contrary, this method is merely a tool to efficiently apply the knowledge obtained from basic research to gain a better insight of the industrial process.

The approach to Mechanism Recognition finds an important field of application in biological processes. In this work two case studies are presented to manifest the advantages and applicability of this method. It is shown how the correct analysis of the process, the state of information, and the models applied to describe the process results in new methods to describe and monitor the process with higher efficiency.

The first case study presented is the monitoring of the Active Sludge Process in Sequencing Batch Reactors. For this, the state of the art model ASM3 extended for two step nitrification-denitrification is reduced to create a simple model which can easily describe the process in a defined range and detect depletion of carbonate matter.

The second case study is *Escherichia coli* batch and fed-batch cultivations. A model obtained from literature is analyzed and reduced. The methodology of Mechanism Recognition allows creating a set of three submodels able to describe batch cultivations with simple systems of Ordinary Differential Equations. Furthermore, the restrictions of the complex model are set under scrutiny to understand its dynamics and limitations.

Finally, special attention is paid to the capability of Mechanism Recognition as a tool to enhance collaboration between basic research and industry.

FIGURE CONTENT

Figure 1.1: Hierarchical modeling scheme. ... 3
Figure 2.1 : *E. coli* transcriptional regulatory network. [53]. ... 14
Figure 2.2: Incremental approach for reaction kinetics identification [58] ... 16
Figure 2.3: Hypothesis-driven research in systems biology [59]. ... 17
Figure 3.1. Behavior of a switching function in dependence of the limiting species. ... 25
Figure 3.2: Three-component monomolecular reaction system, the numbers on the arrows represent the back- and forward reaction constants. ... 26
Figure 3.3: Lumping a monomolecular three-component reaction into a two-component reaction ... 27
Figure 3.4: Phase diagram of full order model (3.12). Comparison with reduced models in a chemostat process ... 30
Figure 4.1: Effect of sensitivities in parameter estimation accuracy. σ_P and σ_y represent standard deviation of parameters and measurements respectively. ... 36
Figure 4.2: Confidence interval from the Lin model, obtained with Montecarlo simulation. ... 39
Figure 4.3: Criteria for optimization [92] ... 40
Figure 4.4: Shape of the confidence interval for different variance values from the Lin model (appendix A). The confidence interval can be approximated by an ellipse near the exact value. ... 41
Figure 4.5: Objective function of a nonlinear model (appendix A) with respect to changes in a two dimensional parameter set. ... 41
Figure 5.1: High level modeling with MOSAIC [46] ... 48
Figure 5.2: Modular structure of the toolbox. The toolbox is designed in a modular ... 49
Figure 6.1: Model fit a) without setting bounds b) with setting bounds for physical parameters. [119] ... 54
Figure 6.2: Comparison experiment/simulation using a) just one model. B) various models [119] ... 55
Figure 6.3 Cleaning strategy based on MR [43] ... 55
Figure 6.4: Flow diagram of MR algorithm ... 64
Figure 7.1: SBR cycle [136] ... 66
Figure 7.2. Nitrification-denitrification process described as a two -step reaction. ... 67
Figure 7.3. Substrate concentration S_S and stored energy *Sto* against time. ... 76
Figure 7.4. Biomass against time. Changes in the biomass are very small (less than 10%). ... 76
Figure 7.5. NO_X concentration against time. ... 77
Figure 7.6. a) Oxygen concentration in the medium against time. ... 77
Figure 7.7: Description of the 5state model in both regimes, with and without substrate. ... 78

Figure 7.8: Minimal length for initialization of MR 82
Figure 7.9. Detection of the regime switching point. 83
Figure 8.1: Integration of the kinetic model proposed by Lin [91] 91
Figure 8.2: Complex model (Lin et al.) fitted to experimental batch cultivation data. 91
Figure 8.3: Comparison between the complex model (dots) vs. the overflow submodel (lines) initializing in four different intervals. 93
Figure 8.4: Comparison between the complex model (dots) vs. the substrate limiting submodel (lines) initializing in four different intervals. 94
Figure 8.5: Comparison between the complex model (dots) vs. the cell starvation submodel (lines) initializing in four different intervals. 95
Figure 8.6: Bioreactor KL2000 at *E. coli* batch cultivation [203] 97
Figure 8.7: EloCheck® 99
Figure 8.8. Calibration curve for glucose determination 100
Figure 8.9. Calibration curve of acetate 101
Figure 8.10: Mechanism of the reactions involved in the assay 102
Figure 8.11: Experimental results batch experiment G1. Part I: Dry biomass and glucose concentrations 107
Figure 8.12: Experimental results batch experiment G1. Part II: Specific concentration of acetic acid 108
Figure 8.13: Experimental results batch experiment G1. Part III: Outgas concentrations 108
Figure 8.14: Experimental results batch experiment G1. Part IV: Metabolite concentration 109
Figure 8.15: OverFlow submodel fitted against experimental data. 110
Figure 8.16: Submodel for the description of growth under substrate limitation fitted against experimental data. 111
Figure 8.17: Starvation condition described by the corresponding submodel fitted against experimental data. 111
Figure 8.18: Experimental validation of the MR approach. 112
Figure 8.19: Identifiability test considering white noise, standard deviation of 5% in all measurements 113

TABLE CONTENT

Table 4.1: Criteria for confidence interval quantification [92]. 40
Table 4.2: Types of sum of square [22] 43
Table 7.1: Reaction rates of the extended ASM3 70
Table 7.2: 9state model constants and its values as shown in the Matlab code 73
Table 7.3: Stoichiometric matrix of the 9state model 73
Table 7.4. Comparison of the computation time. 77
Table 7.5. Singular function evaluations speed 78
Table 8.1: Parameters considered for the model fit 95
Table 8.2. Composition of solution A 102

LIST OF ABBREVIATIONS

Acs	Acetyl-CoA synthase
ADHII	Alcohol Dehydrogenase
AMP	Adenosine monophosphate
AOB	Ammonium Oxidizing Bacteria
ASM	Activated Sludge Model
ASP	Active Sludge Process
BOD	Biological Oxygen Demand
Bpox	Pyruvate oxidase
CAB	Computer Aided Biology
CAPE	Computer Aided Process Engineering
CFD	Computational Fluid Dynamics
COD	Chemical Oxygen Demand
CRB	Cramer-Rao Bound
DAE	Differential Algebraic Equation
DFG	German Research Foundation
DNA	Deoxyribonucleic acid
DOT	Dissolved Oxygen Tension
EDTA	Ethylenediaminetetraacetic acid
EMA	European Medicines Agency
FDA	Food and Drug Administration
FIM	Fisher Information Matrix
GRN	Gene Regulatory Network

HET	Heteroptrophic organisms
HPLC	High-Performance Liquid Chromatography
IA	Incremental Approach
IMM	Interactive Multiple Model
KDD	Knowledge Discovery of Data
LSQ	Least Squares
MBR	Membrane Bioreactor
MBDoE	Model Based Design of Experiments
MD	Model Discrimination
MR	Mechanism Recognition
mRNA	Messenger Ribonucleic Acid
MTT	Thiazolyl Blue
MWF	Multi-Wavelength Fluorescence
MXL	Maximum Likelihood
NAD^+	Nicotinamide adenine dinucleotide (NadH)
NB	Nitrobacter
NBND	Nitrate Bypass Nitrification-Denitrification
NDF	Numerical Differentiation Formula
NS	Nitrosomona
NH^+_4	Ammonia
NIRS	Near-Infrared Spectroscopy
NO^-_2	Nitrite
NO^-_3	Nitrate
NOB	Nitrite Oxidizing Bacteria
NSF	Numerical Differentiation Formula

OC	Orthogonal Collocation
OCFE	Orthogonal Collocation on Finite Elements
ODE	Ordinary Differential Equations
OED	Optimal Experimental Design
OF	OverFlow Metabolism Model
PAT	Process Analytical Technology
PCA	Principal Component Analysis
PCP	Process Constant Parameter
PDE	Partial Differential Equation
PES	Phenazine Ethosulfate
PLS	Partial Least Squares
ppG	Phosphoenol Pyruvate Glyoxylate
ppGpp	Guanosine tetraphosphate
PSO	Particle Swarm Optimization
PSSH	Pseudo Steady State Hypothesis
PTS	phosphotranspherase
QSSA	Quasi Steady State Assumption
RWP	Regime-Wise constant Parameter
SBML	Systems Biology Markup Language
SBR	Sequencing Batch Reactors
SF	Switching Function
SL	Substrate Limitation Model
SQP	Sequential Quadratic Problem
ST	Starvation model
TCA	Tricyclic Acid Cycle

WWTP	Waste Water Treatment Plants

LIST OF SYMBOLS

VARIABLES

A	area	$[m^2]$
Ac	acetate	$\left[\frac{mMol}{L}\right]$
A_{crit}	A-criterion	[ ]
a	linearly independent row vector	[ ]
α	specific cake resistance	$\left[\frac{1}{m^2}\right]$
C	covariance matrix	[ ]
C	concentration	$\left[\frac{mMol}{L}\right], \left[\frac{MolC}{g}\right], \left[\frac{g}{m^3}\right]$
c	substrate consumption coefficient	[ ]
CTR	carbon dioxide transfer rate	$\left[\frac{Mol}{s}\right]$
D	dilution rate	[ ]
D^B	Identifiability threshold	[ ]
Δp	pressure difference	$[bar]$
E	enzyme	[ ]
ε	Distinguishability threshold	[ ]
F	feed rate	$\left[\frac{L}{s}\right], \left[\frac{mL}{h}\right]$
f	Function	[ ]
Φ	objective function	[ ]
g	gravity acceleration	$\left[\frac{m}{s^2}\right]$

Γ	initial velocity of the projectile	$\left[\frac{\mathrm{m}}{\mathrm{s}}\right]$
H	hypothesis	[ ]
η	stochastic error	[%]
η_s	systematic error	[%]
K	limiting constant	$\left[\frac{mg}{L}\right]$
k	monomolecular rate matrix	[ ]
κ	friction constant	$\left[m * \frac{kg}{s}\right]$
L	membrane thickness	$[mm]$
m	mass	$[kg]$
μ	growth rate	$[h^{-1}], [d^{-1}]$
ν	dynamic viscosity	$\left[\frac{N * s}{m^2}\right]$
OTR	oxygen transfer rate	$\left[\frac{Mol}{s}\right]$
P	product	[ ]
$\wp$	probability distribution function	[ ]
Q	uptake	$\left[\frac{g}{L}\right]$
q	specific uptake	$\left[\frac{g}{g * L}\right]$
R	resistance	$[m^{-1}]$
r	reaction rate	[ ]
res	residual	[ ]
S	concentrations of soluble species (substrates and products)/Substrate	[ ]

St	correction constant	[ ]
s	blocked area per unit filtrate volume	$\left[\frac{m^2}{m^3}\right]$
σ	standard deviation	[ ]
t	time	$[s\ ], [min], [h], [d]$
t_{sp}	time span	$[s\ ], [min], [h], [d]$
θ	parameter vector	[ ]
u	input variables vector	[ ]
V	Volume	$[L]$
W	weighting matrix	[ ]
w	constant input variables vector	[ ]
W_m	Culture medium weight	[ ]
X	concentrations of the particulate compounds	[ ]
x	state variables vector	[ ]
Y	yield coefficient	$\left[\frac{g}{g}\right]$
Y	stoichiometric coefficient	[ ]
y	measurement values vector	[ ]
z	reaction invariant	[ ]

SUBSCRIPTS AND SUPERSCRIPTS

0	initial value
α	incoming
aer	aeration phase
anox	anoxic phase
Bio	biomass
C	cake
$calc$	calculated value
cap	capacity
E	experimental
es	estimated
G	general structure
Gluc	glucose
H	heterotrophous
L	lower
M	membrane
max	Maximal
mes	Measured value
nom	nominal
O	Oxygen
S	Substrate

1 INTRODUCTION

1.1 THE GAP BETWEEN RESEARCH AND INDUSTRY

Globalization has changed market conditions drastically. Advances in transport and communication bring companies together in worldwide competition. Cutting edge technology is now essential for chemical and biochemical companies to survive. To achieve this, substantial efforts have to be invested in research and development, not only for direct applications, but also as long term investments to earn basic knowledge.

Industry is forced to make such investments to strive for its success in the world markets, setting new standards in product performance. In the year 2010, BASF invested almost 1.5 billion Euros in research and development [1].

Governments also need to make important investments on research, promoting mostly basic research, which is not attractive to industry because it represents a long term investment. The German Research Foundation (DFG) invested in the same year 2010 approximately 2.3 billion Euros [2], including support to universities, long term projects, and specific research fields.

In spite of the parallel effort of both parties aiming at a common goal, collaboration projects between academia and industry confront many complications. While industry demands mostly fast solution to real process problems, academia is more interested in long term projects offering novel knowledge. It can be said that industry is in search of smart solutions while academy is looking for interesting problems. Finding novel methods to bring industry and the research community together is essential for their efficient development. Basic research offers a strong platform for development of industrial applications, and industry provides not only economic support but also new challenges and interesting applications.

Process modeling in chemistry and biotechnology offers a handful of examples of the advantages of joint work. The development of a complex model, including estimation and validation, may take several years. In addition, model identifiability or observability, and application range cannot be assured beforehand. A company cannot afford to make such long term and uncertain investments. These models have to be developed in basic research. Still, accurate models allow optimal design and operation of plants, reducing energy consumption, hazard, and environmental impact, while allowing better monitoring and control [3]. Today, many of the models and software tools developed in universities and research institutes are used in industry (Aspen®, Gproms®, Matlab®).

In return, industry offers, in addition to economical support, the required facilities for parameter estimation and model validation. The data collected daily in chemical plants provides valuable information to researchers. Additionally, information about large scale processes and long term performance can only be obtained from real plants.

Development of new tools that facilitate the communication and interaction between industry and basic research lead to more efficient collaboration and better individual performance. Instruments to benefit from the advances achieved in basic research by allowing an adequate information transfer between both parties are crucial for an efficient development of modern process technology. Modeling is not an exception. New methods need to be created to bring complex models closer to industry and also to create ways to use the information earned in industry for basic research purposes.

As maximization of process efficiency becomes essential to remaining competitive in the market, complex models, which enable profit increase while fulfilling environmental and safety regulations, are gaining application in industry. Process complexity and safety restrictions have driven design and control to demand accurate and robust models. Current black-box models and heuristic rules cannot provide the information required in modern engineering. Regulations are changing, demanding model-based knowledge of the process. The new regulations of the Process Analytical Technology (PAT) initiative of the Food and Drug Administration (FDA) and the European Medicines Agency (EMA) show the importance that modeling applied to process monitoring and control is gaining in the pharmaceutical and generally in the biotechnological industry. Due to the difficult measurements required, the application of model based control and monitoring is essential.

The FDA makes the following statement in its Guidance for Industry, January 2011 [4]:

"A successful validation program depends upon information and knowledge from product and process development. This knowledge and understanding is the basis for establishing an approach to control of the manufacturing process that results in products with the desired quality attributes. Manufacturers should:

- *Understand the sources of variation*
- *Detect the presence and degree of variation*
- *Understand the impact of variation on the process and ultimately on product attributes*
- *Control the variation in a manner commensurate with the risk it represents to the process and product "*

1.2 HIERARCHICAL MODELING

The contradiction between models in research and industry can also be seen from the point of view of hierarchical modeling. Figure 1.1 depicts the typical layer representation of a chemical process. These three different layers have a diverse level of significance for industry and research, whereas industry is more interested in plant wide behavior aiming at robust and secure process operation, basic research is more interested in the lower layer where the study of microscalar phenomena takes place.

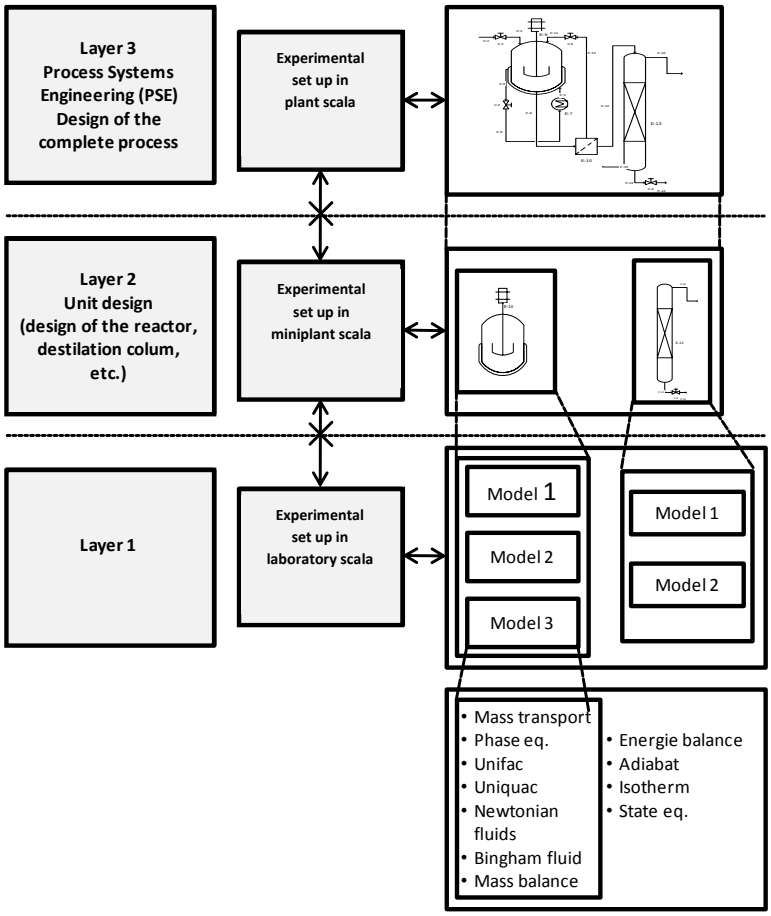

Figure 1.1: Hierarchical modeling scheme.

Particularly in biological systems, a gap can be seen between industry and basic research [5]. Biological systems are extremely complex and very difficult to predict. Depending on the level of system understanding, cells can be described with a simple Boolean equation, from a kinetic down to a genomic level. In addition, regulations in food and pharmaceutical industry are extremely strict. For this reason, industry is only interested in practical, simple, and robust models. On the other hand, the main goal for building a model in research is to gain process information [6]. This second category of models is commonly too complex and requires advanced, expensive and time demanding

measurement techniques. Finally, application of such complex models requires highly trained personnel. Still, industry needs to take advantage of knowledge gained in basic research in general and of application of complex models in particular.

A defined methodology to strategically simplify complex models, considering both the requirements of a particular industrial process and the quality of the data available, is missing in biotechnology. Although many reduction methods are applied for control process purposes [7], a general approach for model reduction considering online and offline measure possibilities, experimental conditions, and deep understanding of the system is not to be found in literature.

1.3 UNDERSTANDING PROCESS DYNAMICS

Mathematical models can be described as the result of an effort to represent behavior of nature with mathematical equations. Despite the inability of mathematics to precisely describe physical phenomena, the approximate description achieved by models has shown to be very useful. In the words of P. G. Box [8]: *"all models are bad, but some are useful"*. Models are applied in all fields of science and have become an essential tool for data acquisition and processing, understanding of complex systems, and prediction of their behavior. In process engineering, models are used for process design, monitoring, control, and optimization.

Modeling and simulation have developed rapidly over the last years [9]. Advanced measurement techniques and fast computer processors enable the creation of very complex models processing enormous amounts of information [10]. Nevertheless, sophisticated models contain an important number of parameters and thus require large amounts of very specific data in order to be identifiable. In most cases, experimental effort for parameter estimation increases exponentially as the model grows in complexity. Not only the measurement techniques become more complicated and expensive but also the identifiability of the parameters is reduced with each new parameter added to the model [11]. Online measurement limitations may also hinder the application of complex models for model-based control. In addition, complex models require costly hardware to make such complicated calculations as well as expensive software to simulate and optimize the model efficiently. Furthermore, with an increasing number of parameters to optimize, initial value consistency gains importance for simulation convergence. Speaking of parameter estimation, a large number of parameters increases the size of the optimization problem and number of local minima [12]. Finally, all the complications mentioned above restrict application to highly trained personnel.

Batch processes commonly show highly nonlinear behavior and require more advanced models for their description. Complications related to batch process simulation and control are well known [13-15]. These dynamic and highly nonlinear processes require accurate first principle models to be properly described. Nonetheless, in rigorous modeling the choice of the mechanistic model to be used for the simulation is based on the dominant physical phenomenon of the process. These phenomena, which dictate the process dynamics, change over time. Hence, the appropriate approach is to simulate the process with various models also changing over time. In other words, models should change based on how and when these phenomena change. This is the principal reason why most dynamic processes can be simulated effectively for short time periods but not for the complete process. Nonetheless, in many cases only certain conditions of the process are of interest. Simplifying the model to adapt it to the strictly important conditions may reduce the complexity drastically. Unfortunately, this cannot be foreseen and the model can only be adjusted once experimental data is available.

1.4 THE BRIDGE BETWEEN INDUSTRY AND RESEARCH

When speaking of industrial systems, there are many processes that operate without detailed model-based knowledge of its dynamics. In the past, predictions were carried out mainly on the basis of empirical knowledge. Experience and over-sizing combined with improvements during operation led to fairly successful results. However, in recent years an increasing trend to bring existing plants to meet new market demands can be established. These demands include, for example, improved quality or compliance with new standards for environmental restrictions. Unfortunately, simple nonlinear regressions based on direct measurements are not suitable for these goals.

On the other hand, complex models present a number of disadvantages which hinder their implementation in industrial processes. Low identifiability, complex measurement techniques, large calculation costs and the need for highly trained staff are only some of the problems to be faced in order to apply complex first principle models to industrial processes. It is well known that mechanistic models offer a number of advantages over "black-box" modeling, e.g. a higher process comprehension and a more accurate scale-up capability [16-21]. Also rigorous models provide the basis needed for efficient quality control. If correctly implemented, mechanistic models help to predict risks, environmental impact and improve design and operation through simulation and optimization. Still, rigorous models developed in basic research are rarely applied in industry. Models have to be tractable, observable, robust, and simple but also accurate and reliable for its use in industrial applications. In order to build models that have all the aforementioned features and are also based on rigorous knowledge of the system, a close cooperation between the research community and industry is essential.

This work represents an important step towards the development of a systematic approach to the adaptation of complex models for their application in industrial processes. Model reduction is a promising approach to close the gap between models developed in basic research and models required in industry.

One of the most difficult decisions to make for a modeler is the level of description accuracy required for a model to be useful [22]. As we will see later in detail, an agreement must be met between model accuracy and modeling, parameter fitting and simulation efforts. Deciding how accurate a model needs to be to accept it as an adequate description of the process is still an open question in engineering.

The difference between the outputs predicted by the model and the outputs measured from the system is called residual [23]. Considering the exact parameters are known, the causes for residual different from zero can be grouped in two main categories [24]:

- uncertainties (stochastic error)
  Disturbances and unknowns are intrinsic errors of the system and cannot be predicted. These show a normalized distribution with norm equal to zero and a variance dependent on the conditions of the system, measuring methods and further unknown factors.
- model structure (systematic, error)
  When the structure of the model is incorrect, meaning it fails to consider all important factors of the process and to represent the correct dynamics, there exists no parameter set, which can make the model fit the data.

Modelers usually tend to build models with too many parameters and to settle with locally optimal parameter values. This trend is slowly changing with the development of efficient global optimization techniques [25]. Global dynamic optimizers offer the possibility to find the definite parameter set which best describes the observations [26]. The most significant contribution of global optimization to model structure analysis is that one can rigorously demonstrate that the model is inconsistent with experimental data regardless of its parameter values. Nonetheless, methods to detect the source of systematic error are required and approaches to detect the instance of the structure causing the error require further development.

Despite many efforts to develop automatic modeling programs [27, 28], the selection of the structure of a model still requires individual analysis of each case and vast experience in modeling added to deep knowledge of the system to be modeled. This is partly overcome by adding new equations and parameters to patch errors in the structure of the model. Nonetheless, these "patches" are usually responsible for unneeded parameter correlation and reduction of model identifiability. To name one example, a straight line can be described exactly by a fifth order polynomial, but such a model will never be

identifiable because there is an infinite combination of values of the polynomial, which can describe a straight line. It is unidentifiable because an infinite combination of parameter sets exists, which fit the system.

Creating new tools to analyze the structure of models and find correct representations of the system is the main goal of this work. To achieve this goal, many disciplines need to be brought together in an effort to attack model defects from different angles to detect failures and to propose solutions. Finding communication paths between the different disciplines to take advantage of the information gained in each case and achieve the best possible model for each system is essential. Furthermore, as will be shown in this manuscript, a combination of simple models may offer important advantages.

1.5 RELATED WORK

Especially in process engineering, the use of models to obtain precise process information based on indirect measurements has been utilized since the beginnings of the discipline. There exists a handful of methods aiming at fast and robust description of processes. Various fields in science require fast calculations to achieve optimal control of systems with high dynamics. From missile tracking to burnout reactions, many approaches have been successfully applied mostly using statistical methods and repeated linear approximations of the system. Furthermore, the use of a combination of more than one model in an effort to describe specific instances of a system or complete processes has been proposed in various forms. Qualitative process theory [29, 30], Interactive Multiple Model (IMM) [31, 32], jump Markov linear systems [33], qualitative algebra and graph theory methods [34], semiquantative simulation [35], variable structure theory [36], are just some examples. However, these methods relay on simple models with no physical foundation with fast, but short term prediction being its ultimate objective.

As limitation by computation burden losses significance due to the increasing capacity of modern microchip architecture and cloud computing systems, the application of large nonlinear models is gaining popularity. Approaches to reject hypothetical reaction pathways in chemistry using first principle models in combination with global optimization have been published [37]. Also online applications like model based fault isolation and identification consider the application of rigorous models to detect malfunctions in the system [38, 39]. These methods use software redundancy with mechanistic models in an effort to detect fault behavior in complex systems. Furthermore, fault detection techniques have rapidly evolved [23] and are being applied in many fields of industry, e.g. PUMon (a tool for online monitoring based on neural networks) is being developed at Bayer [40]. Nonetheless, despite the long story of

similar methods to gain knowledge from limited data sets [41], its application in complex dynamic systems is still limited.

Furthermore, a systematic methodology for the identification of non measurable process variables, using a comparison between different first principle models describing selected regimes in dynamic processes, is not to be found in literature [42].

Mechanism Recognition (MR) differs from all previous approaches in that the physical properties of the system are considered. Most methods for system description with more than one model aim strictly at computation expenses reduction, leaving system understanding aside. On the contrary, MR is concerned with the characteristics of the submodel and its relation to the physical system. Furthermore, MR aims at discerning and selecting the phenomena dictating the dynamics of the system.

MR has been successfully applied for small systems [43]. Still, the first application is limited to models with one state variable. In this example, different models were obtained from literature each one describing a different regime of the process (section 6.1.1). Because of the simplicity of the models applied, no general structure was required and input-output consistency was inherently fulfilled by the single input single output condition of all models. The results obtained suggest that the method can be also applied for systems with a higher number of state variables. Nevertheless, when obtaining models from literature, a continuous computation of all state variables cannot be assured (input-output consistency). Since the models are obtained from different sources, the number and types of state variables contained by each model may differ. Hence, it is not possible to assure calculation of all state variables in every regime. Furthermore, complex models require a general structure to increase its identifiability this cannot be generated for models with different characteristics.

The core of MR is model building, most precisely, submodel building. Once physical meaning of each submodel has been experimentally validated, and its interaction with all other submodels has been understood, the application of MR is straight forward and some of the aforementioned techniques can be applied. Still, the practice of modeling should not be underestimated. Novel software toolboxes for model building modularization and reusability [44, 45] together with efficient integrators [46] facilitate the exercise of modeling significantly. Furthermore, a number of software packages for automatic model building [47] and automatic model reduction [45] confirm the trend to a general, systematic, and automated modeling approach. Nevertheless, modeling is still a field which requires intensive human intervention. The engineer must make use of his knowhow and intuition to be able to develop efficient models which mirror reality and are consistent with scientific evidence. This work provides significant evidence that despite technological advances, modeling is still a challenging and exciting discipline [48]. The challenges of modeling and experimental validation will be discussed, different manners to create and analyze models (chapter 3) and its relation with the observations

of the system (chapter 4) will be presented in an effort to increase the efficiency of model development.

1.6 PROJECT GOAL

The main goal of this project is to find new approaches for a target-oriented model simplification. By these means, complex models created in basic research can be adapted for application in industrial processes. Various methods for model reduction are to be studied in combination with mathematical tools for experimental information quantification (confidence intervals, optimality criteria, etc.) to fulfill specific requirements of particular industrial problem.

Secondly, this works aims at finding new means to accurately describe complex processes based on simple models. In order for a simple model to mirror a complex system, three essential conditions must be fulfilled:

- deep comprehension of the dynamics of the system
    - The complete system, but more important, the phenomenon governing systems behavior must be deeply understood.
- minimal systematic error
    - Equations and structure of the model must describe only the most important dynamics, with the minimal number of parameters possible and minimal systematic (e.g. modelization) error.
- high model identifiability
    - The data set must deliver enough information to estimate the parameter set with high accuracy. It is essential to understand that identifiability depends not only on the data set (state information), but also on the structure of the model.

Now let us assume that a specific variable or process parameter cannot be measured due to physical limitations. Let us also assume that we have created a model, which satisfies the above mentioned conditions. This means that it is able to describe the strictly defined regime of the system with high accuracy. This very special characteristic is exploited by MR. If it is precisely known which regime can be described by the model, a process running outside this regime can be easily detected.

MR provides insight into the system, allowing a deeper understanding of process dynamics and process monitoring to operate in optimal conditions. The biggest challenge for the application of MR is how to create a simple but accurate model specifically adapted to the particular conditions of each regime. This is also the main topic throughout this manuscript.

The validity of the approaches proposed will be tested in two case studies of high relevance in the field of water treatment and recombinant cultivations.

Finally, it is worth recalling that physical understanding of the system, either chemical or biological, is the keystone to this approach. MR does not intend to substitute human reasoning or make up for lack of process knowledge. On the contrary, MR is merely a tool to efficiently apply this knowledge in order to gain a better insight of the system under study.

1.7 ADVANTAGES OF MECHANISM RECOGNITION

The information contained by the complex model has to be used with intelligence to fit the process needs while increasing the identifiability and the observability of the submodels. Furthermore, a reduced model comprehends much more information than the same model built using the classical top down approach (from black-box to grey-box to first principle models). The most important advantages when creating a submodel through an intelligent reduction of a complex model are:

- Specified adaptations for each process:
    - A defined model reduction can be carried out for a specific process. By these means the model is adapted to each particular case. Again, because of the mathematical basis, the information gained can be exported to systems and used for different conditions.
- Phenomenon identification:
    - The model reduction can also be conducted to determine a selected phenomenon of the process. This allows the identification of non measurable variables and increases the information obtained by the experiments.
- Knowledge about the accurate experiment is gained through model reduction:
    - The creation of reduced models and their parameter estimation delivers important information to be implemented in the complex model. For example, the nonlinear interrelation of the states in the complex model can be understood better if the behavior of its reduced models is analyzed.

1.8 THE GOOD, THE BAD, AND THE USEFUL MODEL

It is common to evaluate models as "good" or "bad" and these terms are also used in this work following convention. Still, it is essential to be aware that all the approaches to model evaluations might fail. Although it is true that some special characteristics of a model must be analyzed before using it, experience has shown that it is very difficult to predict the functionality of a model. Particularly in engineering, the most important question to answer is whether or not certain model characteristics can be exploited aiming at specific goals. In many cases, the simplest model has shown to perform much better than complex, nonlinear ones. Reasons for this are explored in this work.

Engineering, being a practice and industry oriented discipline, is mainly interested in usefulness of models. For an engineer the principal aspect to take into account is if a model can bring some advantages in process efficiency or not. For a model to be useful, it is necessary and sufficient that it be robust, reliable, and descriptive.

A model that robustly describes the simplest part of a system properly is far better than a complex model that mirrors the complete process but has a high probability of failure.

2 MODELING

2.1 DEFINITION

This works considers mathematical models exclusively and their application in the description of physical phenomena. For sake of generality, we limit our concept of mathematical model to the definition made by Aris [49]:

"A mathematical model is a representation, in mathematical terms, of certain aspects of a nonmathematical system. The arts and crafts of mathematical modeling are exhibited in the construction of models that not only are consistent in themselves and mirror the behavior of their prototype, but also serve some exterior purpose."

Furthermore, the study of this work is limited to mechanistic models expressed in the form of Differential Algebraic Equation (DAE) systems applied exclusively for description of process engineering in chemistry and biotechnology (2.1). Finally we delimit to controlled physical, chemical, and biological systems.

$$f(\dot{x}(t), x(t), u(t), w, \theta, t) = 0 \qquad \textbf{2.1}$$

where $\dot{x}(t)$ is a vector with the derivatives of the state variables, $x(t)$ is a vector with n_s time-dependent variables which define the system, $u(t)$ a vector of n_u time-dependent input variables, w is a vector with n_w constant input variables, θ is a vector with P parameters, and t represents time.

The initial conditions are also to be defined.

$$f(\dot{x}(t_0), x(t_0), u(t_0), w, \theta, t_0) = 0 \qquad \textbf{2.2}$$

where t_0 is the time at point 0.

Contrary to black box models, mechanistic models are based on physical knowledge of the system to be described. In engineering for example, rigorous modeling includes mass and energy balances, detailed reaction pathways, etc. Models are the core of Computer Aided Process Engineering (CAPE) [50] and Computer Aided Biology (CAB). The quality of every work on simulation, optimization, design, and model based control, depends on the characteristic of the model. Models are evaluated by its simplicity, accuracy, robustness, generality, and computation burden. It is worth reminding, that there is no such thing as the "best" model for all applications. The

"best" model can only be selected after the objective of the simulation and the state of information (chapter 4) has been specified.

In engineering, models are not only used to describe the behavior of systems, they are also essential to map complex systems into smaller dimension more comprehensible to humans. Finally, they also serve to obtain indirect measurements and observe non observable events. This last category of models is also known as software sensors [51]. Software sensors substitute measurements, which are not possible due to physical limitations, with models which predict the behavior of the non measurable variable based on indirect measurements.

2.2 MODEL COMPLEXITY

A common mistake is to consider the most complex model to be the most appropriate for description of a system. In most of the cases it has shown to be quite the opposite. Experience shows that the fewer the parameters in a model, the better [52]. Still, the first solution that comes to mind when a model fails to describe a system is to add new parameters. Instead, this should be considered the last resource and should be done only after all other options have been exhausted.

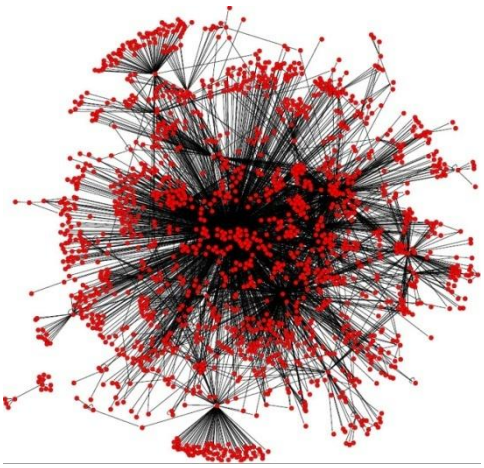

Figure 2.1 : *E. coli* transcriptional regulatory network. [53].

Model complexity is closely related to instability, over parameterization, parameter correlation, and low parameter identifiability. The effort required to develop and fit a model has to be justified by its application. It is useless to apply Computational Fluid Dynamics (CFD) to the simulation of a 1L reactor knowing that the concentration gradients can be neglected. On the other hand, simulating a reaction in a tank with 10,000 L without considering mass transfer limitations may yield catastrophic results. Summarizing, the key dynamics of a system need to be identified, isolated and analyzed before any model is built. Currently, the three conditions (section 1.6) are limited mainly

due to the scarcity of measurement possibilities but also due to the insufficiency of adequate mathematical tools. It is at this point that the MR approach can contribute to modern model building.

A model with hundreds of parameters including exponential, hyperbolic, and discontinuous functions might seem advanced and sophisticated, but this illusion quickly vanishes when the model has to be validated and used for design or optimization. Much better is a correct approximation, than an accurate misconception. The real challenge for modeling is to develop a general and systematic approach to find the simplest manner to describe complex systems aiming at the strictly required accuracy. The meaning of model simplification becomes more important everyday with the increasing complexity of processes analyzed in research Figure 2.1.

2.3 ENGINEERING APPROACH TO COMPLEX SYSTEMS

In chemical engineering, the implementation of different methods to deal with large complex systems has a long history. Engineers have developed methods like hierarchical modeling, model reusability, model inheritance, etc. An extensive discussion of these methods and their application for the simulation of chemical plants is presented by Barton [3]. In biological systems, the modularization of separated instances of the system is not always possible. In traditional process engineering, a pump can be modeled in a modular form and then added to the flow sheet of the plant and reused as many times as needed [54]. Contrary to this, biological systems tend to show different behavior under in vitro conditions compared to their in vivo state [55]. Still, some approaches intend an analysis and modeling of biological systems with methods taken from engineering [56, 57].

An alternative method to create optimal model structures has been published by Bardow [58]. This method called the Incremental Approach (IA) suggests building the model in an inductive manner. In a sense, IA could be considered a hierarchical approach extended to an even lower layer to first principle phenomena. Although its application finds important limitations, e.g. quality of data required and bias, the general concept behind IA is worth our attention. In principle, IA extends the philosophy of hierarchical modeling to the molecular level.

Inverse problem theory is the most common approach for model building and specifically models fit to data. First, the differential model is evaluated (integrated) with a certain parameter set, and then the data is compared against the output previously computed. The residual between model outputs and data is calculated and a new set of parameters is tested. These steps are followed iteratively, usually solving some least square type of optimization problem (section 4.1.1) until the residual is considered to be minimal. An important disadvantage of this approach is that it is not possible to directly

analyze the internal structure of the model. Although, various methods exist to indirectly investigate parameter sensitivities, correlations, and bifurcation among others, a true insight in the structure of the model is still not possible.

Bardow [58] proposes finding the parameter value needed to fit each new data point. Estimating new parameter values for each data assures that the differential equation presents the correct derivative. This process would be similar to fitting one parameter for each measured point independently. The result we obtain is a curve showing the ideal parameter values. This curve, although very noisy in most of the cases and without any physical meaning, is very helpful when building a new set of equations. The modeler can visualize the behavior of the parameters and decide if they can be represented by constants, or algebraic or differential functions.

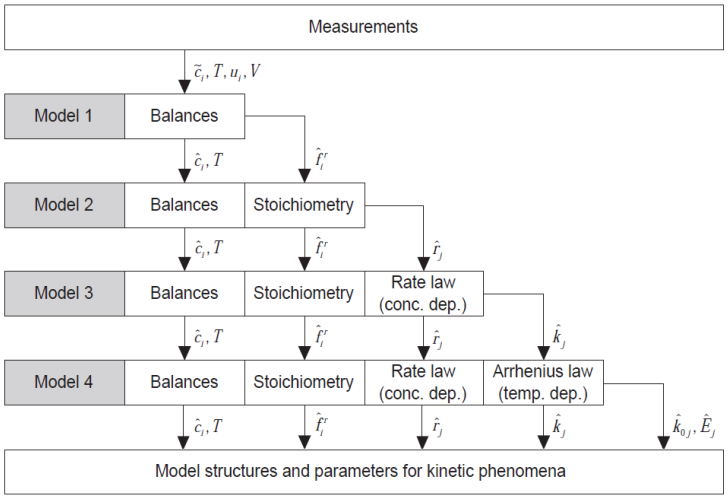

Figure 2.2: Incremental approach for reaction kinetics identification [58]

IA proposes to build the model in a deductive way. A drawback of this approach is that process information is required for each step in the model building process. Still, this method can be very useful if advanced measurement techniques are available. Besides, IA offers a very well described systematic procedure for model building, which is usually underestimated in process engineering and biotechnology Figure 2.2.

2.4 MODELING IN SYSTEMS BIOLOGY

2.4.1 SYSTEMS BIOLOGY

We refer to the definition by Kitano [6, 59]: *"Systems biology aims at understanding biological systems at system level"*. Systems biology emphasizes the fact that the only possible manner to understand living organisms is to consider the system as a whole.

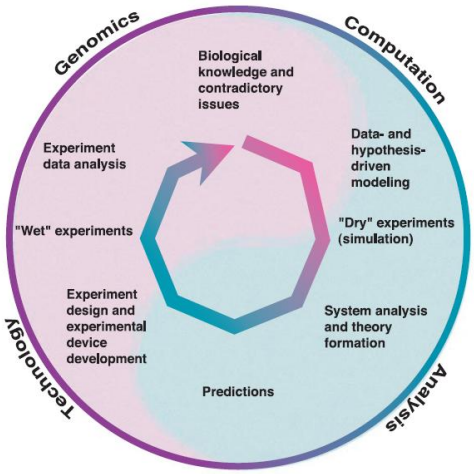

Figure 2.3: Hypothesis-driven research in systems biology [59].

Identifying genes and proteins is only the first step, whereas real understanding can only be achieved by uncovering the structure and dynamics of the system. Kitano states four key properties:

- System structure
  - System structure identification refers to understanding both, the topological relationship of the network components as well as the parameters for each relation.
- System dynamics
  - System behavior analysis suggests the application of standardized techniques such as sensitivity, stability, stiffness, bifurcation, etc.
- The control method
  - System control is concerned with establishing methods to control the state of biological systems.
- The design method
  - System design is the effort to establish new technologies to design biological systems aimed at specific goals, e.g. organ cloning techniques.

The relevance of modeling in systems biology is clearly stated in Figure 2.3.

2.4.2 MODELING OF GENETIC REGULATORY SYSTEMS

System biology has triggered an impressive contest between various methods aimed at an adequate description of the dynamics of living organisms studying its Gene Regulatory Network (GRN), the most representative being [60]:

- Directed and undirected graphs
- Bayesian networks

- Boolean networks
- Generalized logical networks
- Linear and nonlinear differential equations
- Piecewise linear differential equations
- Qualitative differential equations
- Partial differential equations
- Stochastic master equations

Each approach offers different advantages and no definitive method can be defined as the "best" by the systems biology community. The Assessment of Network Inference Methods attempt to analyze all pros and cons of the different GRN inference methods. The goal is to compare the different approaches against equal data sets to obtain quantifiable information of the difference in performance between the methods [61]. Because complete understanding of the system is essential for a proper evaluation, the most promising results have been obtained with simulated data sets, but much work is to be done before an adequate comparison can be achieved.

As stated before, differential equation systems settle the standard modeling method in engineering. For this reason, the most interesting model approaches for MR are the ones based on differential equations. In fact, linear and piecewise linear differential approaches are perfectly suitable for model reduction.

Systems of Ordinary Differential Equations (ODE) have been widely applied for the description of GRN. Usually the system comprises rate equations of the form

$$\frac{dx_i}{dt} = f_i(x, u) \qquad \textbf{2.3}$$

where x can be the vector of concentration of proteins, mRNAs, or other molecules, u the vector of inputs, and f_i is a nonlinear function. Also time delays can be added if necessary. Typical types of equations used are, Monod type, switching, Heaviside, and logoid functions among others. An important advantage of nonlinear ODEs is the possibility to describe multiple steady states and oscillations in the system [62]. Besides the requirement of testing the global convergence of the optimal solution, the bottle neck is still the state information of the parameter set creating identifiability problems. Nevertheless, some successful applications have been published showing the possibilities of ODEs to describe GRN [63].

It is worth recalling that MR aims at simple model building and GRN modeling is far from this. Still, both GRN modeling and model analysis and reduction techniques have shown exponential development in the last years. Therefore, it can be expected, that systematic conversion of complex GRN models in simple submodels suitable for MR will be possible in near future. Someday, detailed descriptions of complete GRN will be

the basis for perfectly defined submodels applied in industry to make fast, robust and accurate predictions of complex processes.

2.5 MATHEMATICAL MODEL FOR A BATCH BIOCHEMICAL REACTOR

MR finds its most important application in dynamic systems. A process in constant change presents different behaviors and governing phenomena also change over time. It is at this stage where the different process conditions can be selected and the submodels can be built. Biochemical batch reactions have been selected to validate MR and its application for the description of industrial processes. For this reason, a short discussion of the general form of the mathematical model is presented.

The biochemical reactions involve consumption of various chemical species (substrates) and production (intermediate or final metabolic products) and biomass growth. Products from a microbial group are often the reactants of other microbial groups. This results in a sequence of individual process steps, which is part of a scheme, where some steps may be independent of those that follow [64].

Assuming that biochemical reactions, generally described through

$$S_i \xrightarrow{r_k} Y_{j,k} * X_j + \sum_{\substack{l=1 \\ l \neq i}}^{l} C_{l,k} * S_l \qquad \textbf{2.4}$$

take place in a batch biochemical reactor, the following differential equations can be derived .

$$\dot{S}_i = \sum_{k=1}^{n} C_{i,k} * r_k(S_1, \cdots, S_l, X_1, \cdots, X_m), \quad i = 1, \dots, l$$
$$\dot{X}_i = \sum_{k=1}^{n} Y_{j,k} * r_k(S_1, \cdots, S_l, X_1, \cdots, X_m), \quad j = 1, \dots, m \qquad \textbf{2.5}$$

where:

$S_i, i = 1, \dots, l$ are the concentrations of the chemical species (substrates and/or products) in the reactor, $X_j, j = 1, \dots, m$ are the concentrations of the microbial masses in the reactor, $r_k(S_1, \cdots, S_l, X_1, \cdots, X_m), k = 1, \dots, n$ are the reaction rates, $C_{i,k}$ and

$Y_{j,k}$ are the stoichiometric coefficients for substrate consumption and microbial growth, respectively.

It should be noted that the consumption of a substrate (e.g. particulate matter) may not be associated with biomass growth. Moreover, a single microbial group may grow on more than one substrate and vice versa. Therefore, in the general case, the number of the substrates involved in a bioreaction scheme will not be equal to the number of microbial masses grown, i.e. $l \neq m$.

Introducing vector notation for the concentrations and the rates

$$\underline{S} = \begin{bmatrix} S_1 \\ \vdots \\ S_i \end{bmatrix}, \underline{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_m \end{bmatrix}, \underline{r}(\underline{S}, \underline{X}) = \begin{bmatrix} r_1(S_1, \cdots, S_l, X_1, \cdots, X_m) \\ \vdots \\ r_n(S_1, \cdots, S_l, X_1, \cdots, X_m) \end{bmatrix}$$

and denoting by $\underline{\underline{C}}$ and $\underline{\underline{Y}}$ the lxn and mxn matrices of the stoichiometric coefficients, model 2.5 takes a more compact form:

$$\begin{aligned} \dot{\underline{S}} &= \underline{\underline{C}} \cdot \underline{r}(\underline{S}, \underline{X}) \\ \dot{\underline{X}} &= \underline{\underline{Y}} \cdot \underline{r}(\underline{S}, \underline{X}) \end{aligned} \qquad \textbf{2.6}$$

3 MODEL REDUCTION

3.1 INTRODUCTION

A model is a poor mathematical representation of a physical system. Lack of accurate knowledge of the process to be modeled, insufficient measurement techniques and extensive computation time hinder an exact representation of the phenomena to be described [65]. Nevertheless, models are widely used in science and their contribution to a better understanding of engineering processes and their proper design, optimization and control is unquestionable. From this it can be deduced that the best model to describe a certain process is not necessarily the most accurate, but the one that describes only the relevant aspects of the system so as to get a good description with minimal effort [66]. Different methods have been developed to detect the key dynamics in order to create an accurate but relatively simple model.

The process can be described as a bottom up approach in the hierarchical modeling sense. Once a detailed model has been built, model reduction leads to model simplification. Because of the information gained from the detailed model, the reduction follows mathematical and physical principles. By these means species are neglected and dynamics are simplified based on their influence on the overall system.

Model reduction is keystone in engineering, a widely applied approach for reduction of nonlinear models is the linearization based on Taylor series, which has proven to be very useful for processes in steady state conditions. Unfortunately, dynamic nonlinear systems require more complex approaches. Model reduction aims at distinguishing the important from the negligible modes in an effort to reduce the model to a more attractable form maintaining its key dynamics [67]. Some of the most important advantages of reducing a model are:

- increased identifiability/observability.
- increment of model robustness
- reduction of model stiffness
- reduction of computation expenses

As a result, not only the experimental effort for parameter estimation is drastically reduced, but, most important, the measurement effort during process monitoring and control is minimized. It may be even possible to convert non observable models into models adequate for model-based process control. A very important difference between steady state and dynamic processes is the selection of the "important" dynamics.

In model reduction for continuous processes based on time scale analysis, the fast modes are neglected. Only the slow modes, which determine the path while reaching the equilibrium point are maintained [68]. On the contrary, when a system is far from equilibrium (as is usually the case in dynamic processes) the fast modes are of major importance and the very slow modes can be considered constant (quasi-steady state). In addition, these new constants may give rise to further reaction invariants, which should be considered because of their model reduction potential.

Literature dealing with model reduction under steady state conditions has been widely published. However, less work has been done in model reduction of dynamic processes. On the one side, most industrial chemical processes run at steady state (constant operating conditions). On the other hand, the condition of stability enables many assumptions which simplify the reduction problem significantly, linearization and Lyapunov stability being just two examples. In biotechnology, most processes are batch or fed-batch processes which allow no steady state assumptions. New approaches to model reduction need to be developed to permit reduction of models for dynamic processes. A representative example of the reduction potential of dynamic models is shown in section 7.3.

3.2 BASIC APPROACHES TO MODEL REDUCTION

The reduction of a particular model maintaining its characteristics is a common task in all fields of engineering. Representative examples are Lumping [69, 70], Sensitivity Analysis [71] and Time-Scale Analysis [64, 72, 73]. In this chapter, a short introduction to some of the methods used for model reduction is presented. Besides many efforts, model reduction techniques still relay on process knowledge and experience to achieve the correct reduction of the model. In most of the cases, a combination of methods is required to achieve the simplest form possible.

3.2.1 REACTION INVARIANTS

In chemical and biochemical models, the method of reaction invariants is applied to detect the modes which stay constant in the process. Firstly, the method finds linear dependencies in the model. Secondly, reaction invariants help detect the invariant manifold which in turn is a valuable tool to find the slow manifold of a system of dynamic equations.

In the general model of equations 2.6, there are $(l + m)$ differential equations that are affected by n reaction rates. As long as $l + m > n$ and the differential equations are independent of each other, there will be $l + m - n$ linear combinations of the

concentrations that are completely unaffected by the reaction rates and therefore completely unaffected by the progress of the chemical reactions. In the literature, these are referred to as reaction invariants; they capture the reaction stoichiometry relations, which are not affected by the reaction rates.

The reaction invariants can be easily calculated from the general form 2.6 .

Assuming $l + m > n$ and

$$Rank\begin{bmatrix} \underline{\underline{C}} \\ \underline{\underline{Y}} \end{bmatrix} = n$$

,

one can find $(l + m - n)$ linearly independent row vectors $a_v, v = 1, \cdots, (l + m - n)$ of length $(l + m)$ such that

$$\alpha_v \begin{bmatrix} \underline{\underline{C}} \\ \underline{\underline{Y}} \end{bmatrix} = 0 \qquad v = 1, \cdots, (l + m - n)$$

This means that the $(l + m - n)$x$(l + m)$ matrix

$$\underline{\underline{A}} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_v \end{bmatrix}$$

has rank $(l + m - n)$ and satisfies

$$\underline{\underline{A}} \begin{bmatrix} \underline{\underline{C}} \\ \underline{\underline{Y}} \end{bmatrix} = 0 \qquad \textbf{3.1}$$

It can then be easily verified, as a result of (2) and (3), that the quantity

$$\underline{z} = \underline{\underline{A}} \begin{bmatrix} \underline{S} \\ \underline{X} \end{bmatrix} \qquad \textbf{3.2}$$

remains constant throughout the entire batch:

$$\dot{\underline{z}} = 0 \qquad \textbf{3.3}$$

and so

$$\underline{\underline{A}} \begin{bmatrix} \underline{S}(t) \\ \underline{X}(t) \end{bmatrix} = \underline{\underline{A}} \begin{bmatrix} \underline{S}(0) \\ \underline{X}(0) \end{bmatrix} \quad \forall t > 0 \qquad \textbf{3.4}$$

3.2.2 SWITCHING FUNCTIONS AND THE REACTION INVARIANT

Switching Functions (SF) are widely applied in biological systems. Its most common form corresponds to the simplified Michaelis-Menten equation 3.5.

$$r = r_m \frac{C_S}{K_S + C_S} \qquad \textbf{3.5}$$

where r, $\mathrm{r_m}$, $\mathrm{K_S}$, C_S are the reaction rate, the maximum reaction rate, the inverse of enzyme affinity and the substrate concentration, respectively.

SF enable activation or deactivation of different reactions depending on the concentration of species. The objective is to create a continuous and smooth function, which is equal to 1, when the concentration of the limiting component is high, and equal to 0 when the limiting component has been depleted. The inhibition functions show the opposite behavior, the equation is equal to 0, in case of high concentrations, and to 1, when the species is not present. Referring to equation 3.5, ϑ will have the value ϑ_m when the concentration of substrate is high and 0 when the concentration of substrate is near 0. We characterize the switching function behavior in three phases Figure 3.1.

- active and constant (species concentration is significantly higher than the limiting constant > 100)
- transition phase (when the ratio species/constant is between 100 and $1*10^{-2}$)
- inactive and constant (species concentration is significantly lower than the limiting constant $<1*10^{-2}$)

This particular characteristic of the SF opens interesting possibilities for model reduction. The presence of SF in a mathematical model may engender temporal reaction invariants during specific time intervals in a process. To be more precise, the process presents different reaction invariants under limitation of different species.

To visualize this idea, considering the simplified Michaelis-Menten equation in 3.5, it can be seen that if $[S] > 100 \; \forall \; t > 0$, then $r \approx r_m = \text{constant}$. We can therefore create reduced versions of the original model, which behave exactly as the original model under specific conditions. This can be very useful when applying general mathematical models to defined conditions. When a model describes a wide range of process conditions all species limitations, which may or may not occur during the process, have to be considered. However, once the conditions of the process are well defined, some limitations may be neglected and new reaction invariants can be detected.

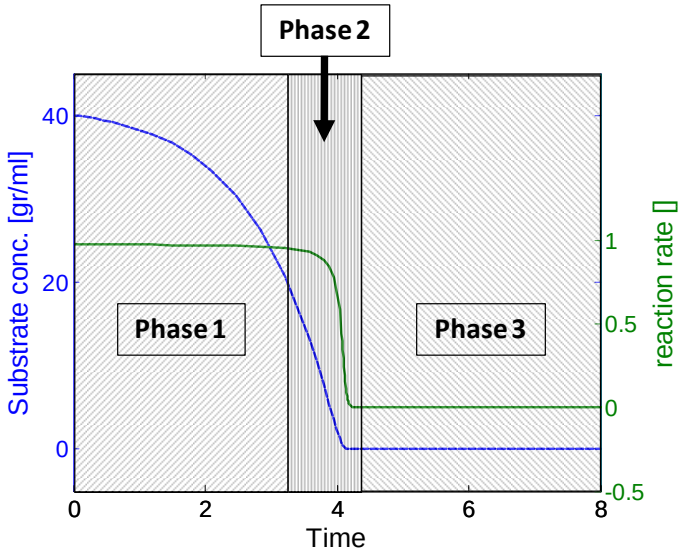

Figure 3.1. Behavior of a switching function in dependence of the limiting species.

Although no general method exists for this procedure, it is possible to achieve important reductions with some engineering experience and mathematical background. An example of the capabilities of this approach is presented in section 7.1.1, were the state of the art model for Active Sludge Process (ASP) is reduced to one third of its size.

3.2.3 SENSITIVITY ANALYSIS

The basic principle of model reduction through sensitivity analysis is to eliminate the components which are not relevant for the accuracy of the model. To correctly reduce the order of the system it is important to eliminate only the species, which have both a weak effect on the model outputs as well as on the important species of the model. Local sensitivities can be easily estimated via finite differences. If the Jacobian matrixes of both state variables and parameters can be obtained, dynamic sensitivities can be computed tailored to the numeric integration of the DAE system.

sDACL [46] is a code for efficient integration of general DAEs with sensitivity generation. Cheap computation is achieved through partial discretization tailored to the integration of general DAEs based on Orthogonal Collocation on Finite Elements (OCFE). This algorithm provides exact sensitivity information of the numeric integration, and can be implemented in any one-step integration method. sDACL has been customized to the staggered method for state and sensitivity integration showing similar computational efficiency. Finally, the algorithm is able to take advantage of the sparsitiy properties common in engineering models. With efficient simulation tools like sDACL, an efficient analysis of the dynamic sensitivities is possible.

3.2.4 LUMPING

Lumping is widely applied when dealing with systems with a large number of components. A process were the advantages of lumping has be demonstrated is steam cracking [74], where the millions of different components need to be grouped in a few categories to allow its simulation. Lumping can be divided in two main categories, proper and improper lumping. When speaking of proper lumping, each one of the reactants to be lumped appears exclusively in one lumped entity. On the other hand, improper lumping considers the possibility of one or more reactants contributing to different lumped entities. Although proper lumping is mostly used in mathematics, most reactions are only to be described precisely by improper lumping methods. A nice illustrative example is of proper lumping given by [70]:

Let us assume we have the following three component system:

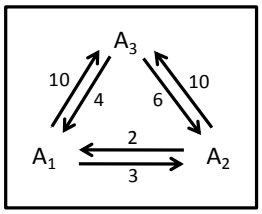

Figure 3.2: Three-component monomolecular reaction system, the numbers on the arrows represent the back- and forward reaction constants.

We consider the system to be described by the following monomolecular reaction scheme:

$$\frac{dA_1}{dt} = -k_{1,1} * A_1 - k_{1,2} * A_2 - k_{1,3} * A_3$$

$$\frac{dA_2}{dt} = -k_{2,1} * A_1 - k_{2,2} * A_2 - k_{2,3} * A_3 \qquad \textbf{3.6}$$

$$\frac{dA_1}{dt} = -k_{3,1} * A_1 - k_{3,2} * A_2 - k_{3,3} * A_3$$

where k is the monomolecular rate matrix

$$\mathrm{k} = \begin{bmatrix} 13 & -2 & -4 \\ -3 & 12 & -6 \\ -10 & -10 & 10 \end{bmatrix} \qquad \textbf{3.7}$$

and fulfils the conditions mentioned bellow:

- Nonnegative rate constant
- Mass conservation
- There exists an equilibrium composition $A_i^* > 0, i = 1,2,\dots,n$ such that: $KA^* = 0$

For the system to be lumpable with proper lumping it is necessary and sufficient to find a K' matrix such that:

$$M * k = k' * M \tag{3.8}$$

In this example:

$$\begin{bmatrix}1 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}13 & -2 & -4\\ -3 & 12 & -6\\ -10 & -10 & 10\end{bmatrix} = \begin{bmatrix}10 & -10\\ -10 & 10\end{bmatrix}\begin{bmatrix}1 & 1 & 0\\ 0 & 0 & 1\end{bmatrix} \tag{3.9}$$

In conclusion, it is possible to represent the system with two components, one representing A_3 and the second representing a combination of A_1 and A_2 Figure 3.3.

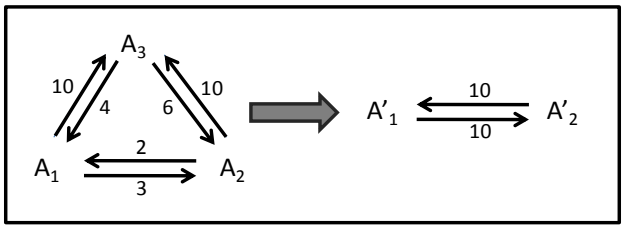

Figure 3.3: Lumping a monomolecular three-component reaction into a two-component reaction

Unfortunately, real processes can rarely be described by proper lumping since for most systems there is no M matrix which fulfils eq. 3.8. On the other hand, the theory of semi-proper and un-proper lumping can become rapidly complicated and are not discussed in this work [75]. However, it is very important to consider the possibility of lumping different species, especially in systems with a large amount of components or cells with similar behavior.

3.2.5 PERTURBATION THEORY

In some problems, it may be the case that a small parameter can be detected (usually ε) such that the solution does not differ significantly if ε = 0. Perturbation theory is a systematic mathematical approach to find the solution for the case ε = 0. Perturbation theory might be seen as the pure representation of model reduction. The model is analyzed to eliminate the parameters which cause very small changes in the dynamics of

the model. Again the real challenge is to find these parameters in complex nonlinear processes. Still, the method of perturbation theory allows fairly accurate results if process knowledge and an iterative method are combined.

A good example is modeling of a projectile thrown vertically into the air. An important question to be answered is if air friction should be considered in the model.

$$\frac{d^2x}{dt^2} + \frac{\kappa\Gamma}{mg} * \frac{dx}{dt} + 1 = 0$$
$$x_0 = 0; \quad \frac{dx_0}{dt} = 1 \qquad \textbf{3.10}$$

where $\kappa = [m * kg/s]$ is the friction constant, $\Gamma = [\mathrm{m/s}]$ the initial velocity of the projectile, $m = [kg]$ its mass, $g = [m/s^2\ ]$ gravity acceleration, and $t = [s]$ the time.

It is possible to consider that the relation $\varepsilon = k\Gamma/mg$ is very small, for low initial velocities, a body with big mass, and low air friction, and apply regular perturbation theory to find an approximate solution to eq. 3.10. The solution obtained.

$$x(t) = t - \frac{t^2}{2} + \varepsilon * \left(-\frac{t^2}{2} + \frac{t^3}{6}\right) + \varepsilon^2 * \left(\frac{t^3}{6} - \frac{t^4}{24}\right) + O\varepsilon^3 \qquad \textbf{3.11}$$

Eq. 3.11 is the exact same solution one would obtain after calculating the Taylor series expansion of the exact solution.

3.2.6 TIME SCALE ANALYSIS

Reaction systems are characterized by an interactive combination of fast and slow reactions. This gives rise to stiff equation systems which are difficult to solve with numerical methods. Substituting the fast reactions by algebraic equations, under the assumption that these species are instantaneously at equilibrium, has proven to be a very effective model reduction method. The Quasi Steady State Assumption (QSSA) has been widely applied in chemical engineering and biochemistry [76]. For the QSSA method empirical knowledge of the system to be analyzed is required. The engineer must decide which modes are fast enough to be considered steady state. Even though some mathematical methods have been developed to identify fast modes in equation systems, a generalized implementation of the method has proved to be difficult.

Stamatelatou [64] published a nice application of time scale analysis for biochemical reaction steps in CSTR. In this paper, Stamatelatou provides a systematic answer to find and remove the fast dynamics of acidogenesis. A simple example taken from this paper should illustrate the concept of the slow manifold.

Consider the following model describes a chemostat process:

$$\frac{dX}{dt} = -D * X + Y * r$$
$$\frac{dS}{dt} = -D * S + F - r \qquad \textbf{3.12}$$

where X and S are the biomass and substrate concentrations respectively, D is the dilution rate, Y is the biomass yield, $F = D * S_0$ is the feed rate of the substrate, S_0 is the substrate concentration in the feed, $r = \mu_{max}/Y_{max} * S/(K\_S + S) * X$ is the reaction rate, μ_{max} is the maximum specific growth rate constant of the biomass, K_S is the saturation constant, and assume that it operates under low dilution rate relative the growth of microorganisms: $D/\mu_{max} \ll 1$.

The work of the experienced engineer is to detect the fast and the slow dynamics. Stamatelatou deliberately selects the wrong dynamics to show the limitations of the approach. Nonetheless for sake of clarity, the correct dynamic is considered to be the fast one. In this simple system, it is very easy to foresee that biomass growth has a slower dynamics than substrate concentration.

Defining the reaction invariant:

$$z = X + Y(S - S_0) \qquad \textbf{3.13}$$

We can reformulate eq. (3.12) with the following two expressions, depending on which is considered to be the fast and the slow dynamic of the process. Considering that the dynamics of the biomass is fast enough to be substituted by an algebraic equation, the equation system can take the form:

$$\frac{dz}{dt} = -D * z$$
$$\frac{dS}{dt} = D * (S_0 - S) - \frac{Y}{X} * r * (S_0 - S + z) \qquad \textbf{3.14}$$
$$X = (z - S_0 + S) * Y$$

If on the contrary, the dynamics of substrate is considered to be faster, the following expression is more appropriate:

$$\frac{dz}{dt} = -D * z$$
$$\frac{dX}{dt} = -D * X + \frac{(z - X)}{(S - S_0)} * r \qquad \textbf{3.15}$$

$$S = \frac{z - X}{Y} + S_0$$

Now let us make a phase diagram where the trajectories of S vs. X are plotted for different initial conditions. In addition, we can see the trajectories of the two reduction hypothesis made (setting z equal to zero).

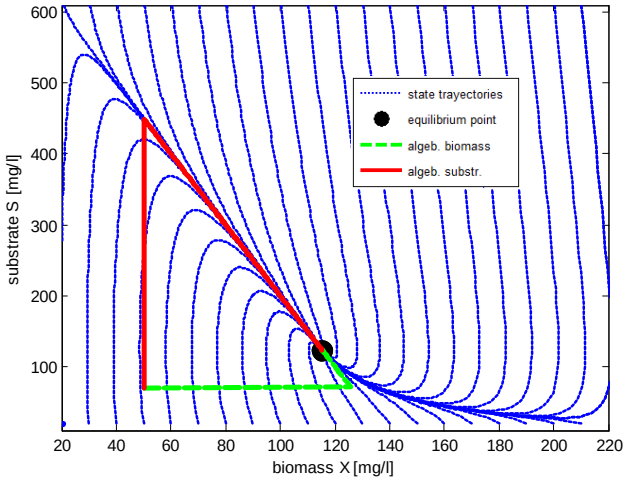

Figure 3.4: Phase diagram of full order model (3.12). Comparison with reduced models in a chemostat process

The method of the invariant manifold can be very effective to reduce stiff systems. Nevertheless, finding the fast and slow dynamics rapidly becomes a difficult task in large systems.

Finally, it is worth recalling, that the field of model reduction is showing important advantages towards reduction of complex nonlinear systems. New techniques in combination with advances in computer technology permit the solution of large systems in a systematic manner [77-80]. In near future, submodel building will become an automatic task in the MR program. By these means, a set of submodels should be automatically available to serve recognition purposes. Still, today the proper reduction of complex models requires much work and collaboration of different disciplines. The efficient recognition of a process regime still depends on the proper building of a set of submodels able to describe the process states accurately while being fast to compute, robust, and differentiable from each other.

4 OPTIMAL EXPERIMENTAL DESIGN

In chapter 3, methods to analyze model structure, parameter set, and its interaction in order to create models striving to defined objectives are discussed. Nevertheless, since the data set against which the model is fitted is at least as important as the model, a discussion regarding how to design experiments, in order to allow a better analysis of models and estimation of the parameter set has great significance.

The principal quality of models is its capacity to mirror a defined system. It is therefore not possible to determine the usefulness of a model without testing its description accuracy. In order to do so, observations of the real system have to be made and the similarities between model predictions, and observations of the real system, are to be compared. Models are created under numerous assumptions and for strictly limited systems. It is important that experiments carried out to test models consider these limitations.

The data set is obtained from observations carried out in defined experiments. Searching for the parameter vector of the model, which best describes the data set, is known as parameter estimation or model fitting. In physics, this is known as the inverse problem. Inverse problem theory studies the appropriate manner to infer the values of the parameter set using the results obtained from observations [24]. The theory of inverse problems has been well described for linear systems, and although it can also be applied to nonlinear problems, it becomes rapidly complex and requires high computation expenses. For this reason, more conventional approaches are applied for engineering applications. Even though some assumption need to be made and the calculations represent merely approximations of the state of information, experience has shown that simple methods deliver fairly accurate results and are useful in many processes [81].

To understand a system and model its behavior, a long iterative approach is required. First the system has to be observed, to achieve this, some experiments are carried out. Next, a hypothesis is made based on the observations. Following, a model is created on the foundation of the hypothesis, information from literature, and models created for the same or similar systems. The model has then to be fitted to the data set. An antithesis is created and the model is adapted. New experiments are required based on the information gained during the development of this process. Variables to be

measured, frequency of sampling, experimental conditions, among others are updated and a new set of experiments is planed and performed. The model is fitted to the new data sets, and this iterative procedure continues until the accuracy of the predictions is accepted or it is considered that the system cannot be described.

Optimal Experimental Design (OED) deals with the selection of the optimal experimental setting. Methods like Principal Component Analysis (PCA) or Partial Least Squares (PLS) search for data correlation to reduce the dimension of the data set [82]. Also more advanced methods in Knowledge Discovery of Data (KDD) like data mining [83] have been developed for treatment of large data sets. These methods study the data characteristics to find new relations between variables and create black-box type models which describe it.

For modeling purposes, alternative methods have been developed which consider the propagation of data uncertainties through the model and its reflection on the parameter set. In literature, these methods are grouped under the field of Model Based Design of Experiments (MBDoE).

MBDoE can be divided in two main categories: 1) parameter accuracy and 2) model discrimination. In parameter accuracy, the optimization problem searches for the experimental settings with the highest state information content over a parameter set, considering the characteristics of the model being fitted to increase model identifiability [81] (section 4.1.2). MBDoE takes advantage of the dynamics of the process to increase the sensitivity of the model outputs with respect to the parameters. Large parameter sensitivities reduce the confidence intervals of the parameter set, increasing the accuracy of the parameter estimation. In other words, it increases the information obtained from the experiment to fit the model parameters with the highest possible accuracy.

In the general case, various models are proposed as possible candidates to describe the system and the model with the highest probability to describe correctly the system is selected. In this case, the optimization problem has a different objective. The goal is to maximize the difference between the outputs of the candidates. The set of experiments is chosen in an effort to maximize the distinguishability between candidates. It is worth to highlight that model distinguishability is closely related to model identifiability, since accurate parameter estimations increase the probability of correct discrimination between candidates.

MBDoE, has been widely investigated and has shown important progress in basic research and industry. Not only theoretical work but also an important number of applications of MBDoE have been published, the most representative examples are [84-87].

4.1 THE EXPERIMENT

In practice, some parameters included in the model are not known beforehand and need to be estimated to enhance model description; the model is fitted to experimental data. The accuracy on the estimated parameters is influenced by two factors.

- The first one is the quality of the data obtained from the experiments. Measurements are inevitably subject to uncertainties. For this reason, data sets should not be considered observations but a "state of information" acquired on observable variables.
- The second one is the sensitivity of the objective function with respect to the parameters. The structure of the model defines the interrelations and effects of the parameter values in the output vector. Outputs which are highly sensitive to changes in the parameters allow a precise estimation of the "exact" parameter set.

An optimal set of experiments should be designed considering both factors mentioned above.

Experiments can be defined by the following conditions [81].

Variables that can be controlled:

- $u_E(t)$ time-dependent control values
- w_E constant control values
- $y(t)$ measurements in the selected time intervals
- $y(0)$ initial conditions
- t_{sp} experiment duration

Parameters that cannot be controlled:

- η_s Systematic error
- η Stochastic error

Systematic errors are errors on measurements which can be detected quantified and reduced. They present a mean value different from zero and error bias. On the other hand, stochastic errors can be defined as the difference between two identical experiments.

Experiment outputs are represented by the vector $y(t)$:

$$y^{mes}(t) = f(y_E^0, u_E(t), \mathrm{w_E}, \eta_s, \eta, t_E) \quad \forall t \in t_{sp} \tag{4.1}$$

Where the sub index E represents the variables of the experiment, and t_{sp} represents the time span of the experiment.

Model outputs are to be compared against the experiment outputs. For this reason, model outputs are represented in a form similar to experiments outputs. Model outputs (model predictions) are represented by the vector $y^{calc}(t)$.

$$y^{calc}(t) = f(\dot{x}(t), x(t), u(t), w, \theta, t) \tag{4.2}$$

Let us considering that all assumptions mentioned above are true, all sources of systematic error have been minimized, and that all control variables have the same value for the model and the experiment. If this is true, the experiments output vector $y(t)$ equals the sum of the models output vector $y^{calc}(t)$and a stochastic error η

$$y^{mes}(t) = y^{calc}(t) + \eta \tag{4.3}$$

The stochastic error η is considered to have multivariate normal distribution with mean zero, where C_y represents the variance covariance matrix of *n* measurements:

$$C_y = \begin{bmatrix} \sigma^2_{y1,y1} & \cdots & \sigma^2_{y1,yn} \\ \vdots & \ddots & \vdots \\ \sigma^2_{yn,y1} & \cdots & \sigma^2_{yn,yn} \end{bmatrix} \tag{4.4}$$

4.1.1 THE MAXIMUM LIKELIHOOD

Many methods for the quantification of the distance between model predictions and the observations can be found in literature [88]. The most common equation to quantify the residual is the Least Squares (LSQ).

$$LSQ = (y^{calc} - y^{mes}(t))^2 \tag{4.5}$$

where y^{calc} and y^{mes} represent the vectors of the calculated output and of the measured value respectively with length n_y.

The LSQ is a quadratic equation making it appropriate for most optimizers and easy to compute. Nevertheless, it fails to consider important factors, as scaling multiple outputs

and considering measurement accuracy. For this reason, the Maximum Likelihood (MXL) is preferred in process engineering:

$$MXL = \frac{1}{2}\frac{\left(y^{calc}(t,x) - y^{mes}(t)\right)^2}{C_y} \qquad \textbf{4.6}$$

The variance-covariance matrix C_y serves two purposes. On the one hand the outputs related to very noisy data has a smaller impact on the criterion. On the other hand, different scales between the outputs are also weighted. For MR, the theory for approximation of the confidence region for parameter estimation, which is based on the MXL criterion, is also essential. The probability density function of the parameter set and its correlation define the confidence region for parameter estimation. The size of the confidence interval is a direct indicator of model identifiability.

4.1.2 MODEL IDENTIFIABILITY

As stated in chapter 3, the meticulous examination of model dynamics and relation between states, allows a better understanding of the behavior of the model, and presents alternative ways to obtain similar dynamics with simpler and more tractable models.

Nevertheless, the reduction of a model depends on its parameter set. For this reason model reduction is closely related to parameter identification. In short words, the model has to be correctly fitted before it can be reduced. Moreover, model reduction affects the identifiability of the model. When a model is reduced some parameters are eliminated. Depending on the correlation of the parameter set, the confidence interval of the new parameters will be reduced. In other words, model reduction can also be applied when complex models present low identifiability. At this point an important question is to determine how to reduce a model correctly without knowing its exact parameter values. Unfortunately, science has no explicit answer to this question.

In the previous section different methods to analyze and alter the structure of models were discussed. However, it is of great importance to keep the objective of modeling in mind. Since a model is used to mirror a physical system, the accuracy of the prediction of the model needs to be quantified. To be more precise, the capability of the model to describe the process behavior excluding the effects of uncertainties has to be evaluated. Observations of the system to be described have to be made in order to test the predictions of a model. These observations contain various sources of error. Summed to human error, even most advanced measurement techniques present uncertainties and produce data with deviations from the physical quantity. Therefore, it is crucial to distinguish model prediction error (systematic error) from observation error (stochastic error.

For these reasons the effects of data variance on the probability distribution function of the parameter set has to be considered. The structure of the model is responsible for the propagation of data variance to the confidence interval of the parameters Figure 4.1. It is the model itself which determines the level of accuracy obtained through parameter estimation considering the quality of the data set. In other words, the quality of the model can only be established after knowing how exact can the parameter set be estimated. This again is directly dependent on the state of information. In conclusion, not only the objective for the implementation of the model, but also the quality of the experimental data have to be considered to select the optimal structure of the model.

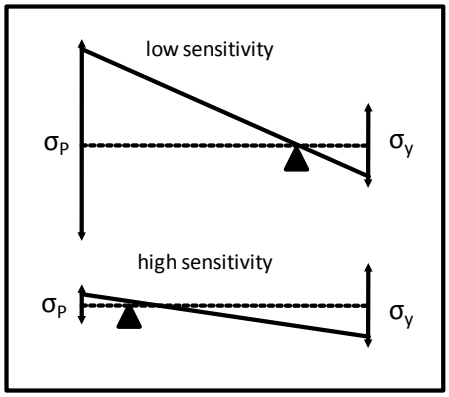

Figure 4.1: Effect of sensitivities in parameter estimation accuracy. σ_P and σ_y represent standard deviation of parameters and measurements respectively.

Asprey [89] defines identifiability as the quality of a model to present a monotonic behavior. In other words each parameter set shows a unique combination of outputs.

To quantify model identifiability the objective function Φ_I is maximized eq. 4.7 keeping the difference of models output smaller than some threshold eq. 4.8.

$$\Phi_I = (\theta - \theta^*)^T W_\theta (\theta - \theta^*) \qquad 4.7$$

s.t.

$$\sum_{i=1}^{n_y} \left(y(u(t),\theta) - y(u(t),\theta^*)\right)^T W_y \left(y(u(t),\theta) - y(u(t),\theta^*)\right) < \varepsilon_y$$

$$\forall u(t) \in U \qquad 4.8$$

$$f(y(t),\dot{x}(t),x(t),u(t),w,\theta,t) = 0$$

$$\theta_i^L \le \theta_i \le \theta_i^U ; i = 1,\ldots,P$$

$$\theta_i^{*L} \leq \theta_i^* \leq \theta_i^{*U}; i = 1,...,P$$

where Φ_I is the objective function to maximize, θ and θ^* the parameter sets with $\theta \neq \theta^*$, and W_θ an arbitrary weighting matrix.. $y(t)$ is a vector of n_y outputs, $\dot{x}(t)$ is a vector with the derivatives of the state variables, $x(t)$ is a vector with n_s time-dependent variables which define the system, $u(t)$ a vector of n_u time-dependent input variables, w is a vector with n_w constant input variables, and the superscripts L and U represent the lower and upper bound respectively.

The approach proposed by Asprey represents a high computational burden. In addition global optimality should be guaranteed, which increases computation time even more. Nonetheless, since identifiability studies are realized in the early stages of model developments and discrimination, this approach might offer important reduction in later experimental and design effort.

The standard procedure to increase mode identifiability is to detect the parameters with small sensitivity and high correlation and take them out of the optimization variables. The number of parameters to be estimated can then be iteratively increased as the identifiability of the model increases. This procedure enables a more accurate parameter estimation in early stages and thus a better design of experiments. Nevertheless, the selection of these parameters, which are assumed to be highly correlated and with low sensitivity, is carried out based on the Fisher Information Matrix (FIM), which is an approximation based on first order derivative information. This very rough approximation is only correct near the vicinity of the parameter values used to calculate the FIM. Therefore, ideally, all parameters should be considered in the parameter estimation.

4.2 THE FISHER INFORMATION MATRIX

4.2.1 THE CONFIDENCE INTERVAL

The propagation of data uncertainties through the model and its effect on parameter estimation has been subject of short discussion in previous sections. When a model is to be fitted to a data set, the accuracy of the estimation of the parameter vector depends not only on the accuracy of the data set, which can be quantified by the variance-covariance matrix, but also on the structure of the model. It is essential to understand two characteristics of the model structure:

- The impact of each parameter on the dynamics of the model, therefore its influence in the outputs.

- The interaction between all parameters of the parameter set, hence how they correlate to each other.

The most precise formulation of the problem is to calculate the probability density function of the estimated parameter set.

Assuming that:

- All parameters are constant
- Input variables behold no uncertainties
- Stochastic error can be described by a normal distribution over model predictions with the exact parameter set (which is unknown)
- There is no correlation between measurement uncertainties
- Measurement uncertainties can be described by a normal distribution with mean equal to zero.

The probability density function of the estimated parameter set can then be estimated with the following equation:

$$\wp(\theta_{es}) = (2\pi)^{-2nN} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{N} \frac{\left(y_{i,k}^{cal}(\theta_{es}) - y_{i,k}^{mes}\right)^2}{\mathrm{C_y}} \right) \prod_{k=1}^{N} (\det C_{res}(t(k)))^{-\frac{1}{2}} \quad \textbf{4.9}$$

where $\wp$ represents the probability density function, θ_{es} the estimated parameter vector, $y_{i,k}^{mes}$ and $y_{i,k}^{cal}$ the measured and the calculated variables respectively, C_y represents the variance of the measurements, and C_{res} represents the variance covariance matrix of the residuals.

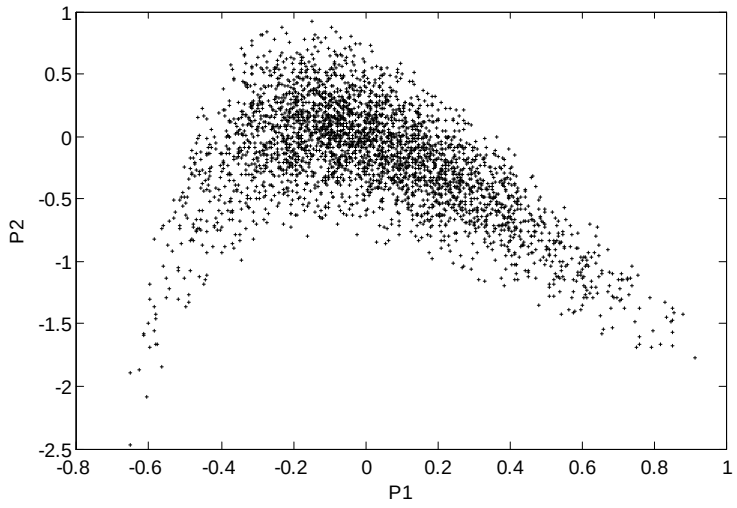

Figure 4.2: Confidence interval from the Lin model, obtained with Montecarlo simulation.

The most effective method to compute the real nonlinear confidence interval is to carry out Montecarlo simulations [90]. Figure 4.2 shows the confidence interval of a two dimensional parameter set. The model used for this example is a model for *E. coli* fed-batch cultivations presented by Lin [91] (see Appendix A), the selected parameters are: q_{Smax}, maximal Substrate uptake capacity, and Y_{XA}, and Yield acetate to biomass. Even though, for this calculation, white noise was added to the data set (simulated normal distributed variance with zero mean), propagation of the uncertainties through the model, results in a none normally distributed confidence interval of the parameter set. This effect is caused by the nonlinearities of the model.

4.2.2 APPROXIMATION OF PARAMETER VARIANCE-COVARIANCE MATRIX

As stated before, the accuracy of the parameter estimation depends on the sensitivity of the objective function to changes in the parameter set. To be able to plan experiments which reduce the confidence interval of the parameter, increasing the accuracy of the parameter estimation, a cheap technique to compute the confidence interval is required. Based on the Cramer-Rao Bound (CRB) theorem:

$$C_p \geq FIM^{-1}(\theta_{es}) \qquad \textbf{4.10}$$

where C_P and FIM are the variance-covariance matrix of the parameter set and the inverse of the FIM respectively, and θ_{es} the vector with P estimated parameters. Hence we can calculate the lowest bound of the variance-covariance matrix in a cheap manner.

In practice, the FIM is computed to approximate the confidence interval and find the optimal experimental set.

$$FIM(\theta_{es}) = \sum_{k=1}^{N} \left(\left(\frac{dy(x,u,t)}{d\theta} \right)_{\theta_{es},t_k} C_y^{-1}(t_k) \left(\frac{dy(x,u,t)}{d\theta} \right)^T_{\theta_{es},t_k} \right) \qquad \textbf{4.11}$$

where θ_{es} is the estimated parameter vector, x is the state variables, t_k is the time point of the measurement, u is the control variables, and C_y is the variance-covariance matrix of the measurements. In other words, FIM gives a linear approximation of the confidence interval region of the parameter set to be estimated.

From equation 4.11 can be deduced, that the FIM depends on the control variables of the experimental setup. We now have all the information we require to build an optimization problem and find the optimal experimental setup which maximizes the state of information obtained from experiments, an approximation to this can be

calculated with the FIM. Finally, the FIM must be compressed into scalar form to create an appropriate objective function for optimization. To achieve this, different criteria are implemented in an effort to create an efficient measurement of the variance-covariance matrix in scalar form. A graphical description of the most common criteria for the two dimensional case can be seen in Figure 4.3.

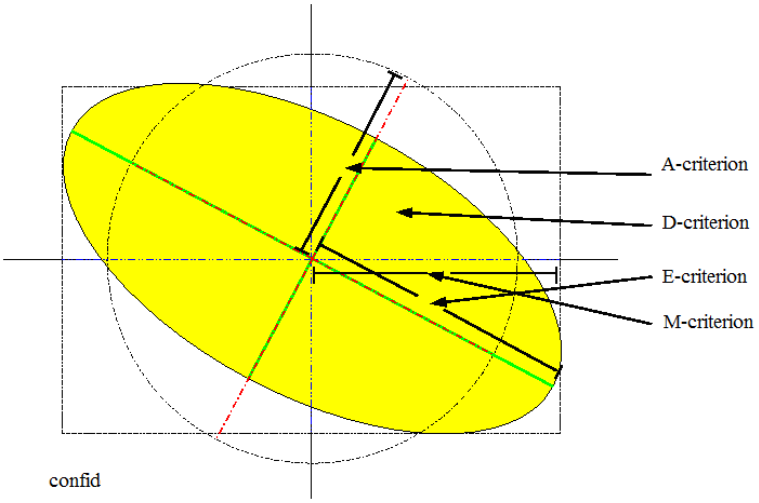

Figure 4.3: Criteria for optimization [92]

Summarizing, the objective of design of experiments for parameter accuracy is to find an experimental set which increases the information content of the data and hence allows a more accurate parameter estimation. The most common criteria are shown in Table 4.1

Table 4.1: Criteria for confidence interval quantification [92].

A- criterion	$\frac{1}{n} trace(C)$	[93]
D- criterion	$det(C^{\frac{1}{n}})$	[94]
E-criterion	$\max(eigenvalue(C))$	[95]
M- criterion	$\max\left(C_{ii}^{-\frac{1}{2}}\right)$	[96]

4.2.3 Limitations of the Fisher Information Matrix

It is essential to consider the limitations of the FIM at all time in order to avoid wrong interpretation of the results or its misuse in experimental design. FIM is the result of

two important mathematic approximations. The validity of these approximations is restricted to a small range near the region of calculation Figure 4.4 shows how the nonlinearities of the model affect the shape of the confidence interval as it grows.

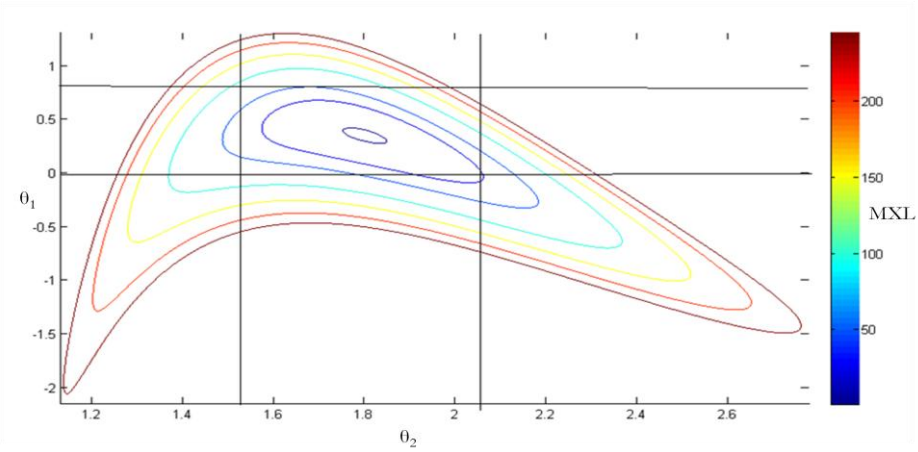

Figure 4.4: Shape of the confidence interval for different variance values from the Lin model (appendix A). The confidence interval can be approximated by an ellipse near the exact value.

Once the variance of the data set has been quantified and set constant, the confidence region of the intervals is determined by the objective function of the parameter estimation program and its sensitivity to changes on its parameters. If we utilize the MXL to estimate the residual of the models prediction, and the model is linear, the resulting objective function will be a quadratic function. This means that we only need to know the Hessian of MXL and the variance covariance matrix of the measurements to calculate the exact confidence interval, again strictly for the case of linear models. On the contrary, nonlinear models cannot be described so easily.

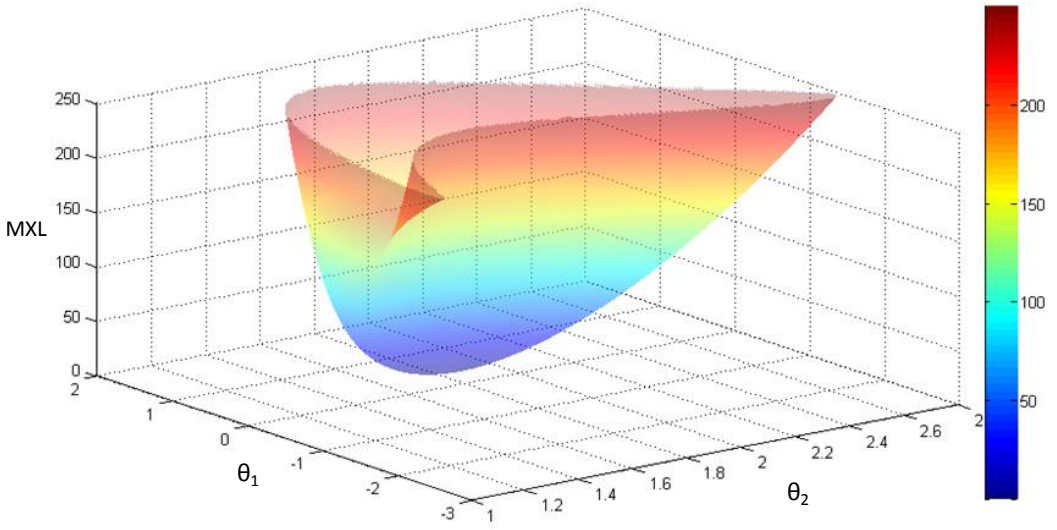

Figure 4.5: Objective function of a nonlinear model (appendix A) with respect to changes in a two dimensional parameter set.

Depicted in Figure 4.5. is the form of the objective function where the shape differs from a parabolic function. Still, to avoid the expensive computation of the Hessian tensor, which may be even impossible for large problems, the gradient information of the state variables with regard to the state vector and to the parameter set is used to obtain a fair approximation. This approximation, which as stated before is only accurate near the vicinity of the exact parameter set, is the FIM.

4.3 MODEL DISCRIMINATION

In Engineering, the use of mechanistic models for the simulation of complex systems is a widespread procedure [48]. Models are created on the basis of the physicochemical phenomena occurring in the process. Rigorous models are, however, compared against the so-called black box models laborious to develop, adapt, and fit. The challenge increases when processes presenting a non-linear and dynamic behavior are to be modeled. As a result, various models, which can differ in structure, size, application and / or complexity, are usually to be found in literature. The questions that arise are:

- How to quantify the capability of the different models to fit the process?
- How to select the model that best fits the needs of the process?
- How to create a minimal set of experiments to achieve question one and two?

In statistics, the field concerned with finding the most appropriate model for a specific process is called Model Discrimination (MD). MD deals with maximizing the probability of a correct model selection considering data uncertainty. To achieve this, an optimal set of experiments is planed with the aid of statistical tools. This approach is widespread in biology and biotechnology [97-99], in physics [100] and in process engineering [101]. Solutions for the adaptation of linear and nonlinear systems and for the discrimination of dynamic models [102-104] can also be found in literature.

The first work to address the problem of finding the most appropriate model to describe a set of data was presented by Hunter [105]. Hunter defined the goal of MD as: *"Perform the experiment which will most strain the incorrect model to explain the data"*. MD is briefly presented in this chapter, for a better understanding of this approach we refer to [52, 81, 106].

The general procedure for MD is rather intuitive. The first step is to select some models which could describe the system, called candidates. Following, experimental setting is designed in an effort to maximize the probability of choosing the correct model. To

achieve this, the experimental conditions which maximize the difference in the outputs of each candidate are selected.

In statistics, Bayes' theorem calculates the precise probability function of a model being the most appropriate to describe the data set. Nevertheless, it requires accurate information of the distribution density function and only quantifies the probability associated with a model prediction. The engineer needs to consider various factors when selecting a model and requires an effective and simple method.

To state the problem we first discuss how to quantify the "goodness" of a model. This is not an easy task. The obvious procedure is to calculate the residual res. The residual is composed of two different types of error.

$$res = \eta + \eta_S \qquad \textbf{4.12}$$

The stochastic error η is caused by the uncertainties in the system and measurement methods and it should ideally be randomly distributed around the true value with mean $\mu = 0$. The systematic error $\eta_S$is caused by defects on the model, e.g. neglected phenomena, not considered external disturbances, incorrect structure, etc.

In order to calculate which model describes the processes with highest certainty possible, the variance of the data needs to be known and all models should have its "exact" parameter set. Both errors, stochastic and systematic, are expressed by a probability distribution function hence discrimination between models can only be correct within a certain probability.

Furthermore, the state of information and the identifiability increase together with the number of experiments. Therefore, the probability of selecting the "best" model increases with each new experiment. For this reason, in order to make the most certain decision, all candidates should be considered throughout the complete discrimination process. However, this represents an enormous experimental effort [106]. Franceschini [52, 81, 106] rather proposes to conduct a gradual discrimination of candidates in an iterative process. This procedure reduces computation costs in each iteration and allows a more specific design in the next iteration.

Table 4.2: Types of sum of square [22]

Error variance	$\sigma_{res}^2 = \frac{MXL}{n-p}$
Akaike's Final Prediction Error	$FPE = \sigma_{res}^2 * \frac{n+p+1}{n-p-1}$
Akaike's Information Criterion	$AIC = n * \ln(\sigma_{res}^2) + 2 * p$

Shortest Data Descriptor	$SDD = \ln(\sigma_{res}^2) + (p+1) * \ln(n)$

Where MXL is the Maximum Likelihood estimation eq. 4.6 value of the minimum sum squared, n the number of measurements, and p the number of parameters.

Unfortunately, the residual only shows the ability of the model to describe the data set. This is by far not the "best" model. Robustness, physical meaning, and extrapolation capacity are a few examples of important model characteristics which should be taken into account during MD.

Other forms of the sum of square intend to quantify further aspects of the models to help proper discrimination. The most representative criteria are presented in Table 4.2. The reader is referred to [107] for detailed summary.

Verheijen [22] proposes five selection criteria:

- Physicochemical criteria: physical meaning of the model
- Flexibility criteria: application in similar processes and reusability.
- Computational criteria: computation expenses and robustness
- Statistical criteria: accuracy of the prediction (level of state information).
- Engineering criteria: usefulness of the model in the real process

Still, he misses to mention the importance of identifiability as selection criterion. Between two candidates with the same description accuracy, the model with the highest identifiability should be selected.

4.3.1 MODEL DISCRIMINATION IN MECHANISM RECOGNITION

For recognition purposes MD is a straight forward procedure, since calculating the MXL is sufficient for selecting the best model. In MR, candidates have been generated to detect different regimes. For this reason description accuracy is the only discrimination criteria to consider. In addition, model distinguishability between the submodels should be guaranteed beforehand (during the model building stage).

During the recognition process, the program for MR computes the minimal length of the interval required to assure, firstly identifiability and next distinguishability. Model distinguishability in each time regime depends on initial conditions and process state information. An essential condition to apply MR to a process is to guarantee model distinguishability in every interval. This has to be considered when building the submodels.

In this section the difficulties to discriminate between various models have been discussed. From this, the complexity of model building can be deduced. A scientist cannot be expected to build correct models, when he is not even able to select the best among a group of models. In order to create an adequate model, the required characteristics of the model and the objective of modeling have to be previously defined.

To summarize, this section intends to emphasize two important aspects in modeling:

- Description ability is not the unique quality of a model worth considering during model building. Many characteristics have to be taken into account and there is never a simple manner to define the "best" model.
- The state information of the system (whether laboratory, pilot plant or real process) is essential for the selection of the optimal features of the model. It is necessary to know the state of information of the data set before or at list while the model is being built.

5 CODE GENERATION, SIMULATION AND OPTIMIZATION

In this project, various software packages and programming languages were used. Model and code generation were supported by MOSAIC (section 5.1.1) and SBPD (section 5.1.2). Simulations were carried out with standard ODE solvers in the simplest cases whereas the integrator sDACL (section 5.2.1) was applied for efficient integration and sensitivity computation with piece-wise constant inputs and parameters. Finally, different optimization algorithms (section 5.3) were required to solve the problems addressed in this thesis.

5.1 CODE GENERATION

5.1.1 MOSAIC

MOSAIC [44, 108] is a modeling environment, which combines equation-based modeling, use of symbolic mathematic language, and code generation, following a new modeling approach for equation reuse and support of different nomenclature conventions. The toolbox supports automatic code generation in a variety of programming languages (C, Fortran, Matlab, among others) and stores the information of the equation systems in the generalized markup languages XML and MathML.

The MOSAIC modeling environment is a tool to implement customized models aiming at:

- the minimization of modeling errors,
- the minimization of the programming effort,
- the avoidance of errors and effort in documentation,
- the encouragement and support of cooperative work.

Furthermore, MOSAIC promotes modular modeling at a high level Figure 5.1. This feature offers important advantages for hierarchical modeling as well as for efficient and simple development and analysis of reduced models in general and submodels in particular.

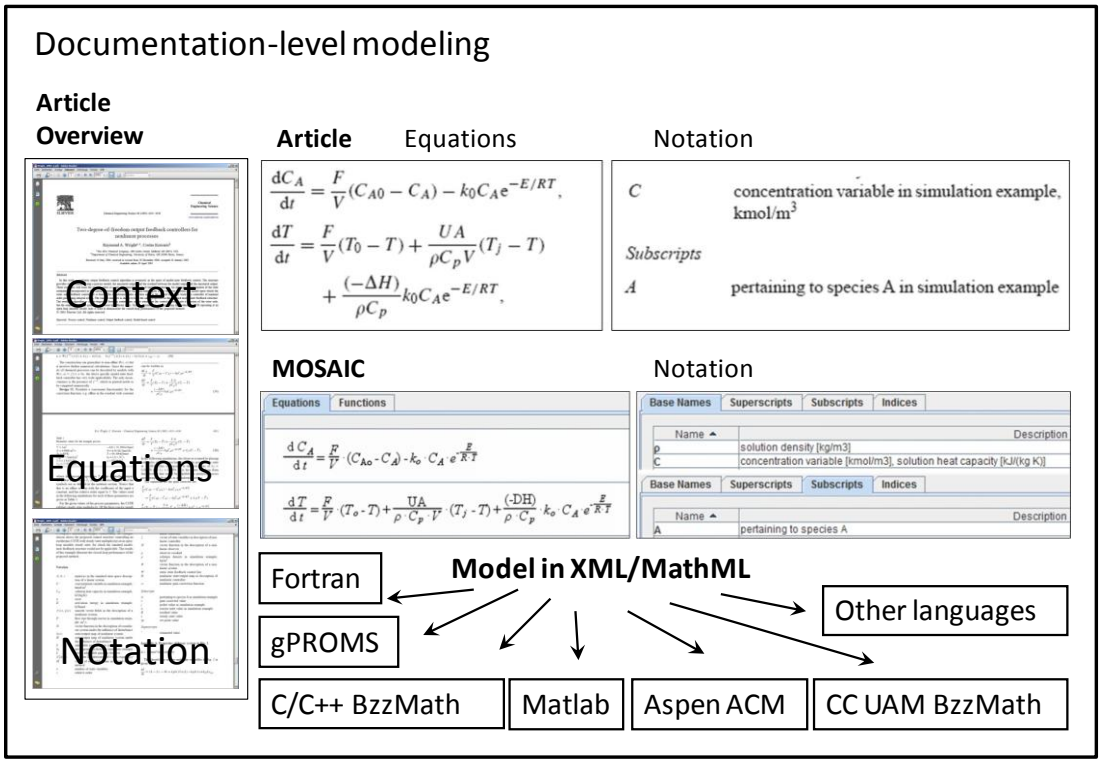

Figure 5.1: High level modeling with MOSAIC [46]

5.1.2 SBPD

SBPD is a Systems Biology Toolbox for MATLAB developed [45]. The toolbox offers an open and extensible environment for the analysis and simulation of biological and biochemical systems limited to ODE systems. An important advantage is support of the Systems Biology Markup Language (SBML) models, enabling model exchange within the systems biology community. Models are represented in an internal model format and can be described by entering biochemical reaction equations. The Systems Biology Toolbox for MATLAB is open source and freely available from http://www.sbtoolbox.org.

The toolbox is built in a modular way, as depicted in Figure 5.2. The base elements are objects of classes SBmodel and SBdata, which are used to represent models and experimental data.

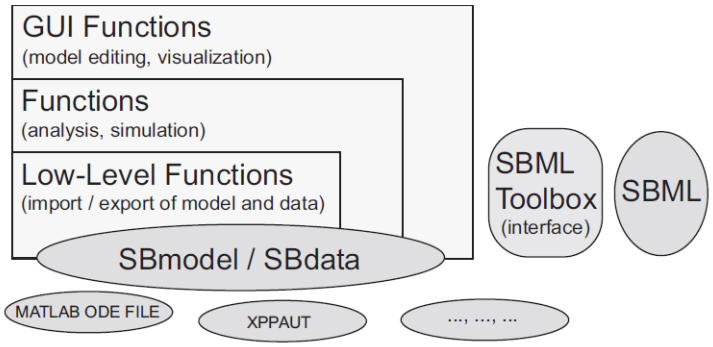

Figure 5.2: Modular structure of the toolbox. The toolbox is designed in a modular way. Optionally, freely available third party software packages are used, e.g. to (i) perform bifurcation analysis and (ii) realize an interface to SBML [45].

5.2 SIMULATION

This work shows that MR is a powerful tool to create simplified models and use them to describe complex processes. Solving the set of differential algebraic equations accurately even in the case of complex systems with highly nonlinear behavior is essential to enable the application of MR. Furthermore, because of the calculation of the FIM and optimization of nonlinear programs using optimality criteria, first and second order dynamic sensitivities offer important advantages [46].

5.2.1 sDACL

Logsdon [109] showed the equivalence of the Orthogonal Collocation (OC) to an implicit Runge–Kutta method, achieving the highest accuracy order. Also, being an implicit numeric method, OC presents great stability properties for systems with index one or higher. OC has shown very good performance especially when solving highly nonlinear models. In addition, extending the method to the OCFE has shown important reductions in computation expenses for the full discretization approach in dynamic optimization, transforming dynamic equations into algebraic ones. Prove of its efficiency solving typical process engineering problems is the important number of solvers for general differential-algebraic optimization problems developed based on this method [110]. OCFE may also be implemented on direct applying the partial discretization approach to solve stiff implicit DAE systems [111]. Besides, solvers based on OCFE have the ability of self-starting in high orders, an important characteristic for efficient partial discretization approach especially when having a high number of discontinuities or even multiple shooting.

For simulation in the MR program, the approach for large-scale dynamic models which takes advantage of the sparsity of the system and generates first and second order sensitivities is applied. The program sDACl developed by Barz [46] at the chair of

process dynamics and operation of the TU-Berlin has been selected for the integration of the DAE systems.

sDACL is a code for efficient integration of general DAEs with sensitivity generation. Cheap computation is achieved through partial discretization tailored to the integration of general DAEs based on OCFE. This algorithm provides exact sensitivity information of the numeric integration, and can be implemented in any one-step integration method. sDACL has been costumized to the staggered method for state and sensitivity integration showing similar computational efficiency. Finally, the algorithm is able to take advantage of the sparsitiy properties common in engineering models.

5.3 OPTIMIZATION

For the application of MR, optimization programs need to be solved including parameter estimation, optimal control time and global optimization. The interface environment Tomlab was utilized, to implement the different optimization programs with the code written in Matlab®. Tomlab is an optimization environment containing different state-of-the-art optimization software packages [112]. The external solvers are distributed as compiled binary MEX DLLs on PC-systems, and compiled MEX library files.

PARAMETER ESTIMATION

For parameter estimation purposes, the family of residual optimizers was selected in general and the Gauss-Newton method in particular. Solving the parameter fit problem with residual information has shown to be a very efficient method to overcome difficult regions caused by parameter correlation [113].

NONLINEAR QUADRATIC PROBLEM

During the optimization of the optimal time lengths, the optimization problem was formulated as solved based on Sequential Quadratic Problem (SQP) algorithm [114].

GLOBAL OPTIMIZATION

Finally, some problems require solution of high nonlinear programs where global optimality is required. For these cases, the approach to stochastic optimization based on Particle Swarm Optimization (PSO) was selected. PSO is a stochastic optimizer which has shown advantages over other stochastic techniques [2,4]. The principle of PSO is based on a population (swarm) formed by a certain number of particles (i), which search for a global optimum in the defined region of n dimensions.

6 AN APPROACH TO MECHANISM RECOGNITION

6.1 A SHORT INTRODUCTION TO MECHANISM RECOGNITION

The method for MR has been developed to recognize different regimes in dynamic processes. In processes which have no steady state condition, the MD approach can be applied to different regimes during the process. Similarly to steady state processes, where a reaction pathway is selected among a number of candidates applying MD, in dynamic processes, different regimes of the process with different pathways could be detected. This is the basic principle of MR. In dynamic processes models which do not fit the complete process data may still fit some time intervals. If the candidates set for discrimination in each time interval are first principle models, the best candidate is also an indicator of the predominating phenomenon in the selected regime. This approach does not only enable a better understanding of the process and its dynamics, but also an accurate description of complex dynamic processes with relatively simple models.

The ultimate goal of process modeling is maximization of prediction accuracy. To achieve this, usually model complexity is increased exponentially with respect to prediction accuracy. Drawbacks of this method are an unavoidable reduction of both, parameter identifiability and model robustness. To overcome these problems, experimental effort needs to be increased drastically and advanced simulation tools (software and hardware) are required. In other words, increasing model complexity cannot be achieved without increasing modeling and experimental effort exponentially. Still, in order to describe the dynamics of a complex system, complex and highly nonlinear models have to be developed.

Instead, dividing the process in different regimes with simpler dynamics and describing them separately, allows process description with multiple but simpler models. Most dynamic processes present two or more quasi steady state regimes, if these regimes can be detected, isolated and modeled by different simpler models, the process is described with remarkably less effort and higher accuracy. MR offers an adequate framework to analyze experimental information and model dynamics. The models for MR need to be

built on the basis of process expertise, knowhow and also taking advantage of modern mathematical tools. It is worth reminding that conclusions drawn by the recognition program depend strictly on the expertise during submodel building.

Consider we have a model which can describe both phenomena in a batch reactor: 1) diffusion and 2) reaction. This model would typically contain highly correlated parameters, which would affect the accuracy of the parameter estimation. We can reduce this model and create two submodels. A submodel 1, which can only describe the change in concentration caused by diffusion, and a submodel 2, which is only capable to describe concentration dynamics caused by reaction. Not only have both submodels less parameters, but because of their limitations (only diffusion or only reaction), a much more accurate detection of the regime where transport limits the process and the regime where reaction limits the process can be achieved. In this example, it would be possible to monitor the minimal speed of the stirrer needed to avoid transport limitation. By these means, it is not only possible to describe the process with simpler models and higher accuracy, but also information about the process is gained.

At this point an approach to MR [42] to detect different dynamics based on simplified models can be of great help. Complex models created to describe complete processes are reduced to form submodels limited to the description of a specified phenomenon. These reduced models, called submodels, have important advantages over the complex model as are:

- Analysis of the complex model
    - Because of the systematic mathematical approaches applied for its reduction, submodels help analyze, understand and identify the complex model.
- Simulation
    - Reduced models (submodels) present a higher identifiability, lower calculation costs and are more robust.
- Process understanding
    - The physical basis of the complex model and of the submodels allows a better understanding of the process and an easier detection of the key dynamics of the system.
- Monitoring
    - In conjunction with fault detection approaches, submodels help detect the influence of non measurable phenomena in the process.

6.1.1 ILLUSTRATIVE EXAMPLE

To illustrate the concept of MR, a short example based on the first published application of MR is presented. Further details of the process and recognition of regimes can be found in [43, 115].

MR was applied to monitor fowling in Membrane Bioreactor (MBR) processes for waste water treatment, which is a highly dynamic process [116, 117]:

- Backflush/relaxation is employed for approx.15 – 60 s every 3 – 10 min of filtration
- Maintenance cleanings are carried out every 2 – 7 days and main cleanings once or twice a year.

Various models for the description of membrane fouling can be found in literature. Nevertheless, most models have been developed and validated under steady state conditions. For the application of MR, five models obtained from literature were selected [118] to fit experimental data obtained in a test cell and in a pilot plant:

- Cake filtration

$$\frac{\Delta p * t * A}{V(t)} = \frac{\nu * \alpha * k}{2} * \frac{V(t)}{A} + \nu * R_M$$
$$R_C = \alpha * k * \frac{V(t)}{A} \qquad \textbf{6.1}$$

where Δp represents the pressure difference, t the time, A the area, V the filtered volume, ν the dynamic viscosity, α the specific cake resistance, k is a dimensionless constant, R is the resistance of the membrane (subscript M) and cake (subscript C) respectively.

- Standard blocking

$$\frac{t}{\dot{V}(t)} = \frac{1}{\dot{V}_0} + \frac{X_P * t}{L * A_0} \qquad \textbf{6.2}$$

where X_P represents volume of solid particles retained per unit filtrate volume, L the membrane thickness, $\dot{V}(t)$ the volumetric flow rate, $\dot{V}_0$ initial volumetric flow rate, and A_0 the initial area.

- Intermediate blocking

$$\frac{t}{\dot{V}(t)} = \frac{1}{\dot{V}_0} + \frac{\Delta p * s * t}{\nu * R_M * \dot{V}_0} \qquad \textbf{6.3}$$

where s represents blocked area per unit filtrate volume.

- Complete blocking

$$\frac{t}{\dot{V}(t)} = V_0 - \frac{\Delta p * s * \dot{V}(t)}{\nu * R_M} \qquad \textbf{6.4}$$

- Cake filtration

$$\frac{\frac{\dot{V}_0}{\dot{V}(t)} - 1}{\dot{V}(t)} = \frac{\alpha * X_0}{A * R_M} - \frac{\alpha * \dot{V}_r}{A * R_M} * \frac{1}{\dot{V}(t)} \qquad \textbf{6.5}$$

The results of model fitting in Figure 6.1 show, that none of the five models is able to fit the data. Furthermore, if physical boundaries of parameter are considered, the residuals obtained are extremely large. The physical explanation for the inability of all models to fit the data is that dynamics of this particular process are governed by different phenomena over time. The only solution to describe the process truthfully is to create a complex model which encloses all phenomena: 1) cake formation, 2) pore constriction, 3) pore blocking. In addition the model must foresee, the time periods where each phenomenon is dominant.

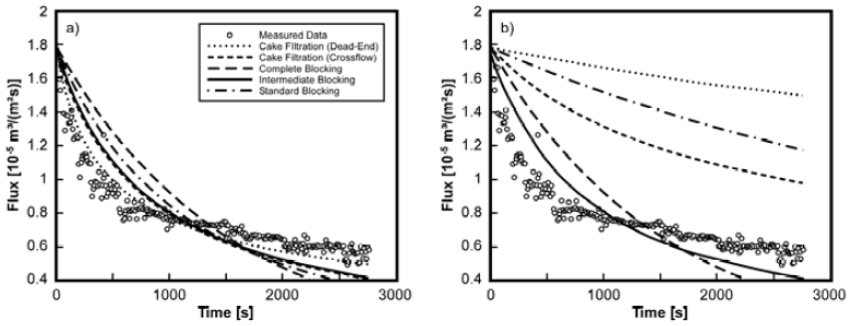

Figure 6.1: Model fit a) without setting bounds
b) with setting bounds for physical parameters. [119]

MR offers an alternative and very practical solution to this problem. To avoid increment on model complexity, the five models obtained from literature are used to describe discrete intervals of the process. By these means, each model is applied only in the time regime, where the model is able to describe the process. This allows not only a good

description of the complex process with simple models, but also enables an indirect detection of the dominant phenomenon in each regime.

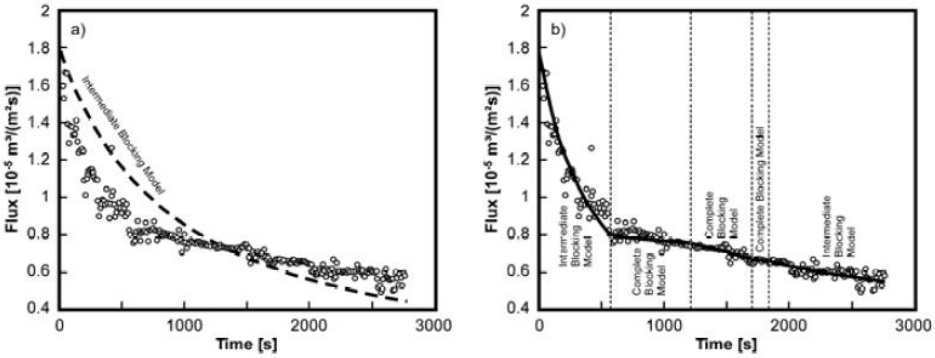

Figure 6.2: Comparison experiment/simulation using a) just one model. B) various models [119]

The program strictly detects the model which describes the present regime with the highest accuracy. Nevertheless, if models enclose physical understanding of the process, an indirect recognition of the dominant phenomenon is possible. In other words, MR can use first principle models to make indirect measurements of non-observable variables and bring new information to light. Figure 6.2 shows a regime detection with mechanism recognition.

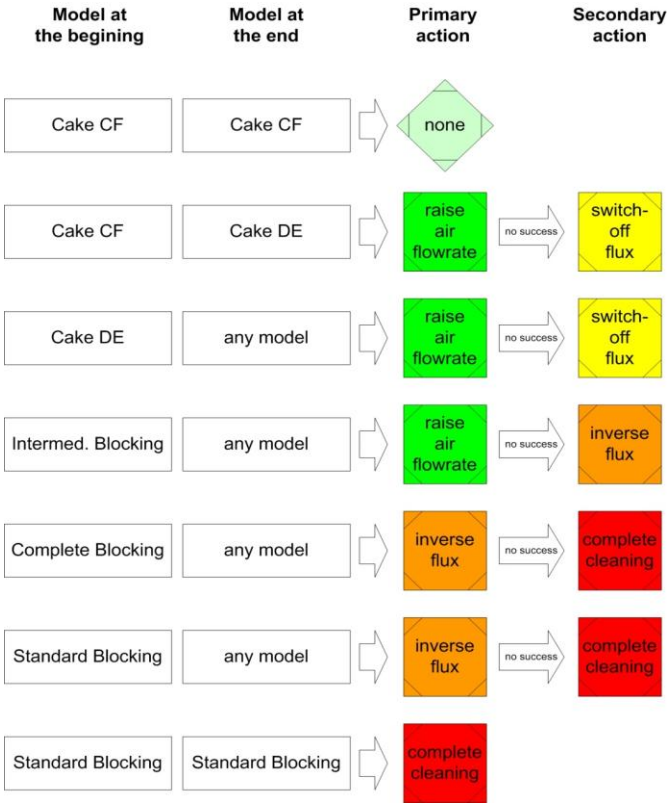

Figure 6.3 Cleaning strategy based on MR [43]

This ability to indirectly detect non-measurable variables can be exploited to monitor processes and allow a better operation. For the case of the MBR process, a scheme for action to take based on MR results was developed. This protocol facilitates a data based process maintenance, which increases the efficiency of the process significantly. As seen in Figure 6.3, the recognition approach can detect the fouling type and based on this a cleaning strategy is selected.

6.2 METHODOLOGY FOR MECHANISM RECOGNITION

In the previous example models with only one variable defining the state of the system are considered. This is a very practical condition since output input consistency between all models is automatically guaranteed. Moreover, it was possible to measure the single state variable at high frequency, which solves the problem of initial conditions for each new regime. Models with different number and types of state variables can still be applied for MR, but only in the special case where all models are observable (no general structure is needed) and the frequency of the measurements is high enough to assure continues initial value information for all state variables (no input-output consistency required). The solution of this type of problems is straight forward and will not be considered in this study. Unfortunately, in practice multiple state variables are to be considered and online measurement information cannot be assured. For this reason an approach to deal with these drawbacks needs to be developed. The method is required to allow a regime recognition also in complex systems with multiple state variables and limited process data.

The present study is constrained to following type of problems:

- Dynamic (non-stationary) process.
- Processes with two or more regimes.
- Detailed model available for the description of the general process of the form:
    - DAE system index zero or one
    - no discontinuities
    - known initial conditions
- Parameter boundaries have to assure global convergence with gradient based optimizers.
- Initial regime is known.
    - no model discrimination is required in the first interval.
- The sequence of the regimes is known.
- The minimal length of each time regime assures model distinguishability.
- The models present no bifurcation.

6.3 PROGRAM STEPS

6.3.1 SUBMODELS

In order to assure a consistent process description throughout all regimes, generate a proper general structure (section 6.3.2), and enable proper transfer of parameter values, all models are derived from one single complex model, which aims at description of the complete process. Model reduction techniques (chapter 3), empirical knowledge, analogies to other models from literature, and the state of information obtained from the process (chapter 4), are combined to create an adequate set of submodels. These submodels should fulfill all the restrictions mentioned previously (input-output consistency, parameter values transmission, general structure, etc) and be easily distinguishable.

Unfortunately, most general model reduction approaches are restricted to systems where Lyapunov stability can be proved. The development of a general method to reduce nonlinear models and find adequate submodels is a very interesting challenge for process modeling and novel promising approaches are already under study [73]. Nonetheless, a general method with reasonable computer expenses is still to be developed. For this reason the reduction of the complex model into submodels is still limited to specific solutions where modeling experience and process understanding are essential. Phenomena delimitation and model adaptation for each regime is still a difficult task. Much care has to be given to adequate description of the system maintaining identifiability and distinguishability between models. Still, this thesis shows the reduction potential of complex models and its advantages.

6.3.2 GENERAL STRUCTURE

It is essential to maximize submodel identifiability to increase the efficiency and accuracy of the MR program. Since identifiability is deeply influenced by the number of parameters to be estimated, it is convenient to distinguish the parameters in two groups:

- Process Constant Parameters (PCP)
- Regime-Wise constant Parameters (RWP)

In the reduction process to build the submodels, some parameters are present in all submodels and remain constant throughout the whole process (PCP), e.g. conductivity, limitation and inhibition constants. Since PCPs do not change over the process, the state of information of the complete process can be used increasing the number of measurements available for its estimation. This is not only important for PCPs, but also RWPs can be estimated with higher accuracy. A very simple reformulation of FIM is

implemented to consider the difference in sate of information between PCPs and RWPs.

The general structure is defined by the parameters and equations which remain unchanged in all submodels. In the ideal case, all submodels consist of the general structure and differ only in one parameter which is responsible for submodel distinguishability. Such a scenario maximizes submodel identifiability, and is known as piece-wise constant parameter modeling [120]. Also some very interesting studies have been published for the case of hybrid systems [121]. Nonetheless, in order to confront cases with multiple RWPs, and a general structure which does not include all the equations of the submodels, MR is required.

6.3.3 SUBMODEL DISTINGUISHABILITY

Model distinguishability is strongly dependent on the modeler ability to create models which enhance the differences between dynamics of different regimes. At this stage only problems where two candidates are set to discrimination are addressed. In addition, enough data information to assure discrimination with statistical meaning is to be guaranteed. Furthermore, the order of the regimes is to be known and only the precise conditions at which changes take place are unknown.

It is important to remind the reader that fast and accurate MD is key stone for process monitoring. The amount of information required to discriminate between two models depends on the candidates and on data quality. The best approach for model discrimination and its ability to detect the next regime with the required promptness needs to be investigated for each particular case. If the length of the interval selected for MD is too short, false switching point detection may occur due to low identifiability. On the other hand, if the interval is too long the sensitivity of MXL to the data points indicating a regime change is decreased. Finally, it is important to consider the possible estimation of initial values in every new regime.

In probability theory, distinguishability quantifies the probability of selecting the model which best describes the observations. Once the certainty desired to consider the discrimination between two candidates is fixed, the quality of the data set required to distinguish between the model outputs can be calculated. The smaller the difference between models outputs, the more precise need to be the measurements. In other words, to assure distinguishability, the average of the square of the difference between the outputs of the model, need to be at least larger than variance of the measured variables. By using eq. 4.8 in compact form, it is possible to quantify the measurements required to fulfill a specified threshold for model distinguishability D^B.

$$\frac{\sum_{i=1}^{n_y}\left(y_1(u(t),\theta_1)-y_2(u(t),\theta_2)\right)^T * \left(y_1(u(t),\theta_1)-y_2(u(t),\theta_2)\right)}{n_y} = D^B \geq \sigma_y^2 \qquad \textbf{6.6}$$

Where y is a vector with n_y outputs of model 1 and 2 respectively, $u(t)$ a vector of n_u time-dependent input variables, and θ the parameter set of model 1 and 2 respectively.

6.3.4 INITIAL INTERVAL

The initial interval is the process stage with less state of information. Because of the lack of data at the beginning of the process, parameter values of previous runs have to be selected as initial guesses. Luckily, in most of the systems, the first regime of the process is known and cases where the initial conditions of a process are not well defined are scarce in process engineering. Mostly, initial concentrations and state variables as temperature and pressure can be defined at the beginning of the process. In addition offline measurements can be carried out before the process is started to increase the information at the beginning of the process. The application of MR with uncertain initial conditions requires addition of backward propagation calculations to define the initial states and is not considered in this work.

During the initial interval, PCPs need to be defined and the boundaries of the RWPs have to be set. In biological systems for example, maximal reaction rates and uptake capacities determined in the initial interval, define the boundary conditions of the successive parameter values throughout the complete process. Hence, it must be assured, that the initial regime offers enough information to identify the general structure of the submodels. This might affect process efficiency since the initial regime could be artificially kept active only to allow identifiability affecting optimal process conditions. Still, the information gained in the initial interval is essential to assure a correct recognition of the process. Again, the effort of this step is reduced as experience and information are accumulated. Furthermore, many processes show great reproducibility and the parameter values of previous runs can be applied to reduce the required length of the initial interval.

6.3.5 MR INITIALIZATION

The recognition process may only be started after the parameter values of the general structure have been estimated and boundaries, as well as linear and nonlinear constraints have been defined. First, the FIM is tested for identifiability.

$$\det(FIM) \geq 0 \qquad \textbf{6.7}$$

Secondly, once the state of information is enough to assure a trouble-free inversion of the FIM, the computation of the some confidence interval criterion is started.

For both case studies presented in chapters 7 and 8, the A criterion (A_{crit}) was selected to quantify the state of information (section 4.2.1). In short words, A_{crit} represents the average variance of the confidence interval considering all unknown parameters. The boundary for A_{crit} to assure model distinguishability is calculated during the tests for model distinguishability (section 4.3). The value obtained at this stage is used to select the time point from which a regime switching point can be detected.

$$A_{crit} \leq D^B \qquad \textbf{6.8}$$

where D^B represents the criterion boundary obtained from the distinguishability test.

Much care should be taken not to initiate the switching point detection before the state of information has been tested and accepted. Once the criterion boundary has been reached (e.g. the number of measurements is enough to make a distinction between the model outputs), the detection of the switching point may be initiated (section 6.3.6).

It is worth reminding that an important assumption is that the general structure of the model is globally correct and structure parameters do not need to be updated. This assumption is essential to assure consistent evaluation of piecewise constant parameter identifiability and model distinguishability.

6.3.6 DETECTION OF SWITCHING POINTS

The position in time, where changes on process conditions induce a change in the regime of the system, is defined as switching point. The accurate detection of the switching points is essential for two reasons:

- fast and accurate detection of regime change
    - avoid undesired or dangerous process conditions
- accuracy of initial values for the next interval
    - obtain the correct initial conditions of the non measurable variables

Once adequate submodels have been developed and the state of information has been analyzed, detection of the switching points is no different from fault diagnosis based on first principle models. Although not necessarily representing a fault on the system, from a mathematical and physical point of view, the program is required to detect an irregular behavior of the system in comparison to some model predictions.

Methods for model-based residual generation, offering so called analytical redundancy have been widely discussed in literature [122]. Typical methods include, on the one side state estimation techniques (the parity space approach, observer based schemes, and

fault detection filter approach) and parameter estimation techniques on the other (parameter identification). In this work we concentrate on the application of parameter identification methods due to the common theoretical background studied in design of experiments. Still MR is not limited to parameter estimation methods and the application of other faster methods like direct evaluation of the objective function offer a fast and fairly accurate result.

Following conditions must be considered during submodel building since they are essential to assure detectability and distinguishability of the switching point during the process [123]:

- knowledge of the normal behavior of the system
- distinguishability of the changing behavior
- availability of at least one observation reflecting regime change
- satisfactory state of information

The evaluation of the validity of the present regime given by the submodel is divided in two main steps:

- residual generation
- decision of the switching point (time and type of the new interval)

To increase the probability of correct detection of the switching point, the method for robust parameter estimation [123] is proposed. In order to decide whether the process is running in the present regime or a change is taking place, a residual calculation is carried out to select one of the following hypotheses:

$$H_0: \theta_n = \theta_e$$
$$H_1: \theta_n \neq \theta_e \qquad \textbf{6.9}$$

where θ_n and θ_e represent the nominal parameter set and the estimated parameter respectively, hypothesis zero assumes there is no regime change whereas hypothesis one assumes the switching point has been reached.

To enable the evaluation of the hypothesis considering the uncertainties of the observations a residual between θ_n and θ_e is calculated taking the variance-covariance matrix of the expected parameter set into account:

$$res = (\theta_n - \theta_e)^T {C_p}^{-1} (\theta_n - \theta_e) \qquad \textbf{6.10}$$

where res is the residual, and C_p is the variance-covariance matrix of the parameter set.

We can also take advantage of the previous calculation of FIM (section 4.2) and substitute C_p by its approximation, the resulting equation is:

$$res = (\theta_n - \theta_e)^T FIM(\theta_n - \theta_e) \quad \textbf{6.11}$$

Finally, a fixed threshold D^B is selected to accept or reject hypothesis zero. Although variable threshold approaches have shown to be more efficient [124], thanks to the implementation of FIM in the equation, the residual also reflect the actual state of information with respect to the parameters. The variation of the threshold is substituted by the variation of FIM, which gives a more accurate description of parameter covariance in the sense that it also considers information from previous intervals PCPs.

6.3.7 INITIAL CONDITIONS OF THE INTERVAL K+1

It is essential to take into consideration that, in real processes, regime changes do not happen instantaneously. Although in many cases the difference in dynamics allows to consider changes in process conditions as instantaneous, the transition phase between two regimes must be considered. This is of relevance for the calculation of the successive regime (k+1), due to the fact that the initial conditions of the regime (k+1) are determined by the endpoint outputs of the previous regime (k). To increase initial condition accuracy and reduce error bias, offline measurements to redefine the state of the process should be carried out whenever possible. In the cases where no observations are available, the initial values of the next interval need to be estimated. It is clear that this represents an addition of parameters to be estimated reducing the identifiability of the problem drastically and tight boundaries of the initial values are required minimize the impact of these new variables in the complexity of the optimization problem.

To find tight boundaries the switching interval can be simulated with both models to create an envelope by selecting the maximal and minimal values for the upper and lower bounds respectively.

6.3.8 DETECTION OF THE NEXT SWITCHING POINT

Once the best candidate has been selected, the initial values and parameters have been estimated, and the arc in the switching interval determined, the algorithm returns to the search for the next switching point (return to 6.3.6). This procedure is repeated until the end of the process is reached.

6.3.9 FLOW DIAGRAM

Following, the steps of the recognition strategy are presented accompanied by a flow diagram Figure 6.4. The flow diagram proposed has been built for the special case of batch processes, but it can still be applied to different cases with slight modifications.

The diagram is divided in three phases:

- Submodel Building
- Initial Interval
- Regime Recognition

The first phase requires the collaborated work of process and model experts to create an adequate set of submodels (section 6.3.1) and cannot be carried out automatically since there is no general solution to proper submodel building. In this stage the models are tested for identifiability considering the state of information obtained from the process and submodel distinguishability.

Once the process has started, the model describing the first regime is tested for identifiability to determine the minimal length required to perform reliable parameter estimations. The parameter set of the general structure is estimated and the boundaries of the RWPs are set. This is also the most critical phase of MR since the nominal values of the parameter set and the general structure are defined.

Finally, with the general structure well defined and the boundaries set, the detection of the switching point is initiated. The algorithm applies one of the switching point detection criteria (section 6.3.6) to select the time point where the process changes from one regime to another. The third phase is recursively repeated until the end of the process is reached.

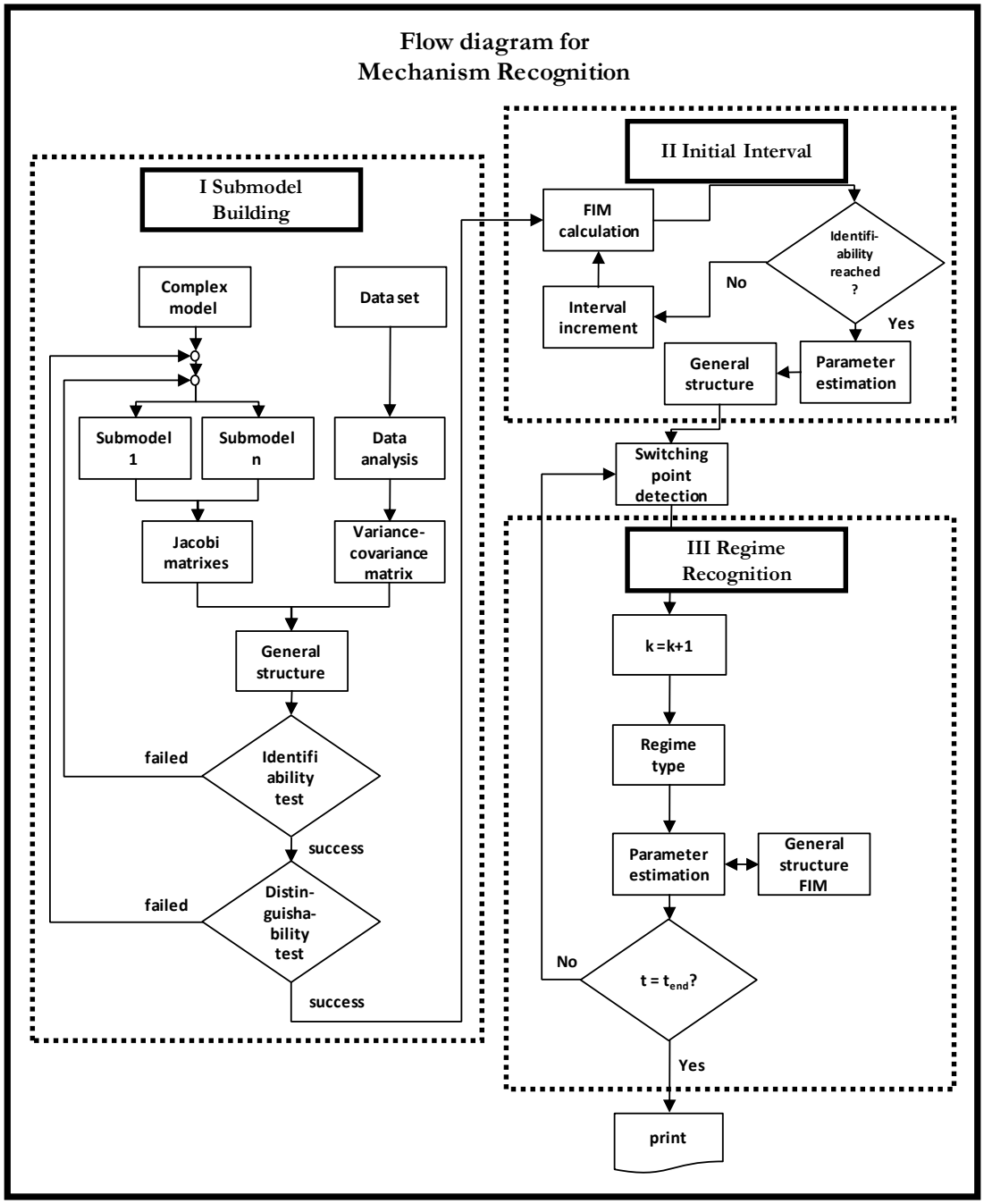

Figure 6.4: Flow diagram of MR algorithm

7 MECHANISM RECOGNITION APPLIED ON SEQUENCING BATCH REACTORS

7.1 INTRODUCTION

7.1.1 ACTIVATED SLUDGE

The most applied method for biological treatment of waste water is the Activated Sludge Process (ASP) [125]. In ASP, carbonaceous organic matter of wastewater (readily biodegradable as well as particulate substrate) provides an energy source for bacterial growth of a mixed population of microorganisms in an aquatic aerobic/anoxic environment. The microbes consume carbon and oxidized end-products that include carbon dioxide and water. In addition, microorganisms may obtain energy by oxidizing ammonia nitrogen to nitrate nitrogen in the process known as nitrification. Moreover, some microorganisms are able to reduce nitrite and nitrate in the anoxic phase as well, this process constituting the second part of the so called nitrification-denitrification process.

The family of Activated Sludge Models (ASM) represents the state-of-the-art model framework for ASP simulation [126]. ASM1 is the most widely used in practice [127], ASM2 [128] is applied to simulate processes that include biological phosphorus removal [78] and the latest version, ASM3 [129], includes the quantification of energy storage in order to describe substrate and oxygen uptake with higher accuracy. Finally, a newer version of ASM3, referred to in this contribution as extended ASM3, where nitrification and denitrification are considered as two-step processes taking into account nitrite as an intermediate, has recently been presented [130].

In order to extend the ASM3, 7 process equations were included, resulting in a stoichiometric matrix with 15 state variables and 20 reaction equations. In addition to a low parameter identifiability, the computation of the extended ASM3 is significantly expensive. These drawbacks represent the main obstacle for efficient optimization and

model-based control. On the other hand, the extended ASM3 describes many states, which are not to be considered in SBR process [131, 132].

7.1.2 SEQUENCING BATCH REACTOR

Continuous Waste Water Treatment plants (WWTP) offer important advantages in comparison to batch processes. The most relevant are; lower energy consumption and operation costs, higher load capacity, and process control stability. Nevertheless, continuous processes confront a major limitation when dealing with an intrinsic characteristic of wastewater treatment processes. Wastewater naturally presents stochastic variations in both time and space. In case of large WWTPs, these variations can be buffered to a certain level by increasing the size of the tanks, or by introducing a flow-stabilization tank upstream. However, small and medium size WWTPs are strongly affected by changing concentrations of contaminants in the wastewater [133]. SBR, because of its high flexibility and operation range, offer an adept solution for such cases [134].

Furthermore, an SBR is able to treat different kinds of effluents such as municipal, domestic, hypersaline, tannery, brewery, dairy wastewaters, and landfill leachates among others [134]. One or more stirred batch reactors with an aeration system are the basis for this kind of process. In an SBR, the retention time, the duration of the aeration and anoxic phases, the settling time, and other conditions can be fitted to a changing quality of load as well as effluent requirements. The SBR can also be considered to be a process which operates with variation in time, whereas a continuous process operates with variations in space [135]. A complete cycle of the sequencing batch process consists of five steps; 1) Idle, 2) Fill, 3) React, 4) Settle and 5) Draw Figure 6.4.

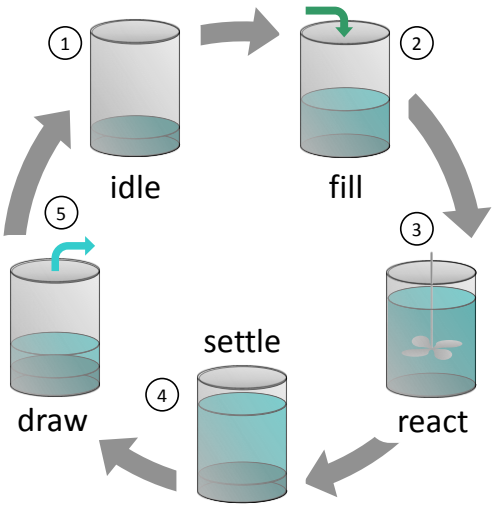

Figure 7.1: SBR cycle [136]

The SBR has gained great popularity in recent years. Advances in process measurement, as well as automation and control, increase the process efficiency, reducing operation costs, while fulfilling the strict environmental regulations [137].

Nowadays, SBR technology has become important in particular for small and medium-sized WWT Plants. When properly designed and operated, an SBR also offers a process with important advantages over continuous processes, not only because of its efficiency and economical aspects [138-140], but also because of its small footprint [135].

Besides many efforts to monitor nutrient removal [78, 141-144] in SBR processes, direct online monitoring of Biological Oxygen Demand (BOD), Chemical Oxygen Demans (COD), Nitrites and Nitrates as well as ammonia is difficult inaccurate and costly. [143]. Indirect methods are required to determine these concentrations for control purposes.

7.1.3 NITRATE BYPASS GENERATION

In the ASP, nitrogen is removed from wastewater by the nitrification/denitrification process. Most of the nitrogen contained in wastewater is converted into ammonia. Ammonia is then converted into molecular nitrogen by a two-step biological processes, namely nitrification followed by denitrification (Figure 7.2).

$$NH_4^+ \rightarrow NO_2^- \rightarrow NO_3^-$$

$$N_2 \leftarrow NO_2^- \leftarrow NO_3^-$$

Figure 7.2. Nitrification-denitrification process described as a two -step reaction.

In the first stage, *Nitrosomonas* and other ammonia oxidizers convert ammonia and ammonium to nitrite, whereas in a second stage, *Nitrobacter* and other nitrite oxidizers finish the conversion of nitrite to nitrate.

Turk and Mavinic [145] proposed the Nitrate Bypass Nitrification-Denitrification (NBND) process, which can be achieved by inhibiting the production of nitrate and proposed various methods for bringing about this effect. Katsogiannis et al [146] showed that a frequent enough change between aerobic and anoxic conditions suppresses nitrate formation. Ammonia is converted to nitrite in the presence of oxygen (nitritation), which is then converted into nitrogen under anoxic conditions before the second oxidation producing nitrate (nitratation) can take place. The NBND has the following advantages over conventional nitrification-denitrification [145]:

- 40% reduction of COD demand during denitrification
- 63% higher rate of denitrification
- 300% lower biomass yield during anaerobic growth
- no apparent nitrite toxicity effects for the microorganisms in the reactor

7.1.4 MONITORING OF WASTEWATER PROCESSES

Online monitoring of WWTPs is a challenging task. Variables measured in standard WWTPs are pH, DO, respirometry, redox capacity and titrimetry, among others [147, 148]. Despite the important advances achieved in recent years, modern measure systems still require laborious maintenance, frequent recalibration and important know how. In addition, environmental regulations demand better process and output quality control at low treatment cost and energy efficient process conditions. The implementation of new approaches to enable an accurate online monitoring of key species like organic matter is of great relevance in modern water treatment technology. Model based control and software sensors offer interesting possibilities for the maximization of process efficiency.

The Activated Sludge Model No. 3 (ASM3) extended for two-step nitrification and denitrification has proven to be a very accurate model to describe ASP [149]. The division of the nitrification-denitrification reaction in a two-step reaction is essential when trying to describe the bypassing nitrate generation process in the ASP [146]. Moreover, the extended ASM3 enables the description of the substrate consumption and the oxygen uptake with a higher precision than the older versions of the ASM family because of the addition of energy storage effects. This model extension facilitates the calculation of the NO_2 concentration as an independent variable.

In this section, diverse approaches to model reduction (chapter 3) are applied in order to develop submodels of the ASM3 while maintaining their characteristic dynamics in each regime. This study is based on the analysis of the process conditions and the available empirical knowledge. The first submodel proposed is named 5state, according to the number of the ordinary differential equations which need to be solved.

The model reduction is based on the principle that an batch cycle should stop once all concentrations comply with the regulations. Furthermore, these concentrations should be reached in the shortest time possible. Consequently, lower output concentrations than required are an indicator of a suboptimal operation and its accurate detection increases process efficiency. Another important assumption is that the bacteria never exhaust their stored energy. Except for the recycle process, the bacteria are always in a medium, which is rich in substrate. Therefore, the stored energy value should be permanently high during the process and never limit bacterial growth.

7.2 SUBMODEL BUILDING

Following the systematic approach to MR, the submodels are created considering process conditions and the characteristics of the complex model to enable a fast

description of the different regimes of the process Figure 6.4. The recognition step should identify two distinct regimes:

1. Regime with readily biodegradable substrate in the medium
2. Regime where organic matter has been depleted

In SBR processes, it is of interest to stop the process once the concentrations of the medium fulfill environmental regulations. An accurate detection of depletion of organic matter offers important improvements in process efficiency. For this reason, the submodel is built to describe the process regime under high substrate concentrations aiming at detection of substrate depletion.

7.3 A PROPOSED 9STATE MODEL

In the first reduction step, the model proposed is named 9state, according to the number of state variables contained in the equation system. Thus there are nine differential equations which describe the basic variables (concentrations), namely: 1.- carbonaceous substrate, 2.- heterotrophic bacteria, 3.- ammonia oxidizers, 4.- nitrite oxidizers, 5.- dissolved oxygen, 6.- ammonia, 7.- nitrite, 8.- nitrate and 9.- stored substrate.

7.3.1 STORAGE

The implementation of energy storage represents the principal improvement of ASM3 in comparison to older versions. Neglecting this equation would impede a proper process description and would result in an incorrect model reduction. Consequently, the storage of substrate and its effects on substrate and oxygen concentration cannot be ignored. For this reason, the proposed 9state model includes some adaptations to the substrate and oxygen uptake equations. The new set of equations is presented in such a way, that both substrate uptake and oxygen uptake increments caused by the storage are now included in the original substrate and oxygen differential equations.

By this means, the relation between both, substrate consumption rate and oxygen consumption rates, to energy storage are linear, which is accurate as long as the substrate concentration is above zero. This could appear to be an inconsistent assumption, though it is almost certain that the process continues after substrate elimination in order to achieve further ammonia degradation. Nevertheless, previous model versions (ASM1 and ASM2) fit the data although they lack a storage variable. In other words, the 9state model responds to the substrate limitation similarly to ASM1, but describes substrate and oxygen uptake as well as energy storage as precisely as the extended ASM3.

ASM3 considers that bacteria store some of the substrate they consume under high substrate concentrations for its later use under substrate limiting conditions. This storage is not only responsible for lower biomass growth under equal substrate consumption, but also for the bacterial growth when no substrate is present in the medium. In the case of an SBR process, it is considered that storage of substrate does not limit bacterial growth. The assumption is based on the fact that in a process with optimal aeration strategy, bacteria never exhaust their stored energy for two reasons:

- In order to minimize process time and costs, the environmental regulations should be fulfilled as soon as substrate is consumed.
- Excepting the idle phase, bacteria are always in a medium rich in substrate. Therefore, the value of the stored energy should be high at any moment during the process and never limit bacterial growth.

If these assumptions are valid, the relation between the substrate used for storage and the substrate used for growth is valid as well. In the extended ASM3, the relation between substrate and biomass growth is described by a second order differential equation. However, as long as the concentration of the stored energy is high, this second-order differential equation can be accurately approximated with a first order differential equation.

7.3.2 REDUCTION OF THE EXTENDED ASM3 MODEL TO A 9STATE MODEL

In this work, the 15 ordinary differential equations of the extended ASM3 include only the process rate variables without their explicit equation. The process rate equations are shown in Table 7.1and have been numbered in the same order as previously presented in [130]:

Table 7.1: Reaction rates of the extended ASM3

Heterotrophic Organisms:
r_1: Hydrolysis
r_2: Aerobic Storage
r_3: Anoxic Storage
r_4: Anoxic Storage of carbonate substrate, NO_2–N_2
r_5: Aerobic Growth of Heterotrophic bacteria
r_6: Anoxic Growth NO_3-NO_2
r_7: Anoxic Growth NO_2-N_2
r_8: Aerobic Endog. Resp. of Heterotrophic bacteria
r_9: Anoxic Endog. Resp. NO_3-NO_2
r_{10}: Anoxic End. Resp. NO_2-N_2
r_{11}: Aerobic Resp. of particulate storage
r_{12}: Anoxic Resp. of particulate storage, NO_3-NO_2
r_{13}: Anoxic Resp. of particulate storage, NO_2-N_2

Ammonium Oxidizing Bacteria (AOB):
r_{14}: Aerobic Growth, Nitritation
r_{15}: Aerobic End. Resp.
r_{16}: Anoxic End. Resp.

Nitrite Oxidizing Bacteria (NOB):
r_{17}: Aerobic Growth, Nitratation
r_{18}: Aerobic End. Resp.
r_{19}: Anoxic Endog. Resp

7.3.3 MATHEMATICAL REPRESENTATION OF THE 9STATE MODEL

ORDINARY DIFFERENTIAL EQUATIONS

The 9 ordinary differential equations are shown in equations 7.1 - 7.9. Their corresponding rate equations are described in 7.10 - 7.14. All the constants K.x have the same value as published in the extended version of ASM3 [130]. The values used for the saturation constants are shown in Table 7.2.

$$\frac{dS_S}{dt} = \left(-\frac{1}{Y_{Haer}} * r_{aae} - \frac{1}{Y_{Hanox}} * (r_{aNO3} + r_{aNO2})\right) * (1 + St_S) \quad 7.1$$

$$\frac{dX_H}{dt} = r_{aae} + r_{aNO3} + r_{aNO2} \quad 7.2$$

$$\frac{dX_{Ns}}{dt} = r_{aaNs} \quad 7.3$$

$$\frac{dX_{Nb}}{dt} = r_{aaNb} \quad 7.4$$

$$\frac{dS_O}{dt} = \quad 7.5$$

$$\frac{dS_{NH4}}{dt} = -\left(-\frac{i_{NSS}}{Y_{Haer}} + i_{NB}\right) r_{aae} - \left(\frac{1}{Y_{A1}} + i_{NB}\right) r_{aaNs} - i_{NB} r_{aaNb} - \left(-\frac{i_{NSS}}{Y_{Hanox}} + i_{NB}\right) r_{aNO3} - \left(-\frac{i_{NSS}}{Y_{Hanox}} + i_{NB}\right) r_{aNO2} \quad 7.6$$

$$\frac{dS_{NO2}}{dt} = \frac{1}{Y_{A1}} r_{aaNs} - \frac{1}{Y_{A2}} r_{aaNb} + \frac{1 - Y_{Hanox}}{1.14 Y_{Hanox}} (r_{aNO3} - r_{aNO2}) \quad 7.7$$

$$\frac{dS_{NO3}}{dt} = \frac{1}{Y_{A3}} r_{aaNb} - \frac{1 - Y_{Hanox}}{1.14 Y_{Hanox}} r_{aNO3} \quad 7.8$$

$$\frac{dX_{Sto}}{dt} = \left(-\frac{1}{Y_{Haer}} * r_{aae} - \frac{1}{Y_{Hanox}} * (r_{aNO3} + r_{aNO2})\right) * (-St_S) \quad 7.9$$

S_S	Readily biodegradable substrate
X_H	Biomass of heterotrophic bacteria
X_{Ns}	Biomass of Nitrosomones
X_{Nb}	Biomass of Nitrobacter
S_O	Oxygen concentration in the medium

S_{NH4}	Ammonia concentration
S_{NO2}	Nitrite concentration
S_{NO3}	Nitrate concentration
X_{Sto}	Energy Storage
Y_{Haer}	Yield coefficient of S_S to X_H in aerobic conditions
Y_{Hanox}	Yield coefficient of S_S to X_H in anoxic conditions
Y_{A1}	Yield coefficient nitrite to nitrosomones in aerobic conditions
Y_{A2}	Yield coefficient nitrite to nitrobacter in aerobic conditions
Y_{A3}	Yield coefficient nitrate to nitrobacter in aerobic conditions
μ_H	Maximal growth of Heterotrophous
μ_{A1}	Maximal growth of nitrosomones
μ_{A2}	Maximal growth of nitrobacter
μ_{H1}	Maximal growth of heterotrophic bacteria on nitrire
μ_{H2}	Maximal growth of heterotrophic bacteria on nitrate
St_S	Substrate to storage constant
St_O	Oxygen to storage
K	Limitation constant

REACTION RATES

Heterotrophic growth on aerobic conditions

$$r_{aae} = \mu_H * \frac{S_S}{S_S + K_S} * \frac{S_O}{S_O + K_{O1}} * \frac{S_{NH4}}{S_{NH4} + K_{NH}} * X_H \quad \textbf{7.10}$$

Growth of nitrosomones

$$r_{aaNs} = \mu_{A1} * \frac{S_O}{S_O + K_O} * \frac{S_{NH4}}{S_{NH4} + K_{NH}} * X_{Ns} \quad \textbf{7.11}$$

Growth of nitrobacter

$$r_{aaNb} = \mu_{A2} * \frac{S_{NO2}}{S_{NO2} + K_{NO21}} * \frac{S_O}{S_O + K_O} * \frac{S_{NH4}}{S_{NH4} + K_{NH}} * X_{Nb} \quad \textbf{7.12}$$

Heterotrophic growth on nitrate

$$r_{aNO3} = \mu_{H1} * \frac{S_S}{S_S + K_S} * \frac{S_{NO3}}{S_{NO3} + K_{NO3}} * \frac{K_{O21}}{K_{O21} + S_O} * \frac{S_{NH4}}{S_{NH4} + K_{NH}} * X_H \quad \textbf{7.13}$$

Heterotrophic growth on nitrite

$$r_{aNO2} = \mu_{H2} * \frac{S_S}{S_S + K_S} * \frac{S_{NO2}}{S_{NO2} + K_{NO2}} * \frac{K_{O22}}{K_{O22} + S_O} * \frac{S_{NH4}}{S_{NH4} + K_{NH}} * X_H \quad \textbf{7.14}$$

Table 7.2: 9state model constants and its values as shown in the Matlab code

SOStar	= 7;	[mgO2/l]	process	K_NH2	= 0.1;	[mgCOD/l]	ASM3
K_La	= 1000;	[d-1]	process	K_S	= 10;	[mgCOD/l]	ASM3
i_NB	= 0.086;	[gN/gCOD]	fitted	K_S1	= 0.1;	[mgCOD/l]	ASM3
mou_H	= 0.6021;	[d-1]	fitted	K_S2	= 0.1;	[mgCOD/l]	ASM3
mou_A1	= 0.6552;	[d-1]	fitted	K_NHH	= 0.05;	[mgN/l]	ASM3
mou_A2	= 0.3468;	[d-1]	fitted	K_O1	= 0.2;	[mgO2/l]	ASM3
Y_Haer	= 0.1302;	[gCOD/gCOD]	fitted	K_NH	= 0.1;	[mgN/l]	ASM3
Y_A1	= 0.1327;	[gCOD/gN]	fitted	K_O	= 0.8;	[mgO2/l]	ASM3
Y_A2	= 0.0985;	[gCOD/gN]	fitted	K_NO21	= 0.5;	[mgO2/l]	ASM3
Y_A3	= 0.0331;	[gCOD/gN]	fitted	K_NO3	= 0.5;	[mgN/l]	ASM3
i_NSS	= 0.01;	[gN/gCOD]	ASM3	K_O21	= 0.2;	[mgO2/l]	ASM3
Y_Hanox	= 0.0632;	[gCOD/gCOD]	fitted	K_NO2	= 0.25;	[mgN/l]	ASM3
mou_H1	= 0.0511;	[d-1]	fitted	K_O22	= 0.2;	[mgO2/l]	ASM3
mou_H2	= 0.0362;	[d-1]	fitted	stS	= 1.7;	[ ]	fitted
K_NH1	= 0.01;	[mgCOD/l]	ASM3	stO	= 0.08;	[ ]	fitted

7.3.4 STOICHIOMETRIC MATRIX

The stoichiometric matrix of the 9state model is presented in Table 7.3. This matrix represents the $(l + m - n)$ matrix of equations. In other words, the analysis of the stoichiometric matrix gives us the information about the number of reaction invariants hidden in the 9state model.

Table 7.3: Stoichiometric matrix of the 9state model

	S_S	X_H	X_{NS}	X_{NB}	S_O	S_{NH4}	S_{NO2}	S_{NO3}	X_{Sto}
r_{aae}	$-\frac{1+\mathrm{St_S}}{Y_{Haer}}$	1	0	0	$-\frac{1-Y_{Haer}}{Y_{Haer}}$	$\frac{i_{NSS}}{Y_{Haer}} - i_{NB}$	0	0	$\frac{\mathrm{St_S}}{Y_{Haer}}$
r_{aaNs}	0	0	1	0	$1-\frac{3.34}{Y_{A1}}$	$-\left(\frac{1}{Y_{A1}} + i_{NB}\right)$	$\frac{1}{Y_{A1}}$	0	0
r_{aaNb}	0	0	0	1	$1-\frac{1.14}{Y_{A2}}$	$-i_{NB}$	$-\frac{1}{Y_{A2}}$	$\frac{1}{Y_{A3}}$	0
r_{aNO2}	$-\frac{1+\mathrm{St_S}}{Y_{Hanox}}$	1	0	0	0	$\frac{i_{NSS}}{Y_{Hanox}} - i_{NB}$	$-\frac{1-Y_{Hanox}}{1.14Y_{Hanox}}$	0	$\frac{\mathrm{St_S}}{Y_{Hanox}}$
r_{aNO3}	$-\frac{1+\mathrm{St_S}}{Y_{Hanox}}$	1	0	0	0	$\frac{i_{NSS}}{Y_{Hanox}} - i_{NB}$	$\frac{1-Y_{Hanox}}{1.14Y_{Hanox}}$	$-\frac{1-Y_{Hanox}}{1.14Y_{Hanox}}$	$\frac{\mathrm{St_S}}{Y_{Hanox}}$

7.3.5 LIMITATIONS OF THE REDUCED MODELS

It should be noted that the proposed 9state model is not valid for the whole range of conditions as the extended ASM3. Some limitations are to put up with in order to reduce the model and speed up the simulation in the region of interest. Moreover, because of the new storage equation, the bacteria can store energy, but cannot use it when there is no more substrate available. This feature converts the 9state model in a perfect indicator of substrate depletion in the reactor.

In the extended ASM3, the ammonia concentration does not limit the energy storage. This results in a consumption of substrate even under ammonia limitation. Once again, because of the coupled equations, the 9tate model predicts substrate consumption only as long as ammonia is present in the medium. Finally, the growth of heterotrophic biomass can be mathematically described as a second order differential equation. For this reason, if the energy stored by the bacteria is low, a time delay can be seen in the growth curve. This time delay is not predicted by the 9state model. Taken into consideration that the storage has a value at least larger than 100 gCOD/m^3, both growth curves match.

7.4 A PROPOSED 6STATE MODEL

The 9state model involves 9 differential equations, with only one of them depending on the process input (oxygen supply). The reaction invariant theory can then be applied to the 8 differential equations, which are not affected by the oxygen input, i.e. to 7.1 - 7.4 and 7.6 - 7.9.

These equations are of the form of eq. 2.6 with $n = 5,\ l + m = 8$. Therefore, it is possible in this case, to find $(l + m) - n = 3$ linearly independent reaction invariants. For example 7.15 - 7.17,

$$S_{NH4} + i_{NSS} * \frac{S_S}{1 + St_S} + i_{NB} * (X_H + X_{NS} + X_{NB}) + \frac{X_{NS}}{Y_{A1}} \tag{7.15}$$

$$S_{NO2} + 2 * S_{NO3} + \frac{1 - Y_{Hanox}}{1.14} * \frac{Y_{Haer}}{Y_{Hanox} - Y_{Haer}} * \left(\frac{S_S}{1 + St_S} + \frac{X_H}{Y_{Haer}}\right) - \frac{X_{NS}}{Y_{A1}} - \frac{2 * Y_{A2} - Y_{A3}}{Y_{A2} * Y_{A3} * X_{NB}} \tag{7.16}$$

$$\frac{S_S}{1 + St_S} + \frac{St_O}{St_S} \tag{7.17}$$

The above quantities remain constant throughout the entire batch. Therefore, it is possible to substitute three differential equations 7.6, 7.7 and 7.9 for three algebraic ones 7.18 - 7.20.

$$S_{NH4} = C_{NH4} - \left(i_{NSS} * \frac{S_S}{1 + St_S} + i_{NB} * (X_H + X_{NS} + X_{NB}) + \frac{X_{NS}}{Y_{A1}}\right) \quad \textbf{7.18}$$

$$S_{NO2} = C_{NO2} - 2 * S_{NO3} - \frac{1 - Y_{Hanox}}{1.14} * \frac{Y_{Haer}}{Y_{Hanox} - Y_{Haer}} * \left(\frac{S_S}{1 + St_S} + \frac{X_H}{Y_{Haer}}\right) + \frac{X_{NS}}{Y_{A1}} - \frac{2 * Y_{A2} - Y_{A3}}{Y_{A2} * Y_{A3} * X_{NB}} \quad \textbf{7.19}$$

$$X_{Sto} = C_{XSto} - \frac{S_S * St_S}{1 + St_S} \quad \textbf{7.20}$$

where C_{NH4}, C_{NO2} and C_{XSto} are the respective constants obtained after solving equations 7.15 - 7.17 for the initial conditions. Based on the method of reaction invariants, we find three equations which are linearly dependent. Therefore, the 6state model mimics the 9state model perfectly well.

7.5 A PROPOSED 5STATE MODEL

Taking a closer look at the growth of the biological matter during one SBR cycle, we can see that the overall change in biomass does not exceed 10%. Based on this observation a further reduction of the mode is possible for the special case of a batch process. We can eliminate these three differential equations 7.2 - 7.4 from our 9state model and consider a constant biomass concentration throughout each cycle of the SBR so as to obtain a model with 6 differential equations. Again it can be shown that this system has one reaction invariant which is the same as indicated for the case of the 9state 7.20. As a result we finally obtain a model with only 5 state equations. This 5state submodel is almost as accurate as the 6state model. However, it can only be applied for batch processes where the change of biomass can be neglected.

7.6 RESULTS

The submodel was set to various conditions so as to confirm their stability and accuracy. The most representative results obtained after the comparison of ASM3 with the 5state model are presented in Figure 7.3-Figure 7.6.

The simulation represents a batch tank ideally mixed, where the aeration can be turned on and off in order to induce either aerobic or anoxic conditions. The initial value of the storage is set to 400 gCOD/m^3. This assumption is based on the fact that the sludge is obtained from previous cycles and bacteria have already stored energy.

The system of DAE is solved with two different integrators:

- sDACL: An in house tool developed to solve DAE systems with OCFE able to also calculate dynamic sensitivities [46].
- ODE15s: An integrator based on the numerical differentiation formulas (NDFs) [150].

The main purpose is to set both models to drastic changes and various limitations. The aeration is turned on and off intermittently to produce a strongly dynamic process. As a result, a constantly changing process is obtained, which makes it very difficult to be described identically by two models with different characteristics. The process conditions prove that the 5state model describes accurately the limitations of dissolved O_2, NO_2^- and NO_3^-

Figure 7.3 shows the results for substrate and storage. The initial value of storage was set to 400 gCOD/m^3 as explained in section 4.2. All reduced versions describe perfectly the behavior of storage even doe it is calculated by an algebraic equation. This proves that the substitution of the differential equation for energy storage by an algebraic equation does not affect the dynamics of the model.

As seen in Figure 7.4, considering constant biomass values throughout the process does not affect the results of the other state variables. Even though growth rate is a crucial parameter of the process biomass concentration, changes in batch processes can be neglected.

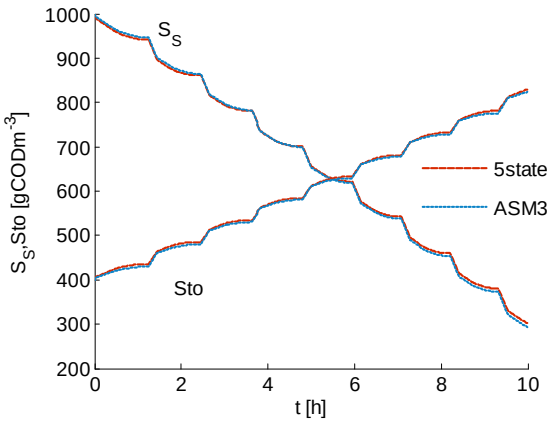

Figure 7.3. Substrate concentration *Ss* and stored energy *Sto* against time.

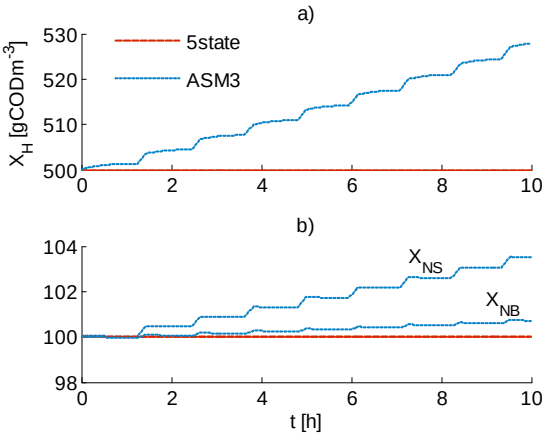

Figure 7.4. Biomass against time. Changes in the biomass are very small (less than 10%).

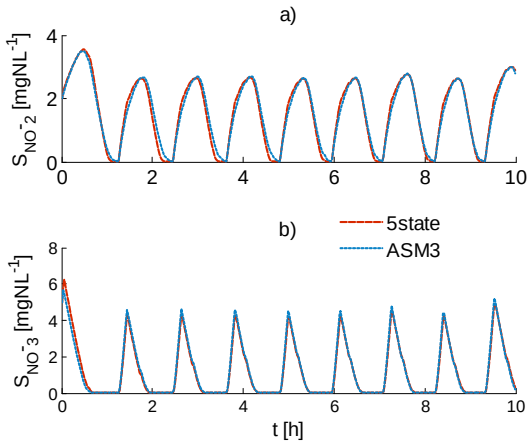

Figure 7.5. NO_X concentration against time.

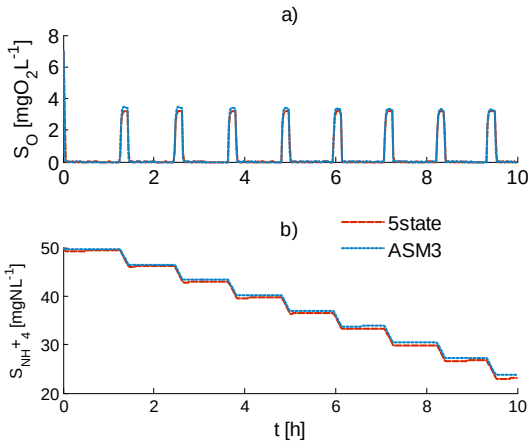

Figure 7.6. a) Oxygen concentration in the medium against time.

b) Ammonia concentration against time.

The concentration of NO_2 and NO_3 are kept under 20 mgN/L. The control variable is the aeration of the tank. Figure 7.5 shows how precise the reduced model describes the curves of nitrite and nitrate even in such drastic conditions. Finally, the results for oxygen and ammonia calculations can be seen in Figure 7.6.

7.6.1 SIMULATIONS RESULTS

The most representative results of the comparison of both models simulated in Matlab® are presented in Table 7.4 and Table 7.5.

Table 7.4. Comparison of the computation time.

	ODE15s CPU time (sec)			
Number of Aer-Anox phases	ASM3	9State	5State	ASM3/5State
1	1.769	0.157	0.156	11.3
2	2.162	0.172	0.14	15.4
3	2.602	0.172	0.172	15.1
4	2.583	0.172	0.156	16.6
5	2.608	0.156	0.172	15.2

The reduced models are up to one order of magnitude faster. For the calculation of the dynamic sensitivities, the Jacobian matrixes have to be solved together with the equation system. Table 7.5 shows the difference of calculation time for the three models.

Table 7.5. Singular function evaluations speed

Evaluation of Jacobians	ASM3	2071.114
	8State	565.364
CPU time 100 evaluations (sec)	5State	321.6
Calculation of the differential eq. system	ASM3	2.225
	8State	0.157
CPU time 1000 evaluations (sec)	5State	0.132

7.7 MECHANISM RECOGNITION IN SBR PROCESSES

We now analyze the capacity of the 5state model to act as an indicator of substrate concentration in the medium. As depicted in Figure 7.7, the 5state model is incapable to describe the behavior of the process after substrate has been depleted. This characteristic of the 5state model is to be expected, since the assumption made for its reduction clearly state the condition of substrate present during the process. The 5state model can be applied as an indirect method to measure substrate concentration.

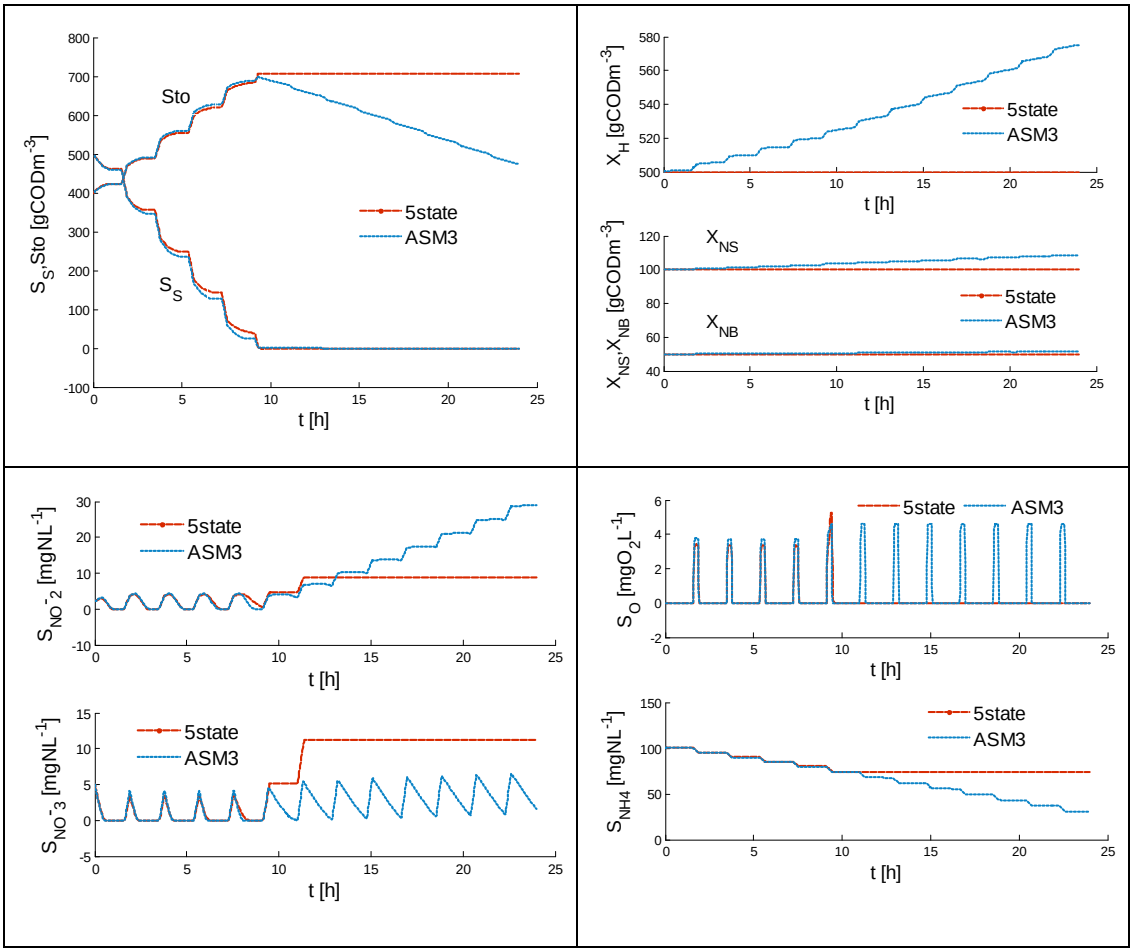

Figure 7.7: Description of the 5state model in both regimes, with and without substrate.

Because of the qualities of the submodel, simulation inaccuracies can be overcome by the switching point detection step. Even for the case where the dynamics of the substrate concentration is not properly predicted, the program can distinguish if the failure on description capacity is due to a substrate mismatch or due to substrate depletion.

7.8 RECOGNITION OF ORGANIC MATTER DEPLETION

7.8.1 CONDITIONS FOR PROPER PROCESS DESCRIPTION WITH MECHANISM RECOGNITION

To ensure successful recognition of the different regimes during the process, the system is required to fulfill conditions stated in section 6.2:

DYNAMIC PROCESS

- The SBR process is characterized by changes in the state of the system throughout the complete run.

PROCESS HAS TO HAVE MORE THAN ONE REGIME

- As previously discussed, two different regimes are considered:
  - biomass growth based on consumption of organic matter.
  - biomass growth under organic matter limitation.

DETAILED MODEL OF THE GENERAL PROCESS AVAILABLE OF THE FORM:

- DAE system index zero or one
- no discontinuities
- known initial conditions

The extended ASM3 fulfils all the above mentioned conditions.

PARAMETER BOUNDARIES HAVE TO ASSURE GLOBAL CONVERGENCE WITH GRADIENT BASED OPTIMIZERS

- The parameter set was tested for global optimality with stochastic optimization based on PSO [60]. Although global optimality cannot be assured, the parameters of the optimizer were set to maximize the probability of global convergence.

INITIAL REGIME IS KNOWN

- Because of the high concentration of organic matter at the beginning of the cycle, it can be guaranteed that the initial regime of the process is biomass growth based on consumption of organic matter.

THE SEQUENCE OF THE REGIMES IS KNOWN.

- Since only two regimes are considered in this process, the sequence is known.

THE MINIMAL LENGTH OF EACH TIME REGIME ASSURES MODEL DISTINGUISHABILITY

- Based on the variance of the data, the form of the general structure and the conditions of the process, it can be assured that the period of biomass growth on consumption of organic matter is long enough to guarantee identifiability and detection of the switching point.

7.8.2 CONDITIONS FOR ACCURATE SWITCHING POINT DETECTION

The assumptions mentioned above assure a successful model description of the process using rigorous submodels with low systematic error. In addition, to allow a proper detection of the switching points in the process, following conditions are to be considered (section 6.3.6):

KNOWLEDGE OF THE NORMAL BEHAVIOR OF THE SYSTEM

- The behavior of the bacteria in each regime is described by the submodels. In addition, the physical foundation of the extended ASM3 offers a basis for evaluation of the dynamics of the process.

DEFINITIVENESS OF THE CHANGING BEHAVIOR

- The distinguishability test proves that, at least for the simulated experiments, the submodels are distinguishable. From this it may be supposed, that the process offers enough amount of information to detect the difference in bacterial behavior under both regimes.

AVAILABILITY OF AT LEAST ONE OBSERVATION REFLECTING REGIME CHANGE

- The combination of measuring variables considered, allows an accurate detection of the change of regime.

SATISFACTORY STATE OF INFORMATION

- The variance and frequency of the simulated experimental data provides the state of information of the system required for MR.

MEASURED VARIABLES

MR was applied considering the following measured variables:

- Ammonia concentration [$mgNL^{-1}$]
- Dissolved Oxygen in the medium [mgO_2L^{-1}]
- Nitrite concentration [$mgNL^{-1}$]
- Nitrate concentration [$mgNL^{-1}$]

GENERAL STRUCTURE

Cells utilize the consumed substrate for a number of functions as are growth, energy production or storage. The reaction pathways responsible for substrate consumption and processing depend on many conditions. From this it can be deduced that the yields for substrate consumption achieved by bacteria vary depending on various factors. Still, the extended ASM3 and consequently in the submodels, consider constant yield coefficients. Taking advantage of MR the yield coefficients are considered to be RWPs. By these means, the program can adapt the yield coefficient in each regime.

The PCPs are:

μ_H	Maximal growth of Heterotrophous
μ_{A1}	Maximal growth of nitrosomones
μ_{A2}	Maximal growth of nitrobacter

The RWP are:

Y_{Haer}	Yield coefficient of S_S to X_H in aerobic conditions
Y_{A1}	Yield coefficient nitrite to nitrosomones in aerobic conditions
Y_{A2}	Yield coefficient nitrite to nitrobacter in aerobic conditions
Y_{A3}	Yield coefficient nitrate to nitrobacter in aerobic conditions

INITIAL INTERVAL

A high concentration of organic matter in the feed water to the tank is guaranteed assuring growth on organic matter as the initial interval.

7.8.3 MR INITIALIZATION

The threshold D^B set at 10 for the A criterion of the general structure is reached after 0.30 days (7.2 hours). From this point on, MR is initiated to detect depletion of organic matter.

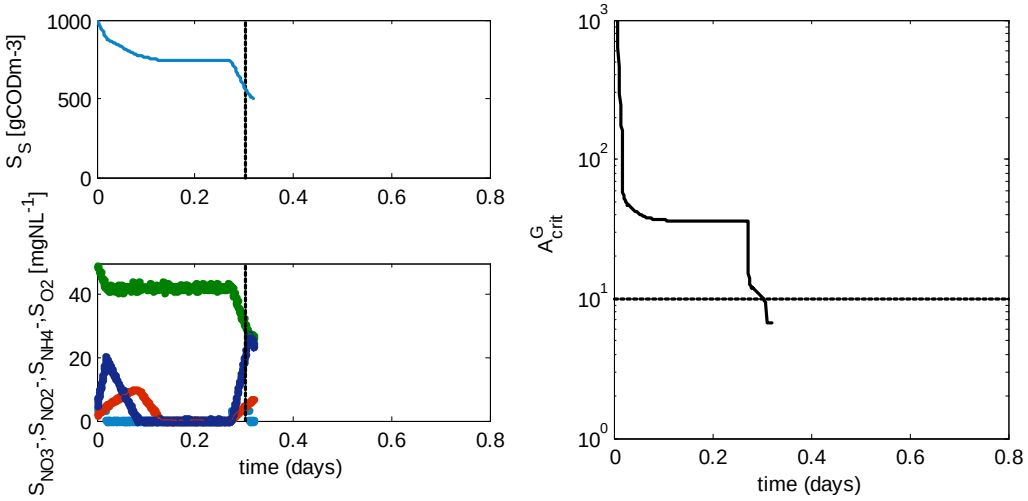

Figure 7.8: Minimal length for initialization of MR

The scenario selected to show the performance of MR is depicted in Figure 7.9 and includes three short aerobic phases as well as three large anoxic intervals. This scenario was selected in order to test MR in a process with very low state of information. The anoxic intervals offer measurements with low information content since 1) the parameters set to estimation control mostly the aerobic phase and 2) the concentrations of oxygen and Nitroxides drops to zero. Still MR shows an incredibly good performance in terms of both accuracy and robustness.

7.8.4 DETECTION OF SWITCHING POINTS

Detection of the switch between overflow and substrate limitation regime was carried out controlling the value of the objective function. The constant threshold was set equal to a MXL value of 0.35. Substrate depletion is accurately detected despite the impossibility to measure the concentration of organic matter online. As it can be seen in Figure 7.9, MR detects the point with high precision. Furthermore, the large slope of *H0* at the depletion region assures a robust and accurate detection of the regime change.

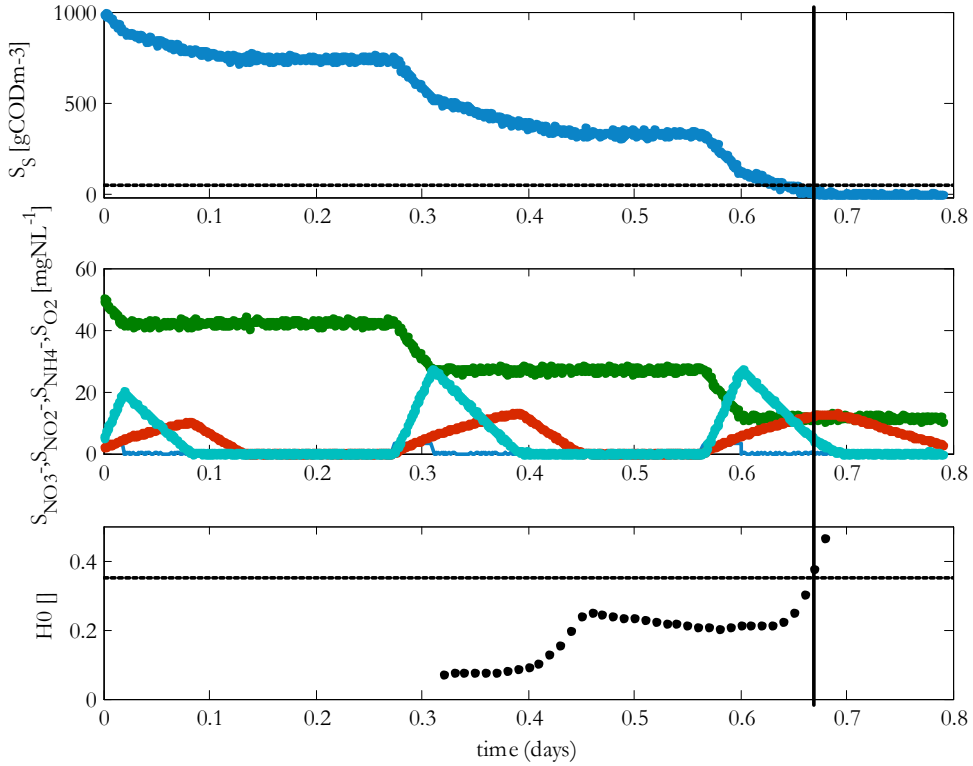

Figure 7.9. Detection of the regime switching point.

SWITCHING INTERVAL

The experimental results confirmed that considering an instant switch of regime allows an accurate description of the process.

7.9 CONCLUSIONS

An assiduous analysis of the process conditions, empirical knowledge, and different approaches to model reduction have been applied to develop a submodel of the extended ASM3. By these means, it was possible to create a model that is significantly simpler and delivers more information about the process. The implementation of MR for description of ASP in SBR processes clearly states the advantages of this approach. The resulting 5state submodel mimics the behavior of the extended ASM3 for SBR process during growth based on organic matter with high accuracy. The results suggest that the 5state model can be applied for the simulation of the nitrate bypass reaction in SBR processes, and thus, for model-based control and online optimization. Moreover, the extended ASM3 was successfully simplified, reducing its calculation costs up to one order of magnitude.

Most important, the limitation of the submodel to describe the process after depletion of organic matter can be exploited. In other words, the reduction of the extended ASM3 creates a model which can indicate the time point where organic matter is present in the reactor. The 5state model works as an indicator of the non measurable state variable readily biodegradable organic matter. The program is able to detect the time point when no more carbonate matter is present in the reactor increasing process efficiency.

8 MECHANISM RECOGNITION IN *ESCHERICHIA COLI* CULTIVATIONS

8.1 *ESCHERICHIA COLI* CULTIVATIONS

Many reports about process optimization of *Escherichia coli* cultivation are describing problems caused by acetate synthesis, which can retard growth and protein production [151-153]. In *E. coli*, acetate synthesis and excretion occurs during the so called 'overflow metabolism' [154] when substrate concentration (namely glucose) is available in large amounts. The threshold value depends on the strain and culture conditions. Furthermore, acetate synthesis already occurs under some conditions at weak excess. The carbon flow through acetyl-CoA is partly shifted towards acetate production instead of entering into the tricarboxylic acid cycle [155]. Acetate is synthesized via phosphotransacetylation. The product acetyl-phosphate is then converted to acetate by the acetate kinase enzyme. Also a direct conversion from pyruvate to acetate with the pyruvate oxidase is possible [156, 157]. Acetate itself can be reconverted to acetyl-CoA either via acetyl-phosphate or by acetyl-CoA synthetase (Acs) via acetyl-AMP. The latter reaction complex is characterized by a 50-fold higher affinity and therefore responsible for acetate reconversion at low concentrations [158]. It could be demonstrated in accelerostat studies, that the first step of overflow mechanism in *E. coli* is the downshift of the ACS system leading to a lower conversion of acetate via acetyl-AMP. However, the reconversion of acetate is regarded as being essential for chemotaxis, proteolysis and pathogenesis [159].

Several attempts have been made to circumvent acetate accumulation. Pulsed feeding strategies based on dissolved oxygen measurements reduced acetate formation increasing growth and product concentrations [160, 161].

The behavior of *E. coli* at substrate excess (overflow metabolism) when acetic acid is formed as carbon storage, the activation of the glyoxylatic shunt, and the response to oxygen limitation, are in the focus of recent development. While more and more regulatory mechanisms are considered in models, the increasing number of equations

and parameters increment model complexity. In addition, since the appropriate detection of intermediates is crucial due to their naturally low levels in bacteria, many compartments of the metabolism cannot be measured accurately or in a sufficient resolution by time.

The number of parameters that can be monitored and hence integrated in models as measured variables is steadily increasing. Nowadays, very advanced techniques are available including online respirometry [162], continuous glucose monitoring [163], on-line high performance liquid chromatography (*on-line* HPLC) [165], multi-wavelength fluorescence (MWF) [166, 167], near-infrared spectroscopy (NIRS)[162, 168, 169], as well as conventional but very efficient techniques like Dissolved Oxygen Tension (DOT) [164] among others. Also methods for describing the physiological state of the cell are developed for monitoring bacterial cultivation like flow cytometry for example [170, 171]. However, there is no possibility to monitor these parameters on-line. Junne [172] reports an approach aiming at monitoring the polarisability and cell length at line and correlate these parameters with the cell viability as it was performed based on flow cytometry studies in batch, fed-batch, and continuous cultures [173]. The great advantage of the method is the absence of cell staining which allows for a much simpler and automated sample preparation. Still efficient methods to obtain sufficient state of process information with online observations are to be developed. For this reason, parallel to advances in measuring techniques, models aiming at an accurate description of bacterial behavior at different levels are being developed.

8.2 MODELS FOR THE DESCRIPTION OF *ESCHERICHIA COLI* CULTIVATIONS

Many approaches have been developed during the last thirty years to create descriptive and predictive models for bacterial organisms. Some of them have been created for the scientifically well-described facultative bacterium *Escherichia coli*. Due to its easy cultivation, this bacterium became the "workhorse" of experimentalists for studying novel microbiological and analytical methods and for the industrial production of proteins and other substances of commercial interest [174].

In recent years, dynamic models including the regulatory interaction on the metabolomic level have been developed. For example, Chassagnole [175] created a kinetic model of the sugar uptake system (phosphotranspherase system – PTS) and the glycolysis following the Embden-Meyerhof-Parnas pathway in *E. coli*. Also, Wang [176] described the catabolite repression based on the sucrose and glycerol transport system in *E. coli*. In this study, the PTS-dependence of the sucrose transport system was considered in the model. More and more knowledge is available concerning the interaction of messenger nucleotides, which act at the DNA replication, translation,

transcription and enzyme activity [177]. Models included the general stringent response in *E. coli*, which has a major impact on the process performance [178]. The integration of signal molecules can serve as bridge between the regulation on the metabolome and the transcriptome. The alarmone guanosine tetraphosphate (ppGpp) acts as a main modulator of gene transcription [179]. It is activated, when the cell starves as in large-scale fed-batch cultivations. Therefore, the simulation of the impact of ppGpp and other nucleotides on the cell's regulation is of great interest for predicting the physiological state of the cell and consider it for modeling [126, 180]. The consideration of alarmones contribute to a more sensitive prediction capacity and a simulation of dynamic cell behavior [181].

For example, a recently developed model of the total central carbon metabolism in *E. coli* comprises of 50 kinetic rate equations and 46 equations that simulate gene expression in the corresponding pathways [182]. This leads to the problem that for a wider application range of the model, parameter estimation becomes a difficult and sometimes even impossible task. However, parameter estimation with a limited set of experimental data leads to unreliable results while multiple solutions exist.

Among recent strategies to overcome these problems are the application of simplified kinetics for expressing the biochemical reaction rate equations, e.g. with linear-logarithmic (so called lin-log) approaches [183, 184]. Also the successful application of piecewise-linear approximations for a multilevel considering model for *E. coli* cultivations has recently been described [7]. Here, the model reduction maintained the dynamic simulation capacity of the model for the starvation response of *E. coli*, although precision of prediction could not be fully maintained. Other approaches to reduce the complexity of complex models included temporal decomposition and establishment of pools which describe the physiologic status of the cell with very few parameters [185]. Still these reduced versions at heuristic level, need to be continuously fitted to new conditions and do not offer nor inside of process conditions, nor of the physiological state of the bacteria. Hence, although there is a great field of application for models for description of *E. coli* cultivations in research and production, their application is very restricted due to the afore mentioned limitations.

8.2.1 DIVISION OF PHYSIOLOGICAL STATES

In an effort to create a tractable model able to describe substrate and oxygen uptake capacities of the cells, Lin [91] proposes a model where cellular response dynamics are lumped into two variables. For the equation system see Appendix A. This model considers time dependent uptake capacities over time and thus enables a better fit of experimental glucose data. Despite the important advances achieved, the model developed by Lin does not consider variation of cell response caused by the physiological state of the bacteria. For this reason, in this section an alternative modeling approach is proposed considering three main cell states, namely overflow,

growth on substrate limitation and starvation (under which the regulatory interaction of the alarmone ppGpp is active).

Recent studies in chemostat and accelerostat *E. coli* cultivations have proven the dynamic development of transcription, protein expression and carbon conversion at different dilution rates [155, 159, 186]. A clear difference of the expression of enzymes in the glycolysis and the TCA cycle depending on the state of the culture was determined. This information can be used to account for altered enzyme concentration at different cultivation conditions (leading to different physiological conditions inside the cell). For example, acetate was shown to be synthesized already at low carbon supply via the pyruvate oxidase (Bpox) from pyruvic acid [156]. Acetate is then converted to acetyl-CoA via the acetyl-CoA synthase (Acs) [159, 187]. The expression of Acs is strongly reduced (more than 30-fold) at high substrate uptake rates (at the onset of acetic acid accumulation). Over expression of this pathway has shown that acetate accumulation could be reduced [188]. This indicates that the pathway provides acetate accumulation at lower substrate uptake rates, although acetate is produced. The low Km value of Acs of 200 μM supports this hypothesis. Despite of this, the usual conversion of accumulated acetate via the acetate kinase to acetyl-phosphate is characterized by a value of the catalytic enzyme of 7-10 mM [158]. The expression of it does not change at altered substrate uptake [159].

Three submodels, each one describing one of the aforementioned physiological states of the cell, are to be developed in an effort to increase prediction accuracy while maintaining model simplicity. By this means it is possible to reduce the number of parameters and increase model flexibility and model identifiability [189]. A further advantage of this approach is the achievement of a reliable mechanistic recognition of non measurable parameters. The simplicity of the models applied, the robustness and a better evaluation of the changing states of the process are potential contributions of MR to the description of *E. coli* cultivations. Besides, MR enables a closer analysis of the dynamics of the model and the correctness of its structure.

Similar to the concept of MR, Veloso [190] developed a software sensors framework to monitor *E. coli* fermentations. He applied an asymptotic observer based on a simplified model with five state variables, namely:

- Biomass concentration
- Substrate concentration
- Acetate concentration
- Oxygen concentrations in the offgas
- Carbon dioxide concentrations in the offgas

combined in the form:

$$\frac{d}{dt}\begin{bmatrix} X \\ S \\ Ac \\ O \\ C \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ -k_1 & -k_2 & 0 \\ 0 & k_3 & -k_4 \\ -k_5 & -k_6 & -k_7 \\ k_8 & k_9 & k_{10} \end{bmatrix} * \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} * X - D * \begin{bmatrix} X \\ S \\ Ac \\ O \\ C \end{bmatrix} + \begin{bmatrix} 0 \\ \left(\frac{F_{in}}{W_m}\right) * S_{in} \\ 0 \\ OTR \\ -CTR \end{bmatrix} \quad \textbf{(8.1)}$$

where X, S, Ac, O, and C represent biomass, glucose, acetate, dissolved oxygen, and dissolved carbon dioxide concentrations, respectively; μ_1, μ_2, and μ_3 are the specific growth rates; k_i are the yield (stoichiometric) coefficients; F_{in} and S_{in} are the substrate feed rate and the glucose concentration in the feeding solution, respectively; D is the dilution rate and W_m is the culture medium weight. CTR is the carbon dioxide transfer rate from liquid to gas phase and OTR is the oxygen transfer rate from gas to liquid phase.

Finally, Veloso divides the process in four possible phases, named regimes:

- regime A:
  - simultaneous oxidative and overflow growth on glucose ($\mu_1, \mu_2 > 0$; $\mu_3 = 0$)
- regime B:
  - oxidative growth on glucose ($\mu_1 > 0$; $\mu_2 = \mu_3 = 0$)
- regime C:
  - simultaneous oxidative growth on acetate and glucose ($\mu_1, \mu_3 > 0$; $\mu_2 = 0$)
- regime D:
  - oxidative growth on acetate ($\mu_3 > 0$; $\mu_1 = \mu_2 = 0$)

This proposition allows building observable models with a reduced number of measured state variables. Similar to the MR approach, Veloso divides the batch process to enable its description with simpler models. From a physical point of view, the growth rate is considered to be constant for each regime. Assuming a piecewise constant growth rate simplifies the model to the extent where observability is obtained.

Although theoretically applicable for process monitoring, the proposed method has some important drawbacks.

- The model is not continuous (discontinuity hinders gradient calculation in the switching point).
- The model proposed for monitoring is simple and has no physical basis nor offers information about the system.
- The results obtained cannot be translated to a first principle model to extend state of information.

- The regimes have to be known beforehand. This supposes important process knowledge impossible to obtain in a real process.

In order to overcome the above mentioned disadvantages, the MR approach is proposed.

8.3 MODELING *ESCHERICHIA COLI* BATCH FERMENTATIONS WITH MECHANISM RECOGNITION

In this section, MR is applied to achieve an efficient mechanistic modeling and simulation of *E. coli* batch fermentations and the simulation framework is validated against experimental data. Fermentation processes are characterized by its dynamic behavior described by parameters such as growth rate, substrate concentration and cellular metabolic activity. Although models able to describe individual batch and fed-batch fermentations exist, they become unreliable when applied to different fermentation conditions. To overcome this drawback, diverse models are used at various regimes enabling not only a better description of the process, but also an improved understanding of non measurable characteristics. Three models compete in different intervals of the process. The candidate models are:

- Overflow metabolism model (OF),
- Growth on substrate Limitation model (SL),
- Starvation model (ST).

Using an adequate model sequence, acetate formation, substrate consumption and cell growth are described with high accuracy. Moreover, the data needed to fit the models are reduced and a standardization of the model to be applied in different process states is enabled.

8.3.1 GENERAL MODEL

The model for the description of *E. coli* K12 fed-batch fermentations was extracted from Lin [91]. This model was originally built in an effort to describe all physiological and regulatory effects, which cause uptake capacity variation of substrate and oxygen, with a relatively simple kinetic model. A series of experiments with glucose pulses were carried out to determine the dynamics of the uptake capacity of glucose during fermentation.

To describe these variations, Lin creates two fictitious enzymes in which all physiological and regulatory effects are put together, the enzymes for glucose uptake and for oxygen uptake to biomass [91]. This model successfully describes the variations of the uptake capacities of substrate and oxygen of determined fermentations. The behavior of the six principal variables: biomass, substrate, acetate, DOT substrate uptake, and oxygen uptake are depicted in Figure 8.1

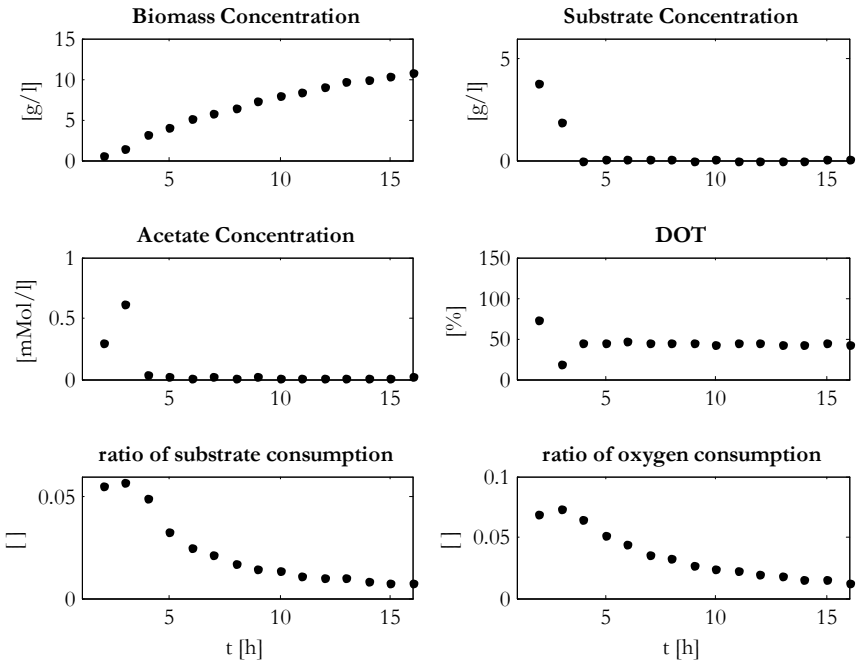

Figure 8.1: Integration of the kinetic model proposed by Lin [91]

Still, the model fails to describe the extremely different behavior of *E. coli* under different process conditions.

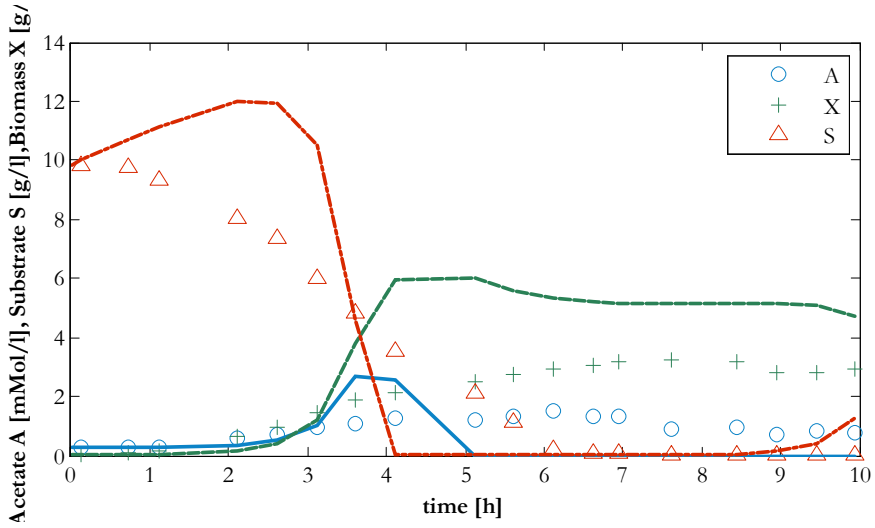

Figure 8.2: Complex model (Lin et al.) fitted to experimental batch cultivation data.

In addition, the model contains discontinues equations which represents a major obstacle for robust simulation and optimization. Moreover, the model fails to describe different experiments. The results of the parameter estimation against real data are depicted in Figure 8.2. The optimization was carried out with a stochastic optimizer PSO (section 5.3) in order to overcome local minima.

8.3.2 SUBMODELS FOR DIVIDING METABOLIC STATES

A reduction of the model in three submodels makes it possible to overcome the previously mentioned drawbacks improving description accuracy and robustness. Each submodel describes one of the three main states of the cell: overflow metabolism Figure 8.3, growth under substrate limitation Figure 8.4, and starvation Figure 8.5.

MODEL FOR DESCRIPTION OF THE OVERFLOW METABOLISM

Drawbacks of acetate production are widely known and have been reported in various contributions. Acetate production is related to growth inhibition [153, 191] and to reduction of recombinant protein production [192, 193]. Lin [91]summarizes two rate-limiting steps contributing to overflow of acetate in aerobic glucose-based cultures of *E. coli*:

- in electron transport system [151]
- rate-limiting steps in the in the TCA cycle [194, 195]

The general model is reduced to limit its description capacity to the specific conditions of the overflow metabolism (OF):

- Substrate concentration is always higher than the minimum amount needed for maintenance.
- Bacteria consume substrate and oxygen to its maximum possible under consideration of substrate and product limitation.
- The expression of both enzymes, for substrate and oxygen uptake capacity, is maximal and constant.

Acetate consumption is considered, since bacteria also consume acetate in overflow conditions where high acetate concentration is present in the medium.

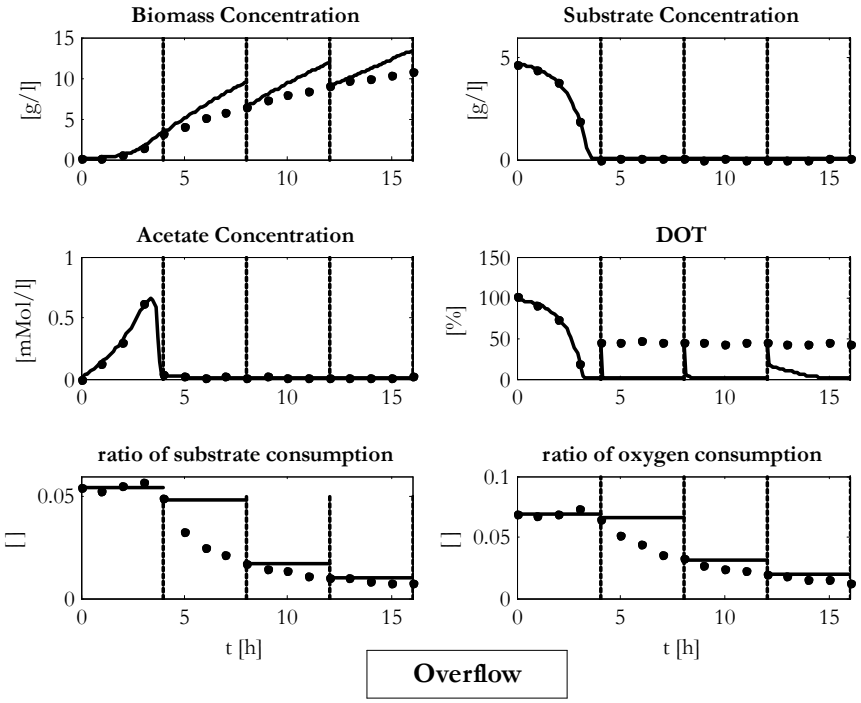

Figure 8.3: Comparison between the complex model (dots) vs. the overflow submodel (lines) initializing in four different intervals.

In Figure 8.3 it can be seen that the submodel describes the initial stage of the process with high precision. Four intervals were selected arbitrarily to test the description capacity of the model in different stages of the process. The prediction accuracy is drastically reduced as the concentration of substrate drops. This suggests that the submodel is adequate for the identification of the bacterial growth with overflow metabolism. For the equation system see Appendix B.

MODEL FOR DESCRIPTION OF GROWTH UNDER SUBSTRATE LIMITATION

Many fermentation strategies as well as metabolic engineering approaches have been applied to reduce the production of acetate [191, 196-202]. It is commonly accepted, that optimal conditions for cultivation are obtained under (moderate) substrate limitation. By these means acetate accumulation is minimized and biomass to substrate yield maximized, while maintaining a relatively high growth rate [155]. A topology of the regulatory network of carbon limitation has been published by Hardiman [181] on which it is possible to estimate to which extend substrate limitation is suitable for application and when starvation and stringent responses are initiated.

Following assumptions were made to create the substrate limitation model:

- Substrate concentration is always higher than the minimum amount needed for maintenance.
- There is no limiting capacity for oxygen consumption.

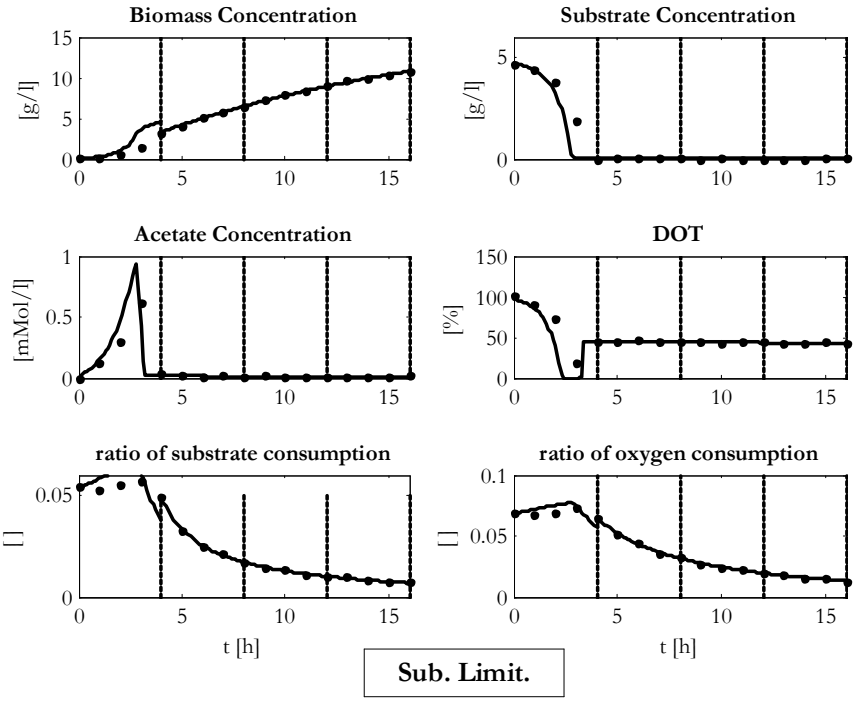

Figure 8.4: Comparison between the complex model (dots) vs. the substrate limiting submodel (lines) initializing in four different intervals.

Contrary to the previous submodel, the submodel for substrate limitation is only able to describe cell behavior at very low substrate concentrations Figure 8.4. The model predicts an extremely fast growth in overflow conditions, because it lacks uptake capacity limitation at overflow metabolism responsible for acetate synthesis and excretion. The detailed equation system is presented in Appendix C.

MODEL FOR DESCRIPTION OF CELL STARVATION

Finally the description of the starvation stage is also considered. The cell starvation is to be avoided at all cost in industrial *E. coli* cultivations. Still large scale reactors inherently imply special concentration gradients in the medium exposing bacteria to undesired conditions. Extreme substrate limitation may trigger the phosphoenol pyruvate glyoxylate (ppG) shunt. Once the stringent response is activated, reactivating bacterial growth requires long periods under substrate excess.

To reduce the models it was assumed that:

- substrate uptake is lower than required by the cell, hence no energy limitation is required
- enzyme rate are minimal and constant

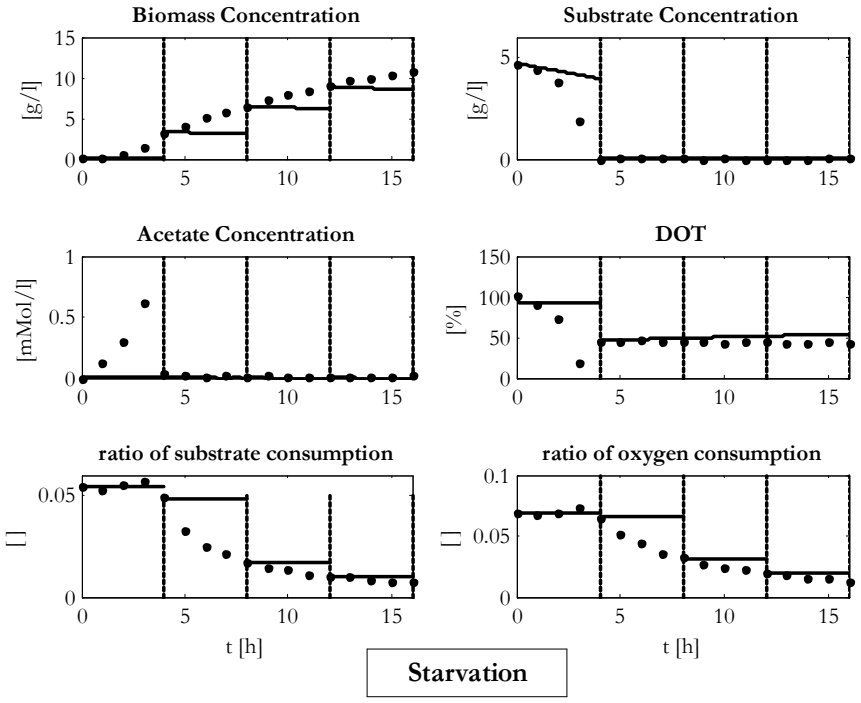

Figure 8.5: Comparison between the complex model (dots) vs. the cell starvation submodel (lines) initializing in four different intervals.

As expected, the model describing cell starvation is not able to predict the behavior at any stage of the process Figure 8.5. This indicates that the process does not reach starvation conditions. For the equation system see Appendix D.

The structure of the submodels is shown in Table 8.1. The identifiability of the submodels will be determined by the parameter boundaries and the consistency of the general structure indicates the following. This again is key to accurate regime recognition.

Unlike the process constant parameters, the regime-wise constant parameters are optimized independently in each regime. The parameters estimated during the recognition process are shown in Table 8.1

Table 8.1: Parameters considered for the model fit

PCPs (general structure, section 6.3.2):

q_{Eod}	specific death rates for the uptake enzyme of oxygen [g/(gl)]
q_{Omax}	Maximal specific oxygen uptake rate [g/(gh)]
q_{Smax}	Maximal specific substrate uptake rate [g/(gh)]
q_{Esd}	specific death rates for the uptake enzyme of substrate [g/(gl)]

RWPs:

Y_{XSox}	Yield of the enzyme for oxidative energyproduction from the biomass growth [-]
Y_{EOX}	Yield of the enzyme for substrate uptake from the biomass growth [-]
Y_{ESX}	Yield of the enzyme for respirationfrom the biomass growth [-]
Y_{XA}	Yield of the enzyme for acetate from the biomass growth [-]
Y_{XSof}	Yield of the enzyme for overflow from the biomass growth [-]

8.4 MATERIAL AND METHODS

(All chemicals in this study were either purchased by Sigma Aldrich GmbH, Munich, Germany, or Carl Roth KG, Karlsruhe, Germany, if not otherwise stated.)

8.4.1 STRAIN AND CULTURE CONDITIONS

Glucose	10 g/l
Yeast -extrakt	5 g/l
K_2HPO_4	3 g/l
KH_2PO_4	1.5 g/l
$(NH_4)_2SO_4$	1.25g/l
$MgSO_4$	0.1 g/l
$Na_2EDTA*7H_2O$	0.037 g/l
NaCl	0.01 G/l
$FeSO_4*7H_2O$	0.001 g/l

Glucose was autoclaved separately to avoid Maillard-reactions.

1000 ml Erlenmeyer flasks were filled with 250 ml of medium. The medium was inoculated with *E. coli* K12 W3110 picked from a frozenstock culture. The flasks were incubated at 37°C and 150 rpm on a longitudinal shaker.

BATCH BIOREACTOR EXPERIMENTS

A reactor KLF2000 (Bioengineering AG, Wald, Switzerland) was used for all fermentations Figure 8.6. To insure sufficient oxygen supply, the reactor was sparged with air at a rate of 1 vvm. Both, the gas inlet and the exhaust gas were steril-filtered with autoclavable cellulose filters.

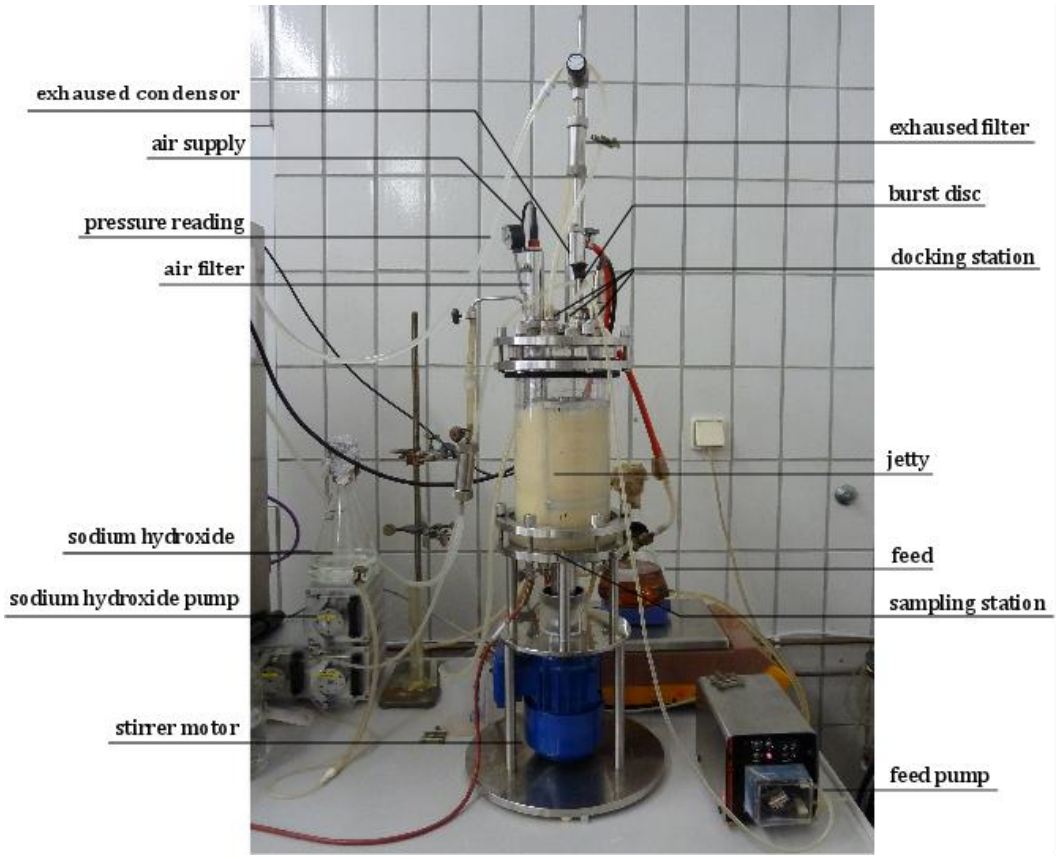

Figure 8.6: Bioreactor KL2000 at *E. coli* batch cultivation [203]

The inlet gas filter was autoclaved separately while the outlet gas filter was autoclaved together with the vessel by a small steam flow. The reactor was sterilized prior to cultivation at 120°C and 1 bar overpressure for 20 min. The level of pH was controlled to pH 7.0 with NaOH 30% w/w during the acetate production phase and HNO_3 10% w/w during the acetate consumption phase. The stirrer was working at 700 rpm and polyethyleneglycol 2000 to 3000 was added in small amounts (~ 1ml) to reduce foam whenever needed.

8.4.2 ONLINE ANALYSIS

A sterilizable temperature detector Pt100 is connected to an electronic measuring and control unit, which is controlling the heating and cooling devices.

An Ingold single-rod measuring cell is used to measure the pH of the medium. The concentration of the dissolved oxygen in the medium is detected with an amperometric (polarographic) Ingold pO_2-electrode (Mettler-Toledo Inc., Mettlach, Germany).

O_2 concentration in the exhaust gas is determined by an OXYGOR 6 N (Maihak AG, Hamburg, Germany) device following the magneto-pneumatic measuring principle.

CO2 concentration in the exhaust gas is measured by a non-depressive infrared photometer with a selective optical-pneumatic receiver UNOR 4 N (Maihak AG, Hamburg, Germany).

The oxygen uptake was calculated based on the exhaust gas analysis as follows:

$$Q_{O2} = \frac{\dot{V}_G (Y_{O_2}^{\alpha} - Y_{O_2}^{\omega})}{V_F * 22.4} \qquad \textbf{8.2}$$

The carbon dioxide production rate was calculated as:

$$Q_{CO2} = \frac{\dot{V}_G (Y_{CO_2}^{\omega} - Y_{CO_2}^{\alpha})}{V_F * 22.4} \qquad \textbf{8.3}$$

Where:

V_F - fluid volume [l]

$\dot{V}_G$ - volumetric gas flow m^3/h

Q_{O2} - oxygen uptake rate [mol/lh]

Q_{CO2} carbon dioxide production rate [mol/lh]

$Y_{O_2}^{\alpha}$ -mole fraction of oxygen in the incoming gas phase [ ]

$Y_{O_2}^{\omega}$ - mole fraction of oxygen in the exhaust gas [ ]

ONLINE-MEASUREMENT OF THE OPTICAL DENSITY WITH ELOCHECK®

EloCheck® Figure 8.7 was also developed by EloSystemsGbR for biomass online monitoring. Because of the implementation of a smaller cuvette with a thickness of 1.6 mm instead of the usual 10 mm, EloCheck® is able to measure the optical density of the sample through a bypass system [204]. EloCheck® needs no sample dilution which makes it possible to reintroduce the measured sample into the reactor. By this means, the device takes a sample and measures it every 15 seconds. Both biomass curves from EloCheck® and offline analysis were compared and a high similarity on the run of the curves was evident. The data of EloCheck® was not used for further analysis because of small deviation due to antifoam agent which caused temporary occlusion in the cuvette.

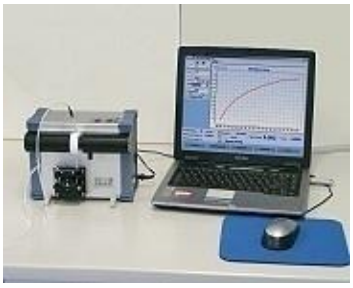

Figure 8.7: EloCheck®

8.4.3 OFFLINE ANALYSIS

SAMPLE PROCEDURE

About 5 ml of crude extract were withdrawn from the reactor every 30 min. Two Eppendorf "Flex-Tubes® 1.5 ml, per 1,000 pcs." were filled with 1 ml sample. The supernatant was separated through centrifugation for 10 min at 13,000 rpm. One flex-tube supernatant was used for the glucose determination and the second one was stored at –20°C for later acetate determination. About 3 ml were used for biomass determination.

BIOMASS DETECTION

Biomass was determined through optical absorption measurement with a Zeiss Spektral Photometer (PM2A, West Germany). The biomass containing sample was diluted until the extinction was in the range of the Lambert-Beer law (0.2 to 0.5). The determination of the extinction was performed at λ = 600 nm. The equation for the linear regression between biomass and extinction was:

Biomass [g/l] = 0.398*extinction [nm]-0.039

ENZYMATIC DETERMINATION OF GLUCOSE

Glucose determination was performed with the liquiUV^{mono} Test (Human Geselschaft für Biochemical und Diagnostica mgH, Wiesbaden, Germany). In this test, the glucose was the substrate for the reaction which is catalyzed by the hexokinase and the glucose-6-phosphat-dehydrogenase at presence of Adenosintriphosphat (ATP) and Nicotinamide Adenine Dinucleotide (NAD^+, NadH). During this chemical conversion NadH, is accumulated:

$$\text{Glucose} + \text{ATP} \longrightarrow \text{Glucose-6-Phosphat} + \text{ADP}$$

$$\text{Gucose-6-Phosphat} + \text{NAD}^+ \longrightarrow \text{Gluconat6-Phosphat} + \text{NADH} + \text{H}^+$$

The amount of NadH is proportional to the concentration of the glucose. The extinction of NadH is measured at λ = 340 nm. A calibration curve consisting of

duplicate measurements at five different concentrations was made prior to each fermentation. In Figure 8.8, the calibration curve for the first experiment is shown:

Concentration [g/l] = 2.776*extinction [nm] – 0.014

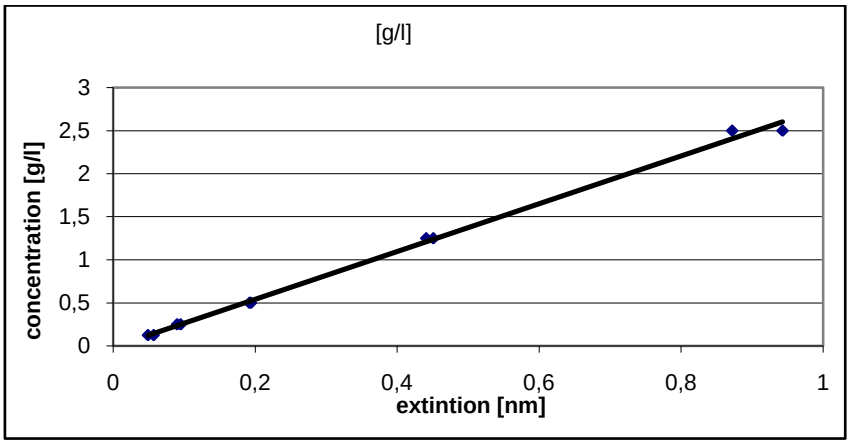

Figure 8.8. Calibration curve for glucose determination

Optical measurement was realized by a Beckman DU640 spectrophotometer (Beckman Coulter Inc., Fullerton, USA).

The specific substrate uptake rate was obtained by dividing the difference off the acetate concentration between two samples, by the biomass at this time.

$$q_{Gluc} = \frac{\frac{C_{Gluc\,n+1} - C_{Glu\,c_n}}{t_{n+1} - t_n}}{C_{Bio\,n} + \frac{C_{Bio\,n+1} - C_{Bio\,n}}{2}} \qquad 8.4$$

where:

q_{Gluc} = specific glucose uptake rate [h^{-1}]

$C_{Glu\,c_n}$ = glucose concentration [g/l]

$C_{Bio\,n}$ = Biomass concentration [g/l]

t_n = time [h]

ENZYMATIC DETECTION OF ACETIC ACID

Acetic acid determination was realized with the UV method, which is based on the reaction from acetic acid to acetyl-CoA [205] (ROCHE Test Kits Cat. No.10148261035, R_BIOFARM AG, Darmstadt, Germany).

$$\textbf{Acetate + ATP + CoA} \longrightarrow \textbf{acetyl-CoA + AMP + pyrophosphate}$$

The determination is realized on a light absorbance wave length of λ = 340 nm.

Samples were taken every 30 min, centrifuged and stored at -20°C for no more than three weeks. Optical measurement was realized by a Beckman DU640 spectrophotometer (Beckman Coulter Inc., Fullerton, USA).

The calibration curve was plotted with the data obtained out of the supplied control solution at five different concentrations Figure 8.9.

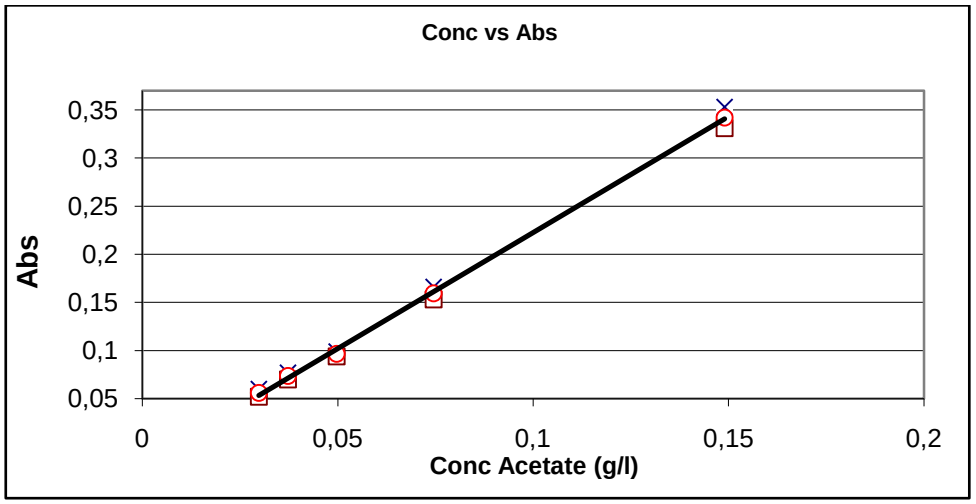

Figure 8.9. Calibration curve of acetate

The concentration of acetate was calculated with the following formula obtained out of the calibration curve:

Acetate concnetration = 2.412*Absorbance – 0.0186 [g/l]

The specific acetate production rate was obtained by dividing the difference off the acetate concentration between two samples, by the biomass at this time.

$$q_{Ac} = \frac{\frac{C_{Ac_{n+1}} - C_{Ac_n}}{t_{n+1} - t_n}}{C_{Bio_n} + \frac{C_{Bio_{n+1}} - C_{Bio_n}}{2}} \qquad 8.5$$

where:

q_{Ac} = specific acetate production rate [mMol/(g*h)]

C_{Ac_n} = acetate concentration [mMol/l]

C_{Bio_n} =Biomass concentration [g/l]

t_n = time [h]

DETERMINATION OF INTRACELLULAR NADH CONCENTRATION

The intracellular NADH concentration was measured by a cycling assay according to Nisselbaum and Green [206] modified by Bernofsky and Swan [207] and improved by San and Bennett [208].

Briefly, a 1 ml sample was pipetted into a microcentrifuge tube. The centrifugation was performed for 10 min at 13,000 rpm. 300 µl of 0.2 M NaOH were added for cell lysis. After 10 min of incubation in a 50°C bath, samples where cooled to 0° C and neutralized by carefully adding 300 µl of 0.1 M HCl while vortexing. Cellular debris was removed by centrifugation for 5 min at 15,000 rpm.

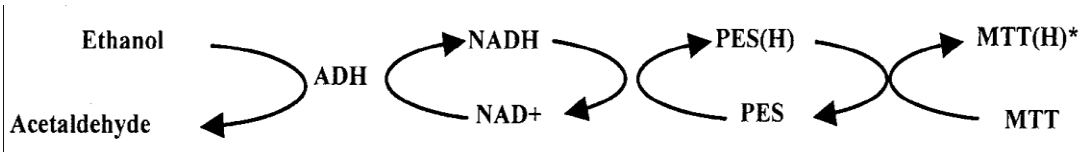

Figure 8.10: Mechanism of the reactions involved in the assay

The cycling assay Figure 8.10 was performed using a reagent mixture consisting of equal volumes of 1.0 M bicine buffer (pH8.0), absolute ethanol, 40 mM EDTA (pH8.0), 4.2 mM 3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide (thiazolyl blue; MTT) and twice the volume of 16.6 mM phenazineethosulfate (PES), previously incubated at 30°C. The following volumes were added to 1 ml cuvettes: 50 ml neutralized extract, 0.3 ml water and 0.6 ml reagent mixture. The reaction was started by adding 50 ml of alcohol dehydrogenase (ADHII isolated from rabbit) of a concentration of 500 or 100 U/ml in 0.1 M Bicine (pH 8.0) buffer. The absorbance at λ = 570 nm was recorded for 10 min. The assay was calibrated with 0.01–0.05 mM standard solution of NAD^+.

Table 8.2. Composition of solution A

Solution	Volume 650 µl
1.0 M bicine buffer (pH8.0)	650 µl
absolute ethanol	650 µl
40mM EDTA (pH8.0)	650 µl
4.2mM MTT	650 µl
16.6 mM PES	1300 µl

1 ml cuvettes were filled with the following mixture:

50 µl extract

0.6 ml solution A

0.3 ml water

To start the reaction:

50 µl of ADH (500 or 100 U/ml in 0.1 M Bicine (pH 8.0) buffer)

The biomass yield was calculated with the following equation:

$$Y_{x/s} = \frac{C_{Bio_{n+1}} - C_{Bio_n}}{C_{Gluc\,n+1} - C_{Glu\,c_n}} \qquad \textbf{8.6}$$

where:

$Y_{x/s}$ = Biomass yield [-]

$C_{Glu\,c_n}$ = glucose concentration [g/l]

C_{Bio_n} =Biomass concentration [g/l]

8.4.4 DATA TREATMENT

The data of every pair of fermentations with the same conditions were considered for calculating a fit-curve. Samples that were considered as outliers were manually excluded. In the case of the biomass curve, oscillations at the beginning of the curve were also attenuated per hand.

Acetate samples were taken every 30 min. The data of every pair of fermentations with the same conditions were considered for calculating a fit-curve whenever possible. The fit curve was calculated in Matlab with a 7th grade polynomial. The acetate production-rate was obtained after dividing the difference of acetate between two samples and dividing the result through the time between both samples.

$$\frac{\Delta Ac}{\Delta t} = \frac{\left(Ac_2 - Ac_1\right)}{\left(t_2 - t_1\right)} \qquad \textbf{8.7}$$

Finally, the acetate building-rate was divided with the biomass to obtain the specific acetate building-rate.

This procedure was also applied to all other parameters which were converted into specific rates.

8.5 EXPERIMENTAL VALIDATION

First we compare submodel performance against a simulated data set. The submodels created in the previous section showed good performance when compared to simulated data. This indicates that it is possible to find combinations of simple models with similar dynamics to more complex versions. MR achieves an accurate detection of the change of regime and finds consistent parameter combinations for both models (OF and SL). Since the complex model used for submodel building is the same model used for data generation, it can be assured that the structure of the model is globally correct (no systematic error). In other words it is possible to prove that the structure of the complex models describes the process dynamics completely and no uncertainties or unknowns are present besides white noise (stochastic error). Certainty of optimal structure for the complex model allows a direct analysis of process phenomena and the precise detection of changing regimes.

Nevertheless, the highest accuracy prediction of these artificially generated observations is clearly obtained with the complex model, which also represents the optimization with the largest number of parameters to fit. One could argue that, also in these cases, MR offers a simplified basis for simulation and optimization. However, the effort of submodel building and MR programming does not justify its application against simulated data. It has been clearly stated in section 1.6 that the objective of the MR approach is to enable industrial application of complex models. Therefore it has to be proven that MR allows using complex models applied to real experimental data. It is essential to test the performance of MR against data sets obtained in real experiments.

8.5.1 CONDITIONS FOR PROPER PROCESS DESCRIPTION WITH MR

MR was tested against batch cultivation with *E. coli* K12 W3110 bacteria. To ensure successful recognition of the different regimes during the process, the system is required to fulfill following conditions.

DYNAMIC PROCESS

- The batch process is characterized by changes in the state of the system throughout the complete run.

PROCESS HAS TO HAVE MORE THAN ONE REGIME

- As previously discussed, the cultivation of *E. coli* is known to present three basic regimes; overflow metabolism, growth under substrate limitation and starvation.

DETAILED MODEL OF THE GENERAL PROCESS AVAILABLE OF THE FORM:

- DAE system index zero or one
- no discontinuities
- known initial conditions

The model applied to describe this process fulfils all the above mentioned conditions

PARAMETER BOUNDARIES HAVE TO ASSURE GLOBAL CONVERGENCE WITH GRADIENT BASED OPTIMIZERS

- The parameter set was tested for global optimality with stochastic optimization based on PSO [209], where the parameters of the optimizer were set to maximize the probability of global convergence.

INITIAL REGIME IS KNOWN

- Because of the high concentration of substrate at the beginning of the process, it is assured that the initial regime of the process is growth under overflow conditions (after a short lag phase).

THE SEQUENCE OF THE REGIMES IS KNOWN.

- The dynamics of substrate consumption and the reaction capacity of the bacteria assure that the regime following overflow conditions is growth under substrate limitation.

THE MINIMAL LENGTH OF EACH TIME REGIME ASSURES MODEL DISTINGUISHABILITY

- Based on the variance of the data, the form of the general structure and the conditions of the process, it can be assured that the period between overflow regime and substrate limitation is long enough to guarantee identifiability and detection of the switching to substrate limitation regime.

8.5.2 CONDITIONS FOR ACCURATE SWITCHING POINT DETECTION

The above mentioned assumptions assure a successful model description of the process using MR and a rigorous model with small systematic error. Nevertheless, to allow a proper detection of the switching points in the process, following conditions are to be considered:

KNOWLEDGE OF THE NORMAL BEHAVIOR OF THE SYSTEM

- The typical behavior of the cell in each regime is known. In addition, the physical foundation of the complex model offers a basis for evaluation of the dynamics of the process.

DEFINITIVENESS OF THE CHANGING BEHAVIOR

- The distinguishability test proves that, at least for the simulated experiments, the submodels are distinguishable. From this it may be supposed, that the process offers enough information to detect the difference in bacterial behavior under both regimes.

AVAILABILITY OF AT LEAST ONE OBSERVATION REFLECTING REGIME CHANGE

- At this stage (model validation) offline measurements are fitted and interpolated to obtain "online" measurements. Again, real time process monitoring is not considered at this stage. Nevertheless, once the submodels have shown to describe the different regimes with precision and allow switching detection, identifiability and distinguishability tests should be repeated considering online measurements exclusively.

SATISFACTORY STATE OF INFORMATION

- The experiment was carried out in duplicate to assure reproducibility and a proper form of the variance-covariance matrix. During MR the data set of one experiment was utilized.

However, due to the nonlinearity of the model and unknown disturbances of the system, the results obtained by the identifiability and distinguishability test cannot guarantee that the afore mentioned conditions are fulfilled. It is worth reminding that the process of MR is strongly dependent of the capability of the complex model to mirror the real process.

8.5.3 DATA SET

BATCH FERMENTATIONS; 16H, 37°C

Experiments G1 and G2 (replicate) were realized at fermentation temperature of 37°C with a precultivation of 16h.

The results depicted in Figure 8.11-Figure 8.14 suggest that the cultivation showed a typical *E. coli* batch cultivation and good reproducibility when compared against its replicate (G2).

The measurements obtained from experiment G1 were selected for MR validation. The acetate concentration reached a maximal concentration at 23.54 mMol/l. Uptake of

acetate started at t = 5.68 h (341min). The specific acetate production rate fitted curve had its peak at t = 94 min with 9.9 mMol/(g*h). Acetate uptake started when the glucose concentration reached 0.5g/l.

Glucose was completely metabolized at t = 6.7 h (400 min). The maximal glucose uptake rate reached 4.24 [g/l/h] at t = 1.8 h (109 min). The biomass curve showed the typical shape of growth in batch fermentation with *E. coli*. The lag-phase endured 2 h (120 min). Exponential growth phase was present until t = 6.6h (397 min) when maximum acetate concentration was measured. This point was also the onset of the stationary phase. The biomass yield had a maximal peak at t = 3h (184 min) at a value of 0.33 g/g.

Specific intracellular NAD^+ concentration reached 0.007 mMol/g at t = 4.1h (247 min). The respiration coefficient had its maximum at the same time. The respiration coefficient reached a minimum at t = 1.7 h (100 min) when the specific acetic acid production rate was maximal.

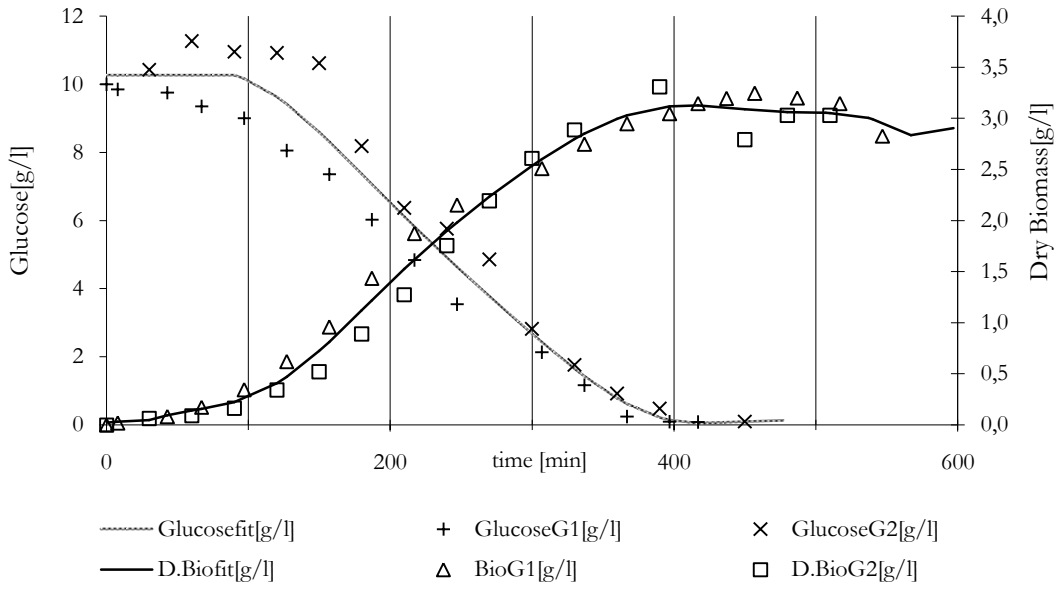

Figure 8.11: Experimental results batch experiment G1. Part I: Dry biomass and glucose concentrations

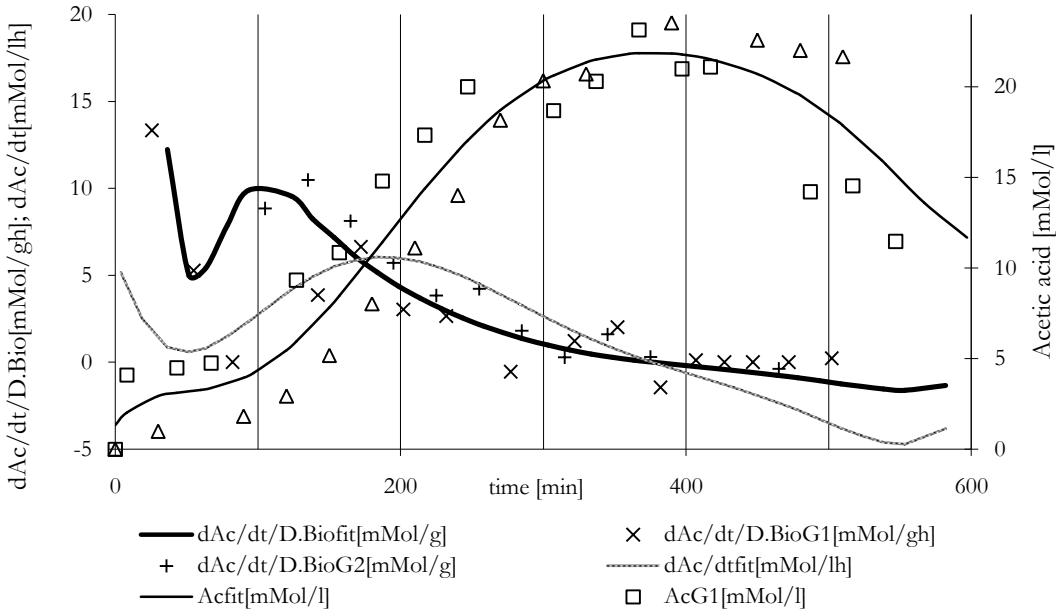

Figure 8.12: Experimental results batch experiment G1. Part II: Specific concentration of acetic acid

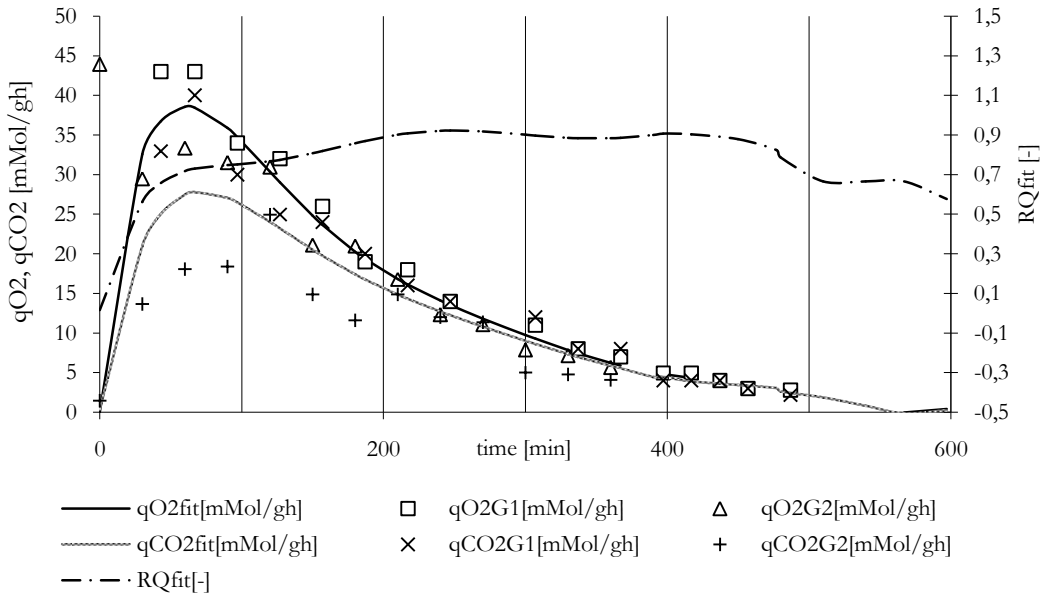

Figure 8.13: Experimental results batch experiment G1. Part III: Outgas concentrations

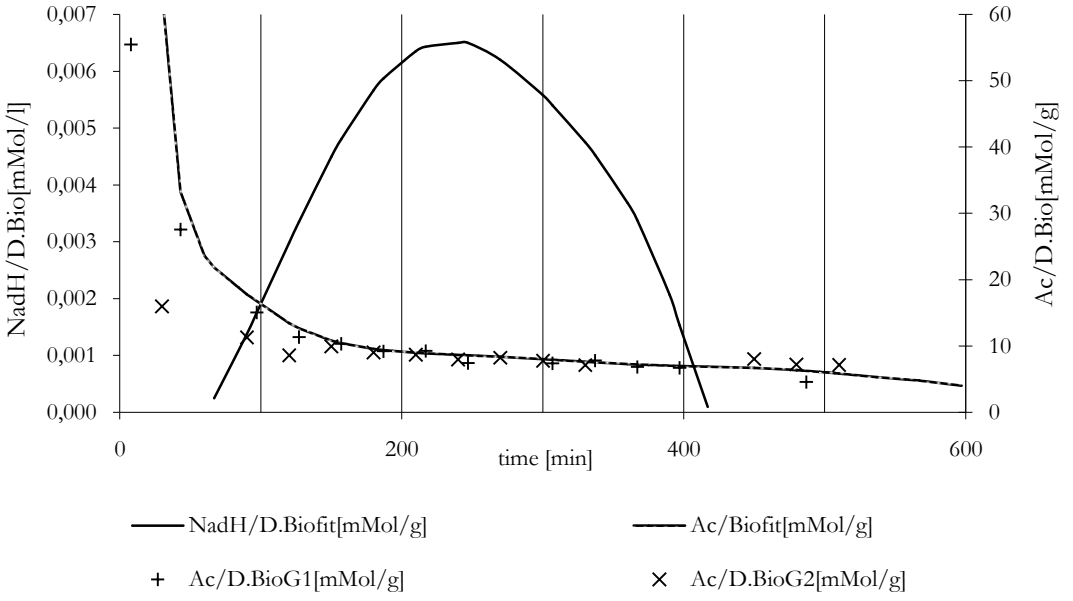

Figure 8.14: Experimental results batch experiment G1. Part IV: Metabolite concentration

8.5.4 RECOGNITION OF OVERFLOW AND SUBSTRATE LIMITATION REGIMES

MR was applied considering the following measured variables:

- Extracellular acetate concentration (mMol/l)
- Extracellular substrate concentration (g/l)
- Biomass concentration (g/l)

INITIAL INTERVAL

As previously discussed, the initial interval is characterized by overflow glucose uptake. The submodel adapted for this regime (OF), has a higher identifiability than the substrate limitation model. This is mainly because the submodel considers the substrate and oxygen uptake enzymes to be constant.

MR INITIALIZATION

Identifiability was obtained after 16 interpolated measurement points same as distinguishability. This permitted initiation of MR after 2.3 h.

DETECTION OF SWITCHING POINTS

Detection of the switch between overflow and substrate limitation regime was carried out with the method for fault detection through parameter identification. The constant threshold was set equal to a residual value of 0.3.

INITIAL CONDITIONS OF THE INTERVAL K+1

Due to assumption of instant switch of the regime, the initial values of the non measured variables had to be included in the parameter estimation problem. Still, the program showed no problem to find an accurate solution.

SWITCHING INTERVAL

The experimental results confirmed that considering an instant switch of regime allows an accurate description of the process. Still, the physical evidence suggests the opposite. The effects of this "mistaken" assumption are buffered due to the estimation of the initial values of the substrate limitation regime (k+1).

8.5.5 SIMULATIONS VS. EXPERIMENTAL DATA

The submodel based on overflow metabolism assumption can fit the data obtained from the experiments with some accuracy, the simulation with the estimated parameters is depicted in Figure 8.15. Nonetheless, it is clear that the submodel lacks the capacity to describe the process after hour 7. Besides, parameters, like Yield from substrate to biomass for example, need to be lower drastically, which is reflected in the wrong description of substrate concentration form hour 5.5. Still, it can be clearly seen, that the submodel can describe the first phase of the process with high accuracy.

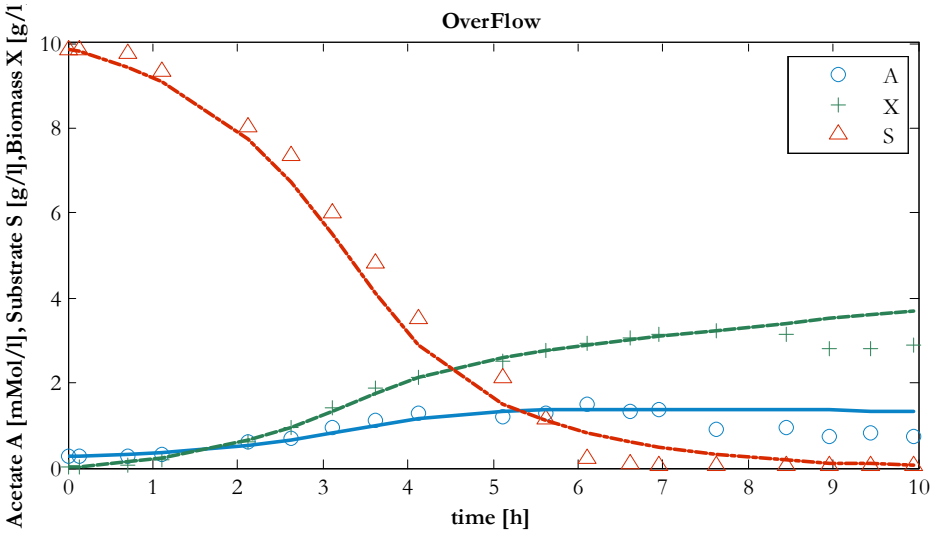

Figure 8.15: OverFlow submodel fitted against experimental data.

Contrary to the overflow submodel, the submodel, which considers that all substrate can be consumed efficiently through the trcyclyc acid cycle, fails to describe the initial phase of the cultivation Figure 8.16. Since the bacteria are not set to any real limitations for substrate uptake, the model consider an unreal growth and substrate consumption. In addition, no acetate is produced and the cultivation stops at hour 7.5 approximately.

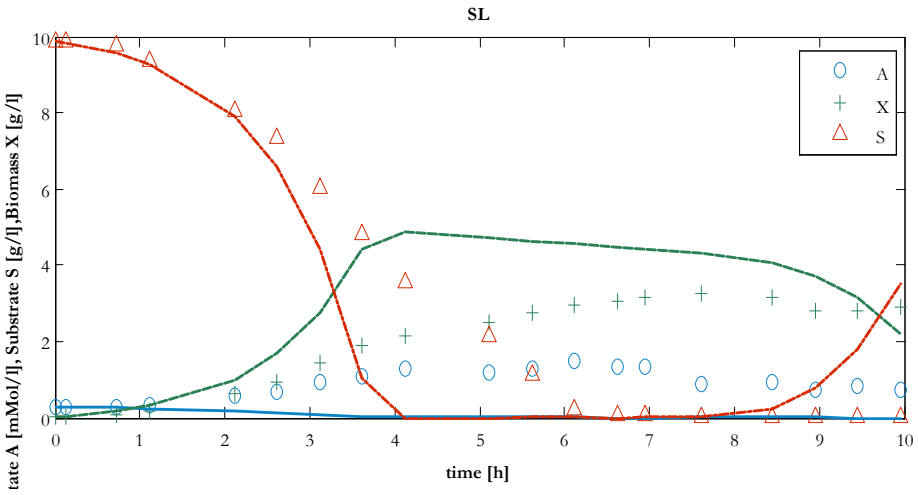

Figure 8.16: Submodel for the description of growth under substrate limitation fitted against experimental data.

Finally, the starvation submodel was also tested Figure 8.17. This submodel has no possibility to describe the fermentation. The optimizer maximizes the yield coefficients to the parameter boundaries without achieving to describe the cultivation.

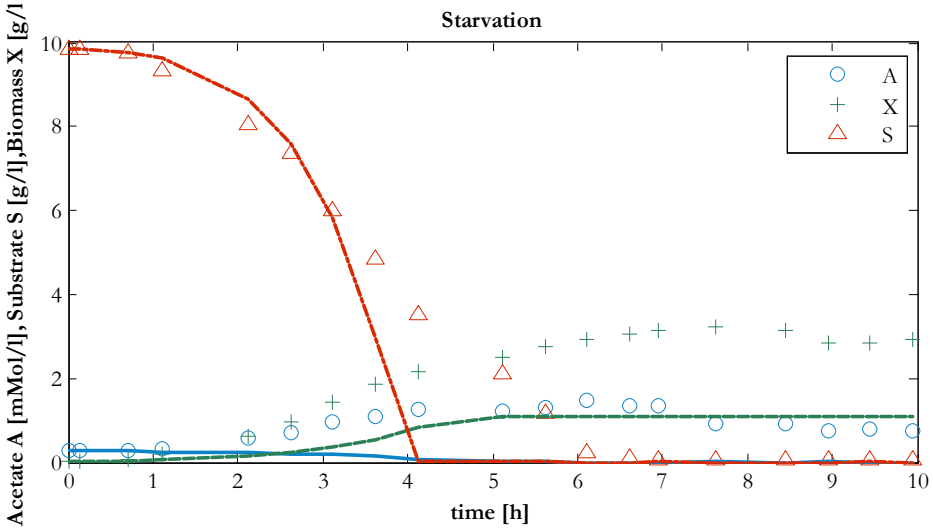

Figure 8.17: Starvation condition described by the corresponding submodel fitted against experimental data.

8.5.6 RESULTS

MR was applied considering the following measured variables:

- Extracellular acetate concentration (mMol/l)
- Extracellular substrate concentration (g/l)
- Biomass concentration (g/l)

The following model sequence was detected by MR Figure 8.18. MR detected the initiation of the regime of growth under substrate limitation at **t = 5.8h (348 min)**. This results are consistent with the evidence shown by metabolite accumulation and acetate uptake **t = 5.7 (341 min)**.

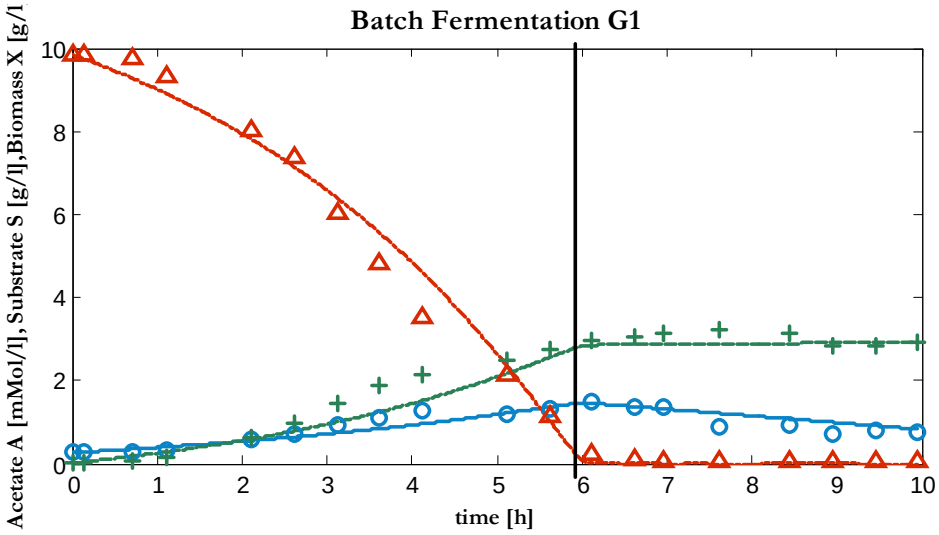

Figure 8.18: Experimental validation of the MR approach.

MR enables the application of Lins model for the description of E. coli batch cultivations. Nevertheless, due to the decrease the process constant parameters in the general structure, and the highly relaxed boundaries required, this set of submodels cannot guarantee adequate recognition of fed-batch processes.

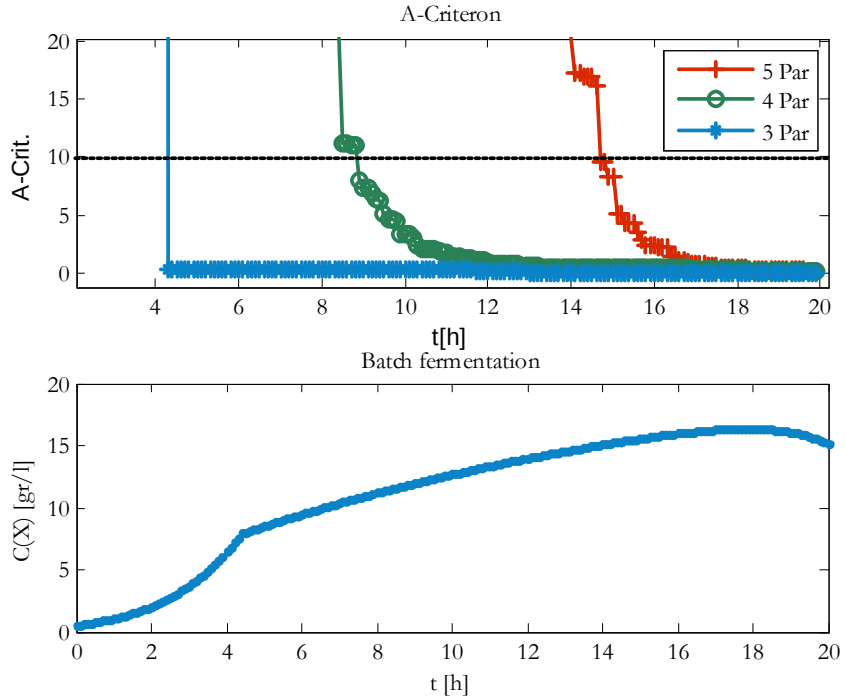

Figure 8.19: Identifiability test considering white noise, standard deviation of 5% in all measurements

To test the identifiability of the submodels with the new highly relaxed boundaries and the reduced general structure we calculate the FIM and evaluate the A-criterion in relation to the number of measurements. The frequency considered for data recollection was 1 min with interpolated data set. The quality of the measurements was simulated with white noise (standard deviation 5%). The results show that the minimal period required to obtain a parameter estimation with less than 10% average standard deviation in all the parameters is more than 4 hours Figure 8.19.

The identifiability is reduced as we add parameters to the regime variable structure. Although this condition can be assured in batch cultivations, it represents a major drawback for fed-batch applications. A new model is required to build a set of submodels with global optimal structure and higher identifiability.

8.6 CONCLUSIONS

MR proved to be able to transform the model originally proposed by Lin into one suitable for parameter estimation and its application in batch processes. The first advantage showed by the application of MR is that the model sequence allows a reduction on the number of parameters to be fitted. Second, an indicator of the current cell state is obtained. MR is able to report whether the metabolism is dominated by overflow (acetate production) or by conversion of metabolites in the citric acid cycle.

However, the model has limitations in its application at processes characterized by a highly dynamic change in the cell physiology and genetic expression faster than in batch processes and in substrate-limited fed batch processes operated at rather constant conditions. When the model had been applied to fed-batch processes where switches of the glucose availability per cell were performed, the submodels were not suitable to reflect the time course of fermentation parameters (data not shown). The reason for this is a likely change in the protein expression in the organism, hence altering the concentration of enzymes of submetabolisms in a different way. The yield coefficients (splits to different submetabolims) are changing. A complex model with a proper description capacity of the dynamic system is required. An improvement of the model considering at least a part of the regulatory behavior of the cell in a model which is based on kinetic rate equations, would reduce the systematic error of the present model. The model should include the glycolysis as the different substrate supply is one characteristic of the different phases. Also, the main carbon metabolism including the TCA as a generator of major branch point intermediates and products should be included. Finally, the acetate synthesis should be modeled in detail including the excretion to the environment and the assimilation. Since the intracellular acetate accumulation is regarded as being crucial and a reason for a change in the carbon

distribution in the metabolism, any suitable model should aim to predict the dynamics of acetate synthesis.

8.7 FUTURE WORK

So far, approaches in literature are rarely covering the complete main carbon metabolism including the submetabolisms mentioned previously. A recent paper [210] describes the model fusion for the glycolysis, the pentose-phosphate pathway, the tricarboxylic acid cycle, the glyoxylate shunt. Further regulatory mechanisms were included as the interaction of the cyclic AMP receptor protein, the catabolite repressor (Cra) , the pyruvate dehydrogenase repressor and the repressor for the operon of acetate kinase. The model was applied for transient phases of an *E. coli* batch process. Another own approach combined a model of glycolysis [175] and the TCA cycle[211] and a module describing acetate synthesis and excretion and reconversion. Each submodel is regarded as having several rate limiting reaction steps depending on the substrate and product concentrations present, but are characterized by a similar degree and change of gene expression. This is introduced as a general description factor for each submodel. This assumption is suitable to reduce the set of parameters and to contribute to the fact that many regulatory patterns of the cells are still not understood and cannot be related to single reactions. Also, the accumulation of intermediates, which in most cases cannot be measured accurately, is not considered.

Again the reduction to create a new set of submodels is required. So far, these models are often applied at continuous cultivations of fed batch processes characterized by a high degree of stability concerning the environment of the cells in the bioreactor. Transient changes cannot be predicted but only in a very narrow range close to the conditions of parameter estimation. It is believed that the combination of the consideration of regulatory patterns, model reduction and mechanism recognition can contribute to the great need of process related models that are able to enhance the degree of control.

Besides the model-based contribution, also the integration of cell physiology measurements that are representing multiple parameters of the cells can be suitable to establish reduced models that increase control possibilities. Many attempts are made to establish more robust *on line* and *at line* methods which are based on optical methods (*on line* microscopy) for the determination of cell size and shape, and also electrooptical methods for measuring the polarizability of cells [172, 189]. The information that can be gained from such methods seems to be very suitable to integrate them as activation or repression factors to submodels. In combination with MR, the magnitude of these factors can be assumed separately for each phase of the process while reducing the number or rate equations needed to include all possible ways of interaction at transient cell stages.

9 CONCLUSIONS AND OUTLOOK

9.1 CONCLUSIONS

The principal achievement of this thesis is the proposal of an alternative approach to analysis and implementation of complex rigorous models in processes with restricted state of information. Firstly, it is shown how model simplification may lead to more efficient process description and better understanding of the system. Secondly, the importance of considering the state of during model building is stated. This work proposes some insight to essential questions in biotechnology and chemical engineering:

- How accurate does a model need to be?
- How can we take advantage of the limitations of the model?
- What is the simplest manner to describe dynamic processes?

Although a general answer to these questions is still not at hand, some important advances can be concluded from the results presented in this manuscript. Attention is directed to make models useful and not as accurate or complex as possible. Moreover, it is shown that, if a model is simplified properly based on a systematic analysis and process know how, a better understanding as well as a more accurate description of the system are enabled.

The approach to Mechanism Recognition represents a step forward in what can be considered the right direction toward the future of efficient modeling. Two case studies prove that the use of deep system understanding as the basis for development of simplified models for description of the process offers important advantages in comparison to model building through direct analysis of data correlations and heuristics.

In the first case, accurate detection of organic matter depletion with standard measuring methods is enabled. MR is able to detect depletion of organic matter allowing a better control of the SBR process for WWTPs increasing process efficiency and water quality. In the second case, a model for the description of *E. coli* cultivation is analyzed and compared against experimental data. The submodels created, present a significantly higher identifiability and allow a better analysis of the complex model as well as the cultivation. In both examples the advances of MR are clearly stated offering a systematic methodology for an improved study of the process and its mathematical representation.

Finally, an additional advantage of process description with multiple submodels is a better understanding of the system. As long as the physical background of the

submodels is understood, an analysis of the description capacity of the submodels is also an indicator of the phenomenon dominating the process. This characteristic can be exploited to detect non-measurable states of the process. As shown in both case studies presented in this thesis, special models can be developed to detect non desired conditions which cannot be measured directly. These simplified versions of the complex models offer a better process understanding and an increased applicability for engineering purposes than their complex versions.

9.2 OUTLOOK

The advantages of MR for small and large scale problems have been presented and proved in two case studies. The method is still in a developing stage and further work is required to maximize efficiency of MR. Nevertheless, it is the scope of this thesis to motivate further research in this field.

The essence of this project was to test a new hypothesis, search for possible manners to perform it, and find examples for its application. Because of the nature of this project, its real success can be quantified by the quality and extension of the outlook. In a sense, the real objective of this work may be understood as developing a realistic proposition of possible manners to develop efficient methods to MR. In order to create a complete and efficient MR program able to facilitate modeling, structure analysis, parameter estimation, design of experiments, and ultimately process monitoring in an efficient and systematic manner, further research is required.

9.2.1 GENERAL THEORY FOR SUBMODEL GENERATION

Model reduction theory, offers a handful of techniques to study the dynamics of models and increase its tractability. Very sophisticated mathematical methods are available in literature. In fact, both case studies presented in this work prove that model reduction techniques in combination with process understanding may give rise to simpler but accurate submodels. Still, significant effort is required to build these reduced versions of the complex model. This represents a major drawback in MR application. Building of submodels, as it stands, is not only time consuming but misses the software tools required for the scientists of different fields to collaborate in the creation of proper versions. Despite important achievements in software tools, which enhance this collaboration, e.g. MOSAIC and SBPD toolbox, much time needs to be invested in the translation of codes and experimental information to create a common modeling environment.

In addition, two main aspects of model reduction require further development to reduce the effort of submodel building:

1) Model reduction of dynamic nonlinear systems.
   As previously stated, in nonlinear systems under dynamic conditions a general approach to reduce large nonlinear systems with reasonable computer expenses is still missing. Heuristics and scientist intuition are still the key to the reduction of such process, resulting not only in large efforts, but also in suboptimal and locally restricted solutions with no general validity.
2) Model reduction based on the state of information.
   Present approaches for model reduction do not take the state of information of the process into account. Furthermore, for MR the difference between the quality of the data set obtained in laboratory experiments in comparison to considerably reduced. This information is very important to create an optimal set of submodels and is not consider in the state of the art methods for model reduction.

9.2.2 SWITCHING POINT IDENTIFICATION

At this early development stage, a simple but robust method taken from the field of model-based fault identification was applied to detect the time point were the switch of regimes takes place. The parameter identification method showed to fulfill the requirements at this stage of development, accurately detecting the change of regimes. Nevertheless, this stage offers large optimization potential. A study of modern techniques and a proper adaptation for the case of MR is required to offer an efficient switching point identification approach.

Advanced methods of model predictive control, e.g. moving horizon, should also be tested since they offer alternative approaches for efficient parameter estimation and accurate regime detection in processes with longer time spans.

9.2.3 GLOBAL OPTIMIZATION

In the case of highly nonlinear models, model structure analysis and parameter estimation require proof of global convergence. In this work a stochastic optimization method was applied in an effort to avoid local minima and increase the probability of global solution. Still, stochastic optimization methods cannot assure convergence to global optimality. At its current development stage, global optimizers for nonlinear dynamic systems require large computer expenses and can by no means be applied for online purposes. Still, global optimizers should be implemented at the early stages of MR. Use of global optimizers is essential in model structure analysis, design of

experiments, and model distinguishability studies. It is only after knowing the global best solution for each case that proper conclusion can be made.

9.2.4 ONLINE MONITORING

Finally, this work proves the great potential on the application of MR to enhance monitoring of none measurable variables in dynamic processes. In particular, detection of quasi steady state regimes, which cannot be detected by standard measurements, offers a great field of application. The present approach has been developed taking into account limitations of online applications in all stages. From this, it can be assumed that MR can be applied to industrial processes where a certain regime must be maintained to assure optimal process conditions, but the data obtained does not offer direct information about the state of the "optimal" regime. An example is *E. coli* cultivations where optimal growth is known to occur under substrate limitation where the state of the cell cannot be measured directly. Still, indirect measurements in combination with the correct set of submodels enable the detection of a change in the state of the cells optimizing feed strategies to maximize cell density and recombinant protein production.

10 Appendix

A. KINETIC MODEL FOR *E. COLI*

MODEL AS PUBLISHED BY LIN [91]

$$\frac{dV}{dt} = F \quad \textbf{A 1}$$

$$\frac{dX}{dt} = \left(-\frac{F}{V} + \mu\right) X \quad \textbf{A 2}$$

$$\frac{dS}{dt} = -\frac{F}{V}(S - S_i) - q_S * X \quad \textbf{A 3}$$

$$\frac{dDOT}{dt} = k_{La}(DOT^* - DOT) - q_O * X * H \quad \textbf{A 4}$$

$$\frac{dAc}{dt} = -\frac{F}{V}A + \left(q_{Ac_p} - q_{Ac_c}\right) X \quad \textbf{A 5}$$

$$\frac{dr_S}{dt} = Y_{E_S/X} * \mu - r_S(q_{E_{Sd}} + \mu) \quad \textbf{A 6}$$

$$\frac{dr_O}{dt} = Y_{E_O/X} * \mu - r_O(q_{E_{Od}} + \mu) \quad \textbf{A 7}$$

The initial values

$$r_{S0} = c_S * \frac{\mu_{max}}{q_{ESd} + \mu_{max}} \quad \textbf{A 8}$$

$$r_{O0} = c_O * \frac{\mu_{max}}{q_{EOd} + \mu_{max}} \quad \textbf{A 9}$$

$$\mu_{max} = \left(q_{S_{max}} - q_m\right) Y_{x/Sox} \quad \textbf{A 10}$$

Calculation of the subsidiary variables:

$$q_{S_{cap}} = q_{S_{max}} \frac{r_S}{r_{S0}} \quad \textbf{A 11}$$

$$q_{O_{cap}} = q_{O_{max}} \frac{r_O}{r_{O0}} \quad \textbf{A 12}$$

$$q_S = q_{S_{cap}} \frac{S}{S + K_S} \quad \textbf{A 13}$$

$$q_{S_{Ox}} = q_S \quad \textbf{A 14}$$

$$q_{mr} = \min[q_m, q_{S_{ox}}] \quad \textbf{A 15}$$

$$q_{S_{oxan}} = \left(q_{S_{ox}} - q_{mr}\right) Y_{x/S_{ox}} * C_S \quad \textbf{A 16}$$

$$q_{S_{oxen}} = q_{S_{ox}} - q_{S_{oxan}} \quad \textbf{A 17}$$

$$q_{O_S} = q_{S_{ox\,en}} * Y_{O/S} \quad \textbf{A 18}$$

To avoid a larger q_{O_S} than allowed by the maximal oxygen uptake capacity of the bacteria, whenever $q_{O_S} > q_{O_{cap}}$ we change the values to $q_{O_S} = q_{O_{cap}}$ and recalculate:

$$q_{S_{oxen}} = \frac{q_{O_{cap}}}{Y_{O/S}} \quad \textbf{A 19}$$

$$q_{S_{ox}} = \frac{q_{mr} * C_S * Y_{X/S\_ox} - q_{S_{oxen}}}{C_S * Y_{O/S_{ox}} - 1} \quad \textbf{A 20}$$

$$q_{S_{oxan}} = q_{S_{ox}} - q_{S_{oxe\,n}} \quad \textbf{A 21}$$

Because of the recalculations, the physical limitation of the real maintenance coefficient has to be checked once more. If $q_{mr} > q_{S_{ox}}$, then $q_{mr} = q_{S_{ox}} = q_{S_{oxen}}$ and $q_{S_{oxan}} = 0$

$$q_{S_{of}} = q_S - q_{S_{ox}} \quad \textbf{A 22}$$

$$q_{A_C} = q_{Ac_{max}} \frac{Ac}{Ac + K_{Ac}} \quad \textbf{A 23}$$

$$q_{S_{ofan}} = q_{Sof} * Y_{X/S_{of}} * C_S \quad \textbf{A 24}$$

$$q_{S_{ofen}} = q_{S_{of}} - q_{S_{ofan}} \quad \textbf{A 25}$$

$$q_{A_{can}} = q_{Ac} * Y_{X/A} * C_{Ac} \quad \textbf{A 26}$$

$$q_{A_{cen}} = q_{A_c} - q_{A_{can}} \quad \textbf{A 27}$$

Finally, if $q_{A_{cen}} > q_{O_{cap}} - q_{O_S}/Y_{OA}$ then, $q_{A_{cen}} = q_{O_{cap}} - q_{O_S}/Y_{OA}$ and:

$$q_{Ac_c} = \frac{q_{A_{cen}}}{1 - Y_{X/A} * C_{Ac}} \quad \textbf{A 28}$$

$$q_{A_{can}} = q_{A_c} - q_{A_{can}} \quad \textbf{A 29}$$

$$q_{A_p} = q_{S_{ofen}} * Y_{Ac/S} \quad \textbf{A 30}$$

$$q_O = q_{O_S} + q_{A_{cen}} * Y_{O/Ac} \quad \textbf{A 31}$$

$$\mu = \left(q_{S_{ox}} - q_{mr}\right) Y_{X/S_{ox}} + q_{S_{of}} * Y_{X/S_{of}} + q_{Ac_c} * Y_{X/Ac} \quad \textbf{A 32}$$

Nomenclature

- μ specific growth rate [h^{-1}]
- Ac acetate concentration [g L^{-1}]
- DOT dissolved oxygen tension [%]
- E_S, E_O concentration of the enzymes substrate/oxygen consumption [gL^{-1}]
- F feed flow rate [L h^{-1}]
- H constant derived from the Henry's law [14,000 L g^{-1}]
- K saturation constant [g L^{-1}]
- k_{La} volumetric oxygen transfer coefficient [h^{-1}]
- q_{Ac} specific acetate production of consumption rate [g g^{-1} h^{-1}]
- q_i specific degradation rate [g g^{-1} h^{-1}]
- q_O specific oxygen uptake rate [g g^{-1} h^{-1}]
- q_S specific substrate (glucose) uptake rate [g $g^{-1}h^{-1}$]
- r_O, r_S ratio of enzyme for oxygen or substrate consumption per biomass
- S substrate (glucose) concentration [g L^{-1}]
- t cultivation time [h]
- V fermenter volume [L]
- X biomass concentration [g L^{-1}]
- Y yield [g g^{-1}]

Subscripts

- 0 Initial concentration
- Ac acetate
- an anabolic
- c consumption
- cap capacity
- d degradation
- E enzyme
- en energetic

- i inlet concentration
- m maintenance
- max maximum
- mr real maintenance
- O oxygen
- of overflow
- ox oxidative
- p production
- S substrate (glucose)
- X biomass

B. OVERFLOW SUBMODEL

The nomenclature is identical to the complete Lin model (appendix A)

$$\frac{dV}{dt} = F \quad \textbf{B 1}$$

$$\frac{dX}{dt} = \left(-\frac{F}{V} + \mu\right) X \quad \textbf{B 2}$$

$$\frac{dS}{dt} = -\frac{F}{V}(S - S_i) - q_S * X \quad \textbf{B 3}$$

$$\frac{dDOT}{dt} = k_{La}(DOT^* - DOT) - q_O * X * H \quad \textbf{B 4}$$

$$\frac{dA}{dt} = -\frac{F}{V}Ac + \left(q_{Ac_p} - q_{Ac_c}\right) X \quad \textbf{B 5}$$

$$r_S = c_S * \frac{\mu_{max}}{q_{ESd} + \mu_{max}} \quad \textbf{B 6}$$

$$r_O = c_O * \frac{\mu_{max}}{q_{EOd} + \mu_{max}} \quad \textbf{B 7}$$

The initial values

$$\mu_{max} = \left(q_{S_{max}} - q_m\right) Y_{x/Sox} \quad \textbf{B 8}$$

CALCULATION OF THE SUBSIDIARY VARIABLES

$$q_{S_{cap}} = q_{S_{max}} \frac{r_S}{r_{S0}} \quad \textbf{B 9}$$

$$q_{O_{cap}} = q_{O_{max}} \frac{r_O}{r_{O0}} \quad \textbf{B 10}$$

$$q_S = q_{S_{cap}} \frac{S}{S + K_S} \quad \textbf{B 11}$$

$$q_{S_{Ox}} = q_S \quad \textbf{B 12}$$

$$q_{mr} = q_m \quad \textbf{B 13}$$

Because of the overflow condition, we assume: $q_{O_S} = q_{O_{cap}}$:

$$q_{S_{oxen}} = \frac{q_{O_{cap}}}{Y_{O/S}} \quad \textbf{B 14}$$

$$q_{S_{ox}} = \frac{q_{mr} * C_S * Y_{X/S\_ox} - q_{S_{oxen}}}{C_S * Y_{O/S_{ox}} - 1} \quad \textbf{B 15}$$

$$q_{S_{oxan}} = q_{S_{ox}} - q_{S_{oxen}} \quad \textbf{B 16}$$

$$q_{S_{of}} = q_S - q_{S_{ox}} \quad \textbf{B 17}$$

$$q_{Ac_c} = q_{Ac_c max} \frac{Ac}{Ac + K_{Ac}} \quad \textbf{B 18}$$

$$q_{S_{ofan}} = q_{Sof} * Y_{X/S_{of}} * C_S \quad \textbf{B 19}$$

$$q_{S_{ofen}} = q_{S_{of}} - q_{S_{ofan}} \quad \textbf{B 20}$$

$$q_{Ac_{can}} = q_{Ac_c} * Y_{X/Ac} * C_{Ac} \quad \textbf{B 21}$$

$$q_{Ac_{cen}} = q_{Ac_c} - q_{Ac_{can}} \quad \textbf{B 22}$$

$$\mu = \left(q_{S_{ox}} - q_{mr}\right) Y_{X/S_{ox}} + q_{S_{of}} * Y_{X/S_{of}} + q_{Ac_c} * Y_{X/Ac} \quad \textbf{B 23}$$

C. SUBSTRATE LIMITATION SUBMODEL

The nomenclature is identical to the complete Lin model (appendix A)

$$\frac{dV}{dt} = F \quad \textbf{C 1}$$

$$\frac{dX}{dt} = \left(-\frac{F}{V} + \mu\right) X \quad \textbf{C 2}$$

$$\frac{dS}{dt} = -\frac{F}{V}(S - S_i) - q_S * X \quad \textbf{C 3}$$

$$\frac{dDOT}{dt} = k_{La}(DOT^* - DOT) - q_O * X * H \quad \textbf{C 4}$$

$$\frac{dAc}{dt} = -\frac{F}{V}Ac + \left(q_{Ac_p} - q_{Ac_c}\right) X \quad \textbf{C 5}$$

$$\frac{dr_S}{dt} = Y_{E_S/X} * \mu - r_S(q_{E_{Sd}} + \mu) \quad \textbf{C 6}$$

$$\frac{dr_O}{dt} = Y_{E_O/X} * \mu - r_O(q_{E_{Od}} + \mu) \quad \textbf{C 7}$$

The initial values

$$r_{S0} = c_S * \frac{\mu_{max}}{q_{ESd} + \mu_{max}} \quad \textbf{C 8}$$

$$r_{O0} = c_O * \frac{\mu_{max}}{q_{EOd} + \mu_{max}} \quad \textbf{C 9}$$

$$\mu_{max} = \left(q_{S_{max}} - q_m\right) Y_{x/Sox} \quad \textbf{C 10}$$

Calculation of the subsidiary variables:

$$q_{S_{cap}} = q_{S_{max}} \frac{r_S}{r_{S0}} \quad \textbf{C 11}$$

$$q_{O_{cap}} = q_{O_{max}} \frac{r_O}{r_{O0}} \quad \textbf{C 12}$$

$$q_S = q_{S_{cap}} \frac{S}{S + K_S} \quad \textbf{C 13}$$

$$q_{S_{Ox}} = q_S \quad \textbf{C 14}$$

$$q_{mr} = q_m \quad \textbf{C 15}$$

$$q_{S_{oxan}} = \left(q_{S_{ox}} - q_{mr}\right) Y_{x/S_{ox}} * C_S \quad \textbf{C 16}$$

$$q_{S_{oxen}} = q_{S_{ox}} - q_{S_{oxan}} \quad \textbf{C 17}$$

$$q_{O_S} = q_{S_{oxen}} * Y_{O/S} \quad \textbf{C 18}$$

$$q_{S_{of}} = 0 \quad \textbf{C 19}$$

$$q_{Ac_C} = q_{Acc_{max}} \frac{Ac}{Ac + K_{Ac}} \quad \textbf{C 20}$$

$$q_{S_{ofan}} = 0 \quad \textbf{C 21}$$

$$q_{S_{ofen}} = 0 \quad \textbf{C 22}$$

$$q_{Ac_{can}} = q_{Ac_c} * Y_{X/Ac} * C_{Ac} \quad \textbf{C 23}$$

$$q_{Ac_{cen}} = q_{Ac_c} - q_{Ac_{can}} \quad \textbf{C 24}$$

$$q_{Ac_p} = 0 \quad \textbf{C 25}$$

$$q_O = q_{O_S} + q_{Ac_{cen}} * Y_{O/Ac} \quad \textbf{C 26}$$

$$\mu = \left(q_{S_{ox}} - q_{mr}\right) Y_{X/S_{ox}} + q_{S_{of}} * Y_{X/S_{of}} + q_{Ac_c} * Y_{X/Ac} \quad \textbf{C 27}$$

D. STARVATION SUBMODEL

The nomenclature is identical to the complete Lin model (appendix A)

$$\frac{dV}{dt} = F \quad \textbf{D 1}$$

$$\frac{dX}{dt} = \left(-\frac{F}{V} + \mu\right) X \quad \textbf{D 2}$$

$$\frac{dS}{dt} = -\frac{F}{V}(S - S_i) - q_S * X \quad \textbf{D 3}$$

$$\frac{dDOT}{dt} = k_{La}(DOT^* - DOT) - q_O * X * H \quad \textbf{D 4}$$

$$\frac{dAc}{dt} = -\frac{F}{V}Ac + \left(q_{Ac_p} - q_{Ac_c}\right) X \quad \textbf{D 5}$$

$$r_S : estimated \quad \textbf{D 6}$$

$$r_O : estimated \quad \textbf{D 7}$$

The initial values

$$\mu_{max} = \left(q_{S_{max}} - q_m\right) Y_{x/Sox} \quad \textbf{D 8}$$

Calculation of the subsidiary variables:

$$q_{S_{cap}} = q_{S_{max}} \frac{r_S}{r_{S0}} \quad \textbf{D 9}$$

$$q_{O_{cap}} = q_{O_{max}} \frac{r_O}{r_{O0}} \quad \textbf{D 10}$$

$$q_S = q_{S_{cap}} \frac{S}{S + K_S} \quad \textbf{D 11}$$

$$q_{S_{Ox}} = q_S \quad \textbf{D 12}$$

$$q_{mr} = q_m \quad \textbf{D 13}$$

$$q_{S_{oxan}} = \left(q_{S_{ox}} - q_{mr}\right) Y_{x/S_{ox}} * C_S \quad \textbf{D 14}$$

$$q_{S_{oxen}} = \frac{q_{O_{cap}}}{Y_{O/S}} \quad \textbf{D 15}$$

$$q_{S_{ox}} = \frac{q_{mr} * C_S * Y_{X/S\_ox} - q_{S_{oxen}}}{C_S * Y_{O/S_{ox}} - 1} \quad \textbf{D 16}$$

$$q_{S_{oxan}} = q_{S_{ox}} - q_{S_{ox\,en}} \quad \textbf{D 17}$$

$$q_{S_{of}} = 0 \quad \textbf{D 18}$$

$$q_{Ac_c} = 0 \quad \textbf{D 19}$$

$$q_{S_{ofan}} = 0 \quad \textbf{D 20}$$

$$q_{S_{ofen}} = 0 \quad \textbf{D 21}$$

$$q_{Ac_{can}} = 0 \quad \textbf{D 22}$$

$$q_{Ac_{cen}} = 0 \quad \textbf{D 23}$$

$$\mu = \left(q_{S_{ox}} - q_{mr}\right) Y_{X/S_{ox}} + q_{S_{of}} * Y_{X/S_{of}} + q_{Ac_c} * Y_{X/Ac} \quad \textbf{D 24}$$

BIBLIOGRAPHY

PERSONAL PUBLICATIONS

Electrooptical monitoring of cell polarizability and cell size in aerobic Escherichia coli batch cultivations
Stefan Junne, M. Nicolas Cruz-Bournazou, Alexander Angersbach and Peter Götz
Journal of Industrial Microbiology & Biotechnology
Volume 37, Number 9, 935-942, DOI: 10.1007/s10295-010-0742-5

From data to mechanistic-based models
M. N. Cruz-Bournazou, H. Arellano-Garcia, J. Schöneberger, G. Wozny
Chemical and Process Engineering
2009, 30, 291–306

A Novel Approach to Mechanism Recognition in Escherichia Coli Fed-Batch Fermentations
M.N. Cruz Bournazou, H. Arellano-Garcia, J.C. Schöneberger, S. Junne, P. Neubauer and G. Wozny
Computer Aided Chemical Engineering
Volume 27, 2009, Pages 651-656 10th

Model reduction of the ASM3 extended for two-step nitrification
M. N. Cruz Bournazou H. Arellano-Garcia G. Wozny, G. Lyberatos, C. Kravaris
Computer Applications in Biotechnology
Volume11 Part 1, 10.3182/20100707-3-BE-2012.0047

Neuer Ansatz zur Mechanismenerkennung in E. coli Fed-Batch-Kultivierungen
M. N. Cruz Bournazou, H. Arellano-Garcia, S. Junne, G. Wozny, P. Neubauer
Chemie Ingenieur Technik
Volume 82, Issue 9, page 1503, September, 2010,DOI: 10.1002/cite.201050496

Modellreduktion des erweiterten ASM3-Modells für die zweistufige Nitrifikation
M. N. Cruz Bournazou M. Sc., H. Arellano-Garcia Dr.-Ing., G. Wozny Prof. , G. Lyberatos Prof., C. Kravaris Prof.
Chemie Ingenieur Technik
Volume 82, Issue 9, page 1540, September, 2010,DOI: 10.1002/cite.201050496

ASM3 extended for two-step nitrification-denitrification: a model reduction for Sequencing Batch Reactors
M. N. Cruz Bournazou, H. Arellano-Garcia, G. Wozny, G. Lyberatos, C. Kravaris
Journal of Chemical Technology and Biotechnology (in press: JCTB 11-0602)

REFERENCES

1. FTD, *BASF erhöht Forschungsausgaben* ed. F.T. Deutschland. 2011, http://www.ftd.de/unternehmen/industrie/:investitionen-beim-chemiekonzern-basf-erhoeht-forschungsausgaben/60007679.html
2. DFG, *DFG-Jahresbericht 2010*, ed. http://www.alphagalileo.org/ViewItem.aspx?ItemId=108368&CultureCode=de. 2010.
3. Barton, P.I., *The modelling and simulation of combined discrete/continuous processes.* PhD diss., University of London, 1992.
4. fda.gov, *Guidance for Industry Process Validation: General Principles and Practice*. 2011.
5. Kitano, H., *Systems Biology: A Brief Overview*. 2002. p. 1662-1664.
6. Kitano, H., *Computational systems biology.* Nature, 2002. **420**(6912): p. 206-210.
7. Ropers, D., V. Baldazzi, and H. de Jong, *Model Reduction Using Piecewise-Linear Approximations Preserves Dynamic Properties of the Carbon Starvation Response in Escherichia coli.* IEEE IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2009.
8. Box, G.E.P. and W.J. Hill, *Discrimination among mechanistic models.* Technometrics, 1967. **9**(1): p. 57-71.
9. Preisig, H.A., *Constructing and maintaining proper process models.* Computers & Chemical Engineering, 2010. **34**(9): p. 1543-1555.
10. Li, D., O. Ivanchenko, N. Sindhwani, E. Lueshen, and A.A. Linninger, *Optimal Catheter Placement for Chemotherapy.* Computer Aided Chemical Engineering. **28**: p. 223-228.
11. Brendel, M., D. Bonvin, and W. Marquardt, *Incremental identification of kinetic models for homogeneous reaction systems.* Chemical Engineering Science, 2006. **61**(16): p. 5404-5420.
12. Balsa-Canto, E., A.A. Alonso, and J.R. Banga, *An iterative identification procedure for dynamic modeling of biochemical networks.* BMC Systems Biology, 2010. **4**(1): p. 11.
13. Rippin, D.W.T., *Batch process systems engineering: a retrospective and prospective review.* Computers & chemical engineering, 1993. **17**: p. 1-13.
14. Kravaris, C., R.A. Wright, and J.F. Carrier, *Nonlinear controllers for trajectory tracking in batch processes.* Computers & chemical engineering, 1989. **13**(1-2): p. 73-82.
15. Peterson, C.L., M. Feldman, R. Korus, and D.L. Auld, *Batch type transesterification process for winter rape oil.* Applied engineering in agriculture (USA), 1991. **110**(3): p. 971-976.
16. Choe, Y.S. and W.L. Luyben, *Rigorous dynamic models of distillation columns.* Industrial & Engineering Chemistry Research, 1987. **26**(10): p. 2158-2161.
17. Gep, B.O.X. and N.R. Draper, *Empirical model building and response surfaces*. 1987, John Wiley and Sons, New York, USA.
18. Atkinson, A.C. and A.N. Donev, *Optimum Experimental Designs*. 1992: Oxford University Press.
19. Sjoberg, J., Q.H. Zhang, L. Ljung, A. Benveniste, B. Delyon, P.Y. Glorennec, H. Hjalmarsson, and A. Juditsky, *Nonlinear Black-Box Modeling in System-Identification-A Unified Overview.* Automatica, 1995. **31**(12): p. 1691-1724.
20. Suykens, J.A.K. and J. Vandewalle, *Nonlinear Modeling: Advanced Black-Box Techniques*. 1998: Kluwer Academic Publishers.
21. Jones, D.R., M. Schonlau, and W.J. Welch, *Efficient Global Optimization of Expensive Black-Box Functions.* Journal of Global Optimization, 1998. **13**(4): p. 455-492.
22. Verheijen, P.J.T., *Model Selection: an overview of practices in chemical engineering.* Dynamic Model Development, 2003: p. 85–104.
23. Frank, P.M., *Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: A survey and some new results.* Automatica, 1990. **26**(3): p. 459-474.
24. Tarantola, A., *Inverse Problem Theory and Methods for Model Parameter Estimation*. 2005, Society for Industrial Mathematics.
25. Esposito, W.R. and C.A. Floudas, *Global optimization for the parameter estimation of differential-algebraic systems.* Ind. Eng. Chem. Res, 2000. **39**(5): p. 1291-1310.

26. Singer, A.B., J.W. Taylor, P.I. Barton, and W.H. Green, *Global dynamic optimization for parameter estimation in chemical kinetics*. The Journal of Physical Chemistry A, 2006. **110**(3): p. 971-976.
27. Mitsos, A., G.M. Oxberry, P.I. Barton, and W.H. Green, *Optimal automatic reaction and species elimination in kinetic mechanisms*. Combustion and Flame, 2008. **155**(1-2): p. 118–132.
28. Herold, S., T. Heine, and R. King. *An Automated Approach to Build Process Models by Detecting Biological Phenomena in (fed-) batch Experiments*. 2010, Computer Applications in Biotechnology, **11**(1).
29. Forbus, K.D., *Qualitative Process Theory*. Artificial Intelligence, 1984. **24**(1-3): p. 85-168.
30. Forbus, K.D., *Qualitative process theory: twelve years after*. Artificial Intelligence in Perspective, 1994. **59**(1): p. 15-123.
31. Blom, H.A.P. and Y. Bar-Shalom, *The interacting multiple model algorithm for systems with Markovian switching coefficients*. Automatic Control, IEEE Transactions on, 1988. **33**(8): p. 780-783.
32. Mazor, E., A. Averbuch, Y. Bar-Shalom, and J. Dayan, *Interacting multiple model methods in target tracking: a survey*. Aerospace and Electronic Systems, IEEE Transactions on, 1998. **34**(1): p. 103-123.
33. Doucet, A., N.J. Gordon, and V. Krishnamurthy, *Particle filters for state estimation of jump Markov linear systems*. Signal Processing, IEEE Transactions on, 2001. **49**(3): p. 613-624.
34. Zhang, W., C. Wu, and C. Wang, *Qualitative Algebra and Graph Theory Methods for Dynamic Trend Analysis of Continuous System*. Chinese Journal of Chemical Engineering, 2011. **19**(2): p. 308-315.
35. Berleant, D. and B.J. Kuipers, *Qualitative and quantitative simulation: bridging the gap*. Artificial Intelligence, 1997. **95**(2): p. 215-255.
36. Utkin, V., *Variable structure systems with sliding modes*. Automatic Control, IEEE Transactions on, 1977. **22**(2): p. 212-222.
37. Bhattacharjee, B., D.A. Schwer, P.I. Barton, and W.H. Green, *Optimally-reduced kinetic models: reaction elimination in large-scale kinetic mechanisms*. Combustion and Flame, 2003. **135**(3): p. 191-208.
38. Escobet, T., D. Feroldi, S. De Lira, V. Puig, J. Quevedo, J. Riera, and M. Serra, *Model-based fault diagnosis in PEM fuel cell systems*. Journal of Power Sources, 2009. **192**(1): p. 216-223.
39. De Lira Ramirez, S., V. Puig Cayuela, A.P. Husar, and J.J. Quevedo Casin. *LPV model-based fault diagnosis using relative fault sensitivity signature approach in a PEM fuel cell*. 2010: IEEE Computer Society.
40. Hotop, R., S. Ochs, and T. Ross, *Überwachung von Anlagenteilen: Neues Werkzeug ermöglicht Statusermittlung*. atp edition, 2010.
41. Dochain, D. and M. Perrier, *Dynamical modelling, analysis, monitoring and control design for nonlinear bioprocesses*. Biotreatment, Downstream Processing and Modelling, 1997: p. 147-197.
42. Cruz-Bournazou, M., H. Arellano-Garcia, J. Schoneberger, and G. Wozny, *From data to mechanistic-based models*. Chemical and Process Engineering Selected full texts, 2009. **30**(2): p. 291-306.
43. Schöneberger, J., H. Arellano-Garcia, and G. Wozny, *A Novel Approach to Automated Mechanism Recognition based on Model Discrimination*. Chemical Engineering Transactions, 2007. **12**
44. Kuntsche, S., T. Barz, R. Kraus, H. Arellano-Garcia, and G. Wozny, *MOSAIC a web-based modeling environment for co de generation*. Computers & Chemical Engineering, 2011. **35** (11): p.2257-2273.
45. Schmidt, H., M.F. Madsen, S. Dano, and G. Cedersund, *Complexity reduction of biochemical rate expressions*. Bioinformatics, 2008. **24**(6): p. 848.
46. Barz, T., S. Kuntsche, G. Wozny, and H. Arellano-Garcia, *An efficient sparse approach to sensitivity generation for large-scale dynamic optimization*. Computers and Chemical Engineering, 2010. **http://dx.doi.org/10.1016/j.compchemeng.2010.10.008**.

47. Violet, N., E. Fischer, T. Heine, and R. King. *A Software Supported Compartmental Modeling Approach for Paenibacillus Polymyxa*. Computer Applications in Biotechnology, 2010. **11**(1).
48. Silvert, W., *Modelling as a discipline*. International Journal of General Systems, 2001. **30**(3): p. 261-282.
49. Aris, R., *Mathematical modeling: a chemical engineer's perspective*. Vol. 1. 1999: Academic Pr.
50. Bayer, B. and W. Marquardt, *Towards integrated information models for data and documents*. Computers & chemical engineering, 2004. **28**(8): p. 1249-1266.
51. Dochain, D., *State and parameter estimation in chemical and biochemical processes: a tutorial*. Journal of process control, 2003. **13**(8): p. 801-818.
52. Stewart, W.E., Y. Shon, and G.E.P. Box, *Discrimination and goodness of fit of multiresponse mechanistic models*. AIChE journal, 1998. **44**(6): p. 1404-1412.
53. Freyre-Gonzalez, J.A. and L.G. Trevino-Quintanilla, *Analyzing Regulatory Networks in Bacteria* Nature Educations, 2010. **3**(9): p. 24.
54. Hady, L. and G. Wozny, *Computer-aided web-based application to modular plant design*. Computer Aided Chemical Engineering, 2010. **28**: p. 685-690.
55. Bailey, J.E., *Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities*. Biotechnology progress, 1998. **14**(1): p. 8-20.
56. Kremling, A., S. Fischer, K. Gadkar, F.J. Doyle, T. Sauter, E. Bullinger, F. Allgöwer, and E.D. Gilles, *A Benchmark for Methods in Reverse Engineering and Model Discrimination: Problem Formulation and Solutions*. Genome Research, 2004. **14**: p. 1773-1785.
57. Csete, M.E. and J.C. Doyle, *Reverse engineering of biological complexity*. Science, 2002. **295**(5560): p. 1664.
58. Bardow, A. and W. Marquardt, *Incremental and simultaneous identification of reaction kinetics: methods and comparison*. Chemical Engineering Science, 2004. **59**(13): p. 2673-2684.
59. Kitano, H., *Systems biology: a brief overview*. Science, 2002. **295**(5560): p. 1662.
60. Spieth, C., N. Hassis, and F. Streichert. *Comparing mathematical models on the problem of network inference*. ACM, 2006.
61. Hecker, M., S. Lambeck, S. Toepfer, E. Van Someren, and R. Guthke, *Gene regulatory network inference: Data integration in dynamic models--A review*. Biosystems, 2009. **96**(1): p. 86-103.
62. Heinrich, R. and S. Schuster, *The regulation of cellular systems*. 1996: Springer Us.
63. Vilela, M., I.C. Chou, S. Vinga, A. Vasconcelos, E. Voit, and J. Almeida, *Parameter optimization in S-system models*. BMC systems biology, 2008. **2**(1): p. 35.
64. Stamatelatou, K., L. Syrou, C. Kravaris, and G. Lyberatos, *An invariant manifold approach for CSTR model reduction in the presence of multi-step biochemical reaction schemes. Application to anaerobic digestion*. Chemical Engineering Journal, 2009. **150**(2-3): p. 462-475.
65. Box, G.E.P., *Statistics as a catalyst to learning by scientific method part II-a discussion*. Journal of Quality Technology, 1999. **31**(1): p. 16-29.
66. Raduly, B., A.G. Capodaglio, and D.A. Vaccari, *Simplification of wastewater treatment plant models using empirical modelling techniques*. Young Researchers 2004, 2004: p. 51.
67. Okino, M.S. and M.L. Mavrovouniotis, *Simplification of mathematical models of chemical reaction systems*. Chem. Rev, 1998. **98**(2): p. 391-408.
68. Fraser, S.J., *The steady state and equilibrium approximations: A geometrical picture*. The Journal of Chemical Physics, 1988. **88**: p. 4732.
69. Kuo, J.C.W. and J. Wei, *Lumping Analysis in Monomolecular Reaction Systems. Analysis of Approximately Lumpable System*. Industrial & Engineering chemistry fundamentals, 1969. **8**(1): p. 124-133.
70. Kaufmann, J., H. Weiss, and Schering (Berlin West), *Statistische Methoden in der experimentellen Forschung Kolloquium, Berlin Kurzfassungen von Vorträgen der Wintersemester 1988/89, 1989/90 und 1990/91*. 1991, Berlin: s.n. 254 S.
71. Liao, J.C. and J. Delgado, *Advances in metabolic control analysis*. Biotechnology Progress, 1993. **9**(3): p. 221-233.
72. Roussel, M.R. and S.J. Fraser, *Invariant manifold methods for metabolic model reduction*. Chaos, 2001. **11**: p. 196-206.

73. Kazantzis, N., C. Kravaris, and L. Syrou, *A new model reduction method for nonlinear dynamical systems.* Nonlinear Dynamics, 2010. **59**(1): p. 183-194.
74. Ranzi, E., M. Dente, A. Goldaniga, G. Bozzano, and T. Faravelli, *Lumping procedures in detailed kinetic modeling of gasification, pyrolysis, partial oxidation and combustion of hydrocarbon mixtures.* Progress in energy and combustion science, 2001. **27**(1): p. 99.
75. Dokoumetzidis, A. and L. Aarons, *Proper lumping in systems biology models.* Systems Biology, IET, 2009. **3**(1): p. 40-51.
76. Segel, L.A. and M. Slemrod, *The quasi-steady-state assumption: a case study in perturbation.* SIAM review, 1989. **31**(3): p. 446-477.
77. Danos, V., J. Feret, W. Fontana, R. Harmer, and J. Krivine. *Abstracting the ODE semantics of rule-based models: Exact and automatic model reduction*. 2009.
78. Kim, Y.H., C.K. Yoo, and I.B. Lee, *Optimization of biological nutrient removal in a SBR using simulation-based iterative dynamic programming.* Chemical Engineering Journal, 2008. **139**(1): p. 11-19.
79. Kevrekidis, I.G., C.W. Gear, J.M. Hyman, P.G. Kevrekidid, O. Runborg, and C. Theodoropoulos, *Equation-Free, Coarse-Grained Multiscale Computation: Enabling Mocroscopic Simulators to Perform System-Level Analysis.* Communications in Mathematical Sciences, 2003. **1**(4): p. 715-762.
80. Nagy, T. and T. TurÃ¡nyi, *Reduction of very large reaction mechanisms using methods based on simulation error minimization.* Combustion and Flame, 2009. **156**(2): p. 417-428.
81. Franceschini, G. and S. Macchietto, *Model-based design of experiments for parameter precision: State of the art.* Chemical Engineering Science, 2008, **63** (19): p 4846-4872.
82. Georgescu, R., C.R. Berger, P. Willett, M. Azam, and S. Ghoshal. *Comparison of data reduction techniques based on the performance of SVM-type classifiers*: IEEE, 2010.
83. Witten, I.H. and E. Frank, *Data Mining: Practical machine learning tools and techniques*. 2005: Morgan Kaufmann.
84. Korkel, S., E. Kostina, H.G. Bock, and J.P. Schloder, *Numerical Methods for Optimal Control Problems in Design of Robust Optimal Experiments for Nonlinear Dynamic Processes.*
85. Küpper, A., M. Diehl, J.P. Schlöder, H.G. Bock, and S. Engell, *Efficient moving horizon state and parameter estimation for SMB processes.* Journal of Process Control, 2009. **19**(5): p. 785-802.
86. Barz, T., V. Löffler, H. Arellano-Garcia, and G. Wozny, *Optimal determination of steric mass action model parameters for [beta]-lactoglobulin using static batch experiments.* Journal of Chromatography A.
87. Franceschini, G. and S. Macchietto, *Validation of a model for biodiesel production through model-based experiment design.* Industrial & Engineering Chemistry Research, 2007. **46**(1): p. 220-232.
88. Arlot, S. and P. Massart, *Data-driven calibration of penalties for least-squares regression.* The Journal of Machine Learning Research, 2009. **10**: p. 245-279.
89. Asprey, S.P. and S. Macchietto, *Statistical tools for optimal dynamic model building.* Computers and Chemical Engineering, 2000. **24**(2-7): p. 1261-1267.
90. Binder, K. and D.W. Heermann, *Monte Carlo simulation in statistical physics: an introduction*: Springer.
91. Lin, H.Y., B. Mathiszik, B. Xu, S.O. Enfors, and P. Neubauer, *Determination of the maximum specific uptake capacities for glucose and oxygen in glucose-limited fed-batch cultivations of Escherichia coli.* Biotechnol Bioeng, 2001. **73**(5): p. 347-57.
92. Korkel, D.M.S., *Numerische Methoden fuer Optimale Versuchsplanungsprobleme bei nichtlinearen DAE-Modellen*, in *Ruprecht-Karls-Universitaet Heidelberg*. 2002.
93. Kerr, M.K. and G.A. Churchill, *Experimental design for gene expression microarrays.* Biostatistics, 2001. **2**(2): p. 183-201.
94. Gadkar, K.G., R. Gunawan, and F.J. Doyle Iii, *Iterative approach to model identification of biological networks.* BMC Bioinformatics, 2005. **6**(155).

95. Felix Oliver Lindner, P. and B. Hitzmann, *Experimental design for optimal parameter estimation of an enzyme kinetic process based on the analysis of the Fisher information matrix.* Journal of Theoretical Biology, 2006. **238**(1): p. 111-123.
96. Imhof, L.A., *Maximin designs for exponential growth models and heteroscedastic polynomial models.* Ann. Statist, 2001. **29**(2): p. 561-576.
97. Cooney, M.J. and K.A. McDonald, *Optimal dynamic experiments for bioreactor model discrimination.* Applied Microbiology and Biotechnology, 1995. **43**(5): p. 826-837.
98. Swinnen, I.A.M., K. Bernaerts, K. Gysemans, and J.F. Van Impe, *Quantifying microbial lag phenomena due to a sudden rise in temperature: a systematic macroscopic study.* International Journal of Food Microbiology, 2005. **100**(1-3): p. 85-96.
99. Valdramidis, V.P., A.H. Geeraerd, J.E. Gaze, A. Kondjoyan, A.R. Boyd, H.L. Shaw, and J.F.V. Impe, *Quantitative description of Listeria monocytogenes inactivation kinetics with temperature and water activity as the influencing factors; model prediction and methodological validation on dynamic data.* Journal of Food Engineering, 2006. **76**(1): p. 79-88.
100. Hung, K.K., P.K. Ko, C. Hu, and Y.C. Cheng, *A physics-based MOSFET noise model for circuit simulators.* Electron Devices, IEEE Transactions on, 1990. **37**(5): p. 1323-1333.
101. Patel, S. and K.K. Pant, *Experimental study and mechanistic kinetic modeling for selective production of hydrogen via catalytic steam reforming of methanol.* Chemical Engineering Science, 2007. **62**(18-20): p. 5425-5435.
102. Box, G.E.P. and H.L. Lucas, *design of experiments in non-linear situations.* Biometrika, 1959. **46**(1-2): p. 77-90.
103. Box, G.E.P., J.S. Hunter, and W.G. Hunter, *Statistics for experimenters*. 1978: Wiley New York.
104. Espie, D. and S. Macchietto, *The optimal design of dynamic experiments.* AIChE Journal, 1989. **35**(2): p. 223-229.
105. Hunter, W.G. and A.M. Reiner, *Designs for discriminating between two rival models.* Technometrics, 1965. **7**(3): p. 307-323.
106. Buzzi-Ferraris, G. and F. Manenti, *Kinetic models analysis.* Chemical Engineering Science, 2009. **64**(5): p. 1061-1074.
107. Linhart, H. and W. Zucchini, *Model selection*. 1986: John Wiley & Sons.
108. Zerry, R.u.W., G., *MOSAIC: ein Modellierungsserver für die Verfahrenstechnik*. 2005.
109. Logsdon, J.S. and L.T. Biegler, *Accurate solution of differential-algebraic optimization problems.* Industrial & engineering chemistry research, 1989. **28**(11): p. 1628-1639.
110. Zavala, V.M., C.D. Laird, and L.T. Biegler, *A fast moving horizon estimation algorithm based on nonlinear programming sensitivity.* Journal of Process Control, 2008. **18**(9): p. 876-884.
111. Vaca, M., A. Jimenez-Gutierrez, and R. Monroy-Loperena, *Application of collocation techniques to the design of petlyuk distillation columns.* Revista Mexicana de Ingeniería Química, 2006. **5**(1): p. 121-127.
112. Holmstrom, K., *The TOMLAB Optimization Environment in Matlab 1*. 1999.
113. Binder, T. and E. Kostina, *Approximate Gauss-Newton Method for Robust Parameter Estimation.* PAMM, 2009. **10**(1): p. 531-532.
114. Thanedar, P.B., J.S. Arora, C.H. Tseng, O.K. Lim, and G.J. Park, *Performance of some SQP algorithms on structural design problems.* 1986.
115. Drews, A., H. Arellano-Garcia, J. Schöneberger, J. Schaller, G. Wozny, and M. Kraume, *Model-based recognition of fouling mechanisms in membrane bioreactors.* Desalination, 2009. **236**(1-3): p. 224-233.
116. Kraume, M., U. Bracklow, M. Vocks, and A. Drews, *Nutrients removal in MBRs for municipal wastewater treatment*. 2005. p. 391-402.
117. Gimbel, R., A. Nahrstedt, and S. Panglisch, *Filtration zur Partikelabtrennung bei der Wasserreinigung.* Chemie Ingenieur Technik, 2008. **80**(1 2): p. 35-47.
118. Drews, A., H. Arellano-Garcia, J. Schoneberger, J. Schaller, M. Kraume, and G. Wozny, *Improving the Efficiency of Membrane Bioreactors by a Novel Model-based Control of Membrane Filtration.* Computer Aided Chemical Engineering, 2007. **24**: p. 345.

119. Drews, A. and M. Kraume, *Process improvement by application of membrane bioreactors.* Chemical Engineering Research and Design, 2005. **83**(3): p. 276-284.
120. Goshen-Meskin, D. and I.Y. Bar-Itzhack, *Observability analysis of piece-wise constant systems. I. Theory.* Aerospace and Electronic Systems, IEEE Transactions on, 1992. **28**(4): p. 1056-1067.
121. Bemporad, A., G. Ferrari-Trecate, and M. Morari, *Observability and controllability of piecewise affine and hybrid systems.* Automatic Control, IEEE Transactions on, 2000. **45**(10): p. 1864-1876.
122. Isermann, R., *Supervision, fault-detection and diagnosis methodsâ€"a short introduction.* Fault-Diagnosis Applications, 2011: p. 11-45.
123. Frank, P.M. and X. Ding, *Survey of robust residual generation and evaluation methods in observer-based fault detection systems.* Journal of process control, 1997. **7**(6): p. 403-424.
124. Sturza, M.A. and A.K. Brown. *Comparison of fixed and variable threshold RAIM algorithms*. Navigation, Vol. 35, No. 4, Winter 1988-89.
125. Gray, N.F., *Activated sludge: theory and practice*. 1990: Oxford University Press.
126. Gernaey, K.V., M.C.M. van 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, 2004. **19**(9): p. 763-783.
127. Petersen, B., K. Gernaey, M. Henze, and P.A. Vanrolleghem, *Evaluation of an ASM 1 model calibration procedure on a municipal-industrial wastewater treatment plant.* Journal of Hydroinformatics, 2002. **4**(1): p. 15-38.
128. Gujer, W., M. Henze, T. Mino, T. Matsuo, M. Wentzel, and G. Marais, *The activated sludge model no. 2: biological phosphorus removal.* Water science and technology, 1995. **31**(2): p. 1-11.
129. Gujer, W., M. Henze, T. Mino, and M. Van Loosdrecht, *Activated sludge model No. 3.* Water Science & Technology, 1999. **39**(1): p. 183-193.
130. Sin, G., D. Kaelin, M.J. Kampschreur, I. Takacs, B. Wett, K.V. Gernaey, L. Rieger, H. Siegrist, and M.C. van Loosdrecht, *Modelling nitrite in wastewater treatment systems: a discussion of different modelling concepts.* Water science and technology: a journal of the International Association on Water Pollution Research, 2008. **58**(6): p. 1155.
131. Velmurugan, S., W.W. Clarkson, and J.N. Veenstra, *Model-Based Design of Sequencing Batch Reactor for Removal of Biodegradable Organics and Nitrogen.* Water Environment Research, 2010. **82**(5): p. 462-474.
132. Balku, S., M. Yuceer, and R. Berber, *Control vector parameterization approach in optimization of alternating aerobic-anoxic systems.* Optimal Control Applications and Methods, 2009. **30**(6): p. 573-584.
133. Artan, N. and D. Orhon, *Mechanism and design of sequencing batch reactors for nutrient removal.* 2005: Intl Water Assn.
134. Mace, S. and J. Mata-Alvarez, *Utilization of SBR technology for wastewater treatment: an overview.* Ind. Eng. Chem. Res, 2002. **41**(23): p. 5539-5553.
135. Office of Water Washington, D.C., *Wastewater Technology Fact Sheet Sequencing Batch Reactors.* EPA 832-F-99-073, 1999.
136. Cruz Bournazou, M.N., H. Arellano-Garcia, G. Wozny, G. Lyberatos, and C. Kravaris, *ASM3 extended for two-step nitrification-denitrification: a model reduction for Sequencing Batch Reactors.* Journal of Chemical Technology and Biotechnology 2012**,** DOI: 10.1002/jctb.3694.
137. Wilderer, P.A., R.L. Irvine, and M.C. Goronszy, *Sequencing batch reactor technology*. 2000: Intl Water Assn.
138. Luccarini, L., G.L. Bragadin, G. Colombini, M. Mancini, P. Mello, M. Montali, and D. Sottara, *Formal verification of wastewater treatment processes using events detected from continuous signals by means of artificial neural networks. Case study: SBR plant.* Environmental Modelling & Software. **25**(5): p. 648-660.
139. Ketchum, L.H., *Design and physical features of sequencing batch reactors.* Water Science and Technology, 1997. **35**(1): p. 11-18.

140. Bungay, S., M. Humphries, and T. Stephenson, *Operating strategies for variable flow sequencing batch reactors.* Water and Environment Journal, 2007. **21**(1): p. 1-8.
141. Upadhyayula, S., *Alkylation of Benzene with Isopropanol over SAPO-5: A Kinetic Study.* The Canadian Journal of Chemical Engineering, 2008. **86**: p. 207-213.
142. Mak, W.C., C. Chan, J. Barford, and R. Renneberg, *Biosensor for rapid phosphate monitoring in a sequencing batch reactor (SBR) system.* Biosensors and Bioelectronics, 2003. **19**(3): p. 233-237.
143. Barton, P.I. and C.K. Lee, *Modeling, simulation, sensitivity analysis, and optimization of hybrid systems.* ACM Transactions on Modeling and Computer Simulation (TOMACS), 2002. **12**(4): p. 256-289.
144. Akin, B.S. and A. Ugurlu, *Monitoring and control of biological nutrient removal in a sequencing batch reactor.* Process Biochemistry, 2005. **40**(8): p. 2873-2878.
145. Turk, O. and D.S. Mavinic, *Preliminary assessment of a shortcut in nitrogen removal from wastewater.* Canadian Journal of Civil Engineering/Revue Canadienne de Genie Civil, 1986. **13**(6): p. 600-605.
146. Katsogiannis, A.N., M.E. Kornaros, and G.K. Lyberatos, *Adaptive optimization of a nitrifying sequencing batch reactor.* Water Research, 1999. **33**(17): p. 3569-3576.
147. Bourgeois, W., J.E. Burgess, and R.M. Stuetz, *On-line monitoring of wastewater quality: a review.* Journal of chemical technology and biotechnology, 2001. **76**(4): p. 337-348.
148. Vanrolleghem, P.A. and D.S. Lee, *On-line monitoring equipment for wastewater treatment processes: state of the art.* Automation in Water Quality Monitoring, 2003. **47**(2): p. 1-34.
149. Kaelin, D., R. Manser, L. Rieger, J. Eugster, K. Rottermann, and H. Siegrist, *Extension of ASM3 for two-step nitrification and denitrification and its calibration and validation with batch tests and pilot scale data.* Water Research, 2009. **43**(6): p. 1680-1692.
150. Shampine, L.F., M.W. Reichelt, and J.A. Kierzenka, *Solving index-I DAEs in MATLAB and Simulink.* SIAM review, 1999. **41**(3): p. 538-552.
151. El-Mansi, E.M.T. and W.H. Holms, *Control of carbon flux to acetate excretion during growth of Escherichia coli in batch and continuous cultures.* Journal of general microbiology, 1989. **135**(11): p. 2875.
152. Diez-Gonzalez, F. and J.B. Russell, *The ability of Escherichia coli O157:H7 to decrease its intracellular pH and resist the toxicity of acetic acid.* Microbiology, 1997. **143**: p. 1175-1180.
153. Luli, G.W. and W.R. Strohl, *Comparison of growth, acetate production, and acetate inhibition of Escherichia coli strains in batch and fed-batch fermentations.* Applied and environmental microbiology, 1990. **56**(4): p. 1004.
154. Wolfe, A., *The Acetate Switch [Review].* Microbiol Mol Biol Rev, 2005. **69**(1): p. 12-50.
155. Kayser, A., J. Weber, V. Hecht, and U. Rinas, *Metabolic flux analysis of Escherichia coli in glucose-limited continuous culture. I. Growth-rate-dependent metabolic efficiency at steady state.* Microbiology, 2005. **151**(3): p. 693.
156. Abdel-Hamid, A.M., M.M. Attwood, and J.R. Guest, *Pyruvate oxidase contributes to the aerobic growth efficiency of Escherichia coli.* Microbiology, 2001. **147**(6): p. 1483.
157. Phue, J.N. and J. Shiloach, *Impact of dissolved oxygen concentration on acetate accumulation and physiology of E. coli BL21, evaluating transcription levels of key genes at different dissolved oxygen conditions.* Metab Eng, 2005. **7**(5-6): p. 353-63.
158. Kumari, S., C.M. Beatty, D.F. Browning, S.J.W. Busby, E.J. Simel, G. Hovel-Miner, and A.J. Wolfe, *Regulation of acetyl coenzyme A synthetase in Escherichia coli.* Journal of bacteriology, 2000. **182**(15): p. 4173.
159. Nahku, R., K. Valgepea, P.J. Lahtvee, S. Erm, K. Abner, K. Adamberg, and R. Vilu, *Specific growth rate dependent transcriptome profiling of Escherichia coli K12 MG1655 in accelerostat cultures.* Journal of biotechnology, 2010. **145**(1): p. 60-65.
160. Akesson, M., E.N. Karlsson, P. Hagander, J.P. Axelsson, and A. Tocaj, *On-Line Detection of Acetate Formation in Escherichia coli Cultures Using Dissolved Oxygen Responses to Feed Transients.* Biotechnology and Bioengineering, 1999. **64**(5): p. 590-598.
161. Zawada, J. and J. Swartz, *Maintaining Rapid Growth in Moderate-Density Escherichia coli Fermentations.* Biotech Bioeng, 2005. **89**(4): p. 407-415.

162. Tosi, S., M. Rossi, E. Tamburini, G. Vaccari, A. Amaretti, and D. Matteuzzi, *Assessment of in line near infrared spectroscopy for continuous monitoring of fermentation processes.* Biotechnology progress, 2003. **19**(6): p. 1816-1821.
163. Kovatchev, B.P., L.A. Gonder-Frederick, D.J. Cox, and W.L. Clarke, *Evaluating the accuracy of continuous glucose-monitoring sensors.* Diabetes Care, 2004. **27**(8): p. 1922.
164. Whiffin, V.S.C., M. J. Cord-Ruwisch, R., *Online detection of feed demand in high cell density cultures of Escherichia coli by measurement of changes in dissolved oxygen transients in complex media* Biotechnology and Bioengineering, 2004. **85**(4): p. 422-433
165. Turner, C.G., M. E. Thornhill, N. F., *Closed-loop control of fed-batch cultures of recombinant Escherichia coli using on-line HPLC.* Biotechnology and Bioengineering 1994. **44**(7): p. 819-829
166. Chorvat Jr, D. and A. Chorvatova, *Multi wavelength fluorescence lifetime spectroscopy: a new approach to the study of endogenous fluorescence in living cells and tissues.* Laser Physics Letters, 2009. **6**(3): p. 175-193.
167. Marose, S., C. Lindemann, and T. Scheper, *Two Dimensional Fluorescence Spectroscopy: A New Tool for On Line Bioprocess Monitoring.* Biotechnology progress, 1998. **14**(1): p. 63-74.
168. Gonzalez-Vara, Y.R., *Enhanced production of L-(+)-lactic acid in chemostat by Lactobacillus casei DSM 20011 using ion-exchange resins and cross-flow filtration in a fully automated pilot plant controlled via NIR.* Biotechnology and bioengineering, 2000. **67**(2): p. 147-156.
169. Scarff, M., S.A. Arnold, L.M. Harvey, and B. McNeil, *Near infrared spectroscopy for bioprocess monitoring and control: current status and future trends.* Critical reviews in biotechnology, 2006. **26**(1): p. 17-39.
170. Hewitt, C.J. and G. Nebe-Von-Caron, *The application of multi-parameter flow cytometry to monitor individual microbial cell physiological state.* Physiological Stress Responses in Bioprocesses, 2004: p. 197-223.
171. Shapiro, H.M., *Microbial analysis at the single-cell level: tasks and techniques.* Journal of microbiological methods, 2000. **42**(1): p. 3-16.
172. Junne, S., M.N. Cruz Bournazou, A. Angersbach, and P. Götz, *Electrooptical monitoring of cell polarizability and cell size in aerobic Escherichia coli batch cultivations.* Journal of Industrial Microbiology and Biotechnology, 2010: p. 1-8.
173. Want, A., O.R.T. Thomas, B. Kara, J. Liddell, and C.J. Hewitt, *Studies related to antibody fragment (Fab) production in Escherichia coli W3110 fed batch fermentation processes using multiparameter flow cytometry.* Cytometry Part A, 2009. **75**(2): p. 148-154.
174. Merten, O.W., *Recombinant protein production with prokaryotic and eukaryotic cells: a comparative view on host physiology: selected articles from the meeting of the EFB Section on Microbial Physiology, Semmering, Austria, 5th-8th October 2000*. 2001: Springer Netherlands.
175. Chassagnole, C., N. Noisommit-Rizzi, J.W. Schmid, K. Mauch, and M. Reuss, *Dynamic modeling of the central carbon metabolism of Escherichia coli.* Biotechnology and Bioengineering, 2002. **79**(1): p. 53-73.
176. Wang, F.S., *A modified collocation method for solving differential-algebraic equations.* Applied mathematics and computation, 2000. **116**(3): p. 257-278.
177. Pesavento, C. and R. Hengge, *Bacterial nucleotide-based second messengers.* Curr Opin Microbiol, 2009. **12**(2): p. 170-6.
178. Andersson, L., S.J. Yang, P. Neubauer, and S.O. Enfors, *Impact of plasmid presence and induction on cellular responses in fed batch cultures of Escherichia coli.* Journal of Biotechnology, 1996. **46**(3): p. 255-263.
179. Magnusson, L.U., A. Farewell, and T. Nystrom, *ppGpp: a global regulator in Escherichia coli.* Trends Microbiol, 2005. **13**(5): p. 236-42.
180. Lin, H.Y. and P. Neubauer, *Influence of controlled glucose oscillations on a fed-batch process of recombinant Escherichia coli.* J Biotechnol, 2000. **79**(1): p. 27-37.
181. Hardiman, T., K. Lemuth, M.A. Keller, M. Reuss, and M. Siemann-Herzberg, *Topology of the global regulatory network of carbon limitation in Escherichia coli.* Journal of biotechnology, 2007. **132**(4): p. 359-374.

182. Usuda, Y., Y. Nishio, S. Iwatani, S.J. Van Dien, A. Imaizumi, K. Shimbo, N. Kageyama, D. Iwahata, H. Miyano, and K. Matsui, *Dynamic modeling of Escherichia coli metabolic and regulatory systems for amino-acid production.* J Biotechnol, 2010. **147**(1): p. 17-30.
183. Costa, R.S., D. Machado, I. Rocha, and E.C. Ferreira, *Hybrid dynamic modeling of Escherichia coli central metabolic network combining Michaelis-Menten and approximate kinetic equations.* Biosystems, 2010. **100**(2): p. 150-7.
184. Srinivasan, B., D. Bonvin, E. Visser, and S. Palanki, *Dynamic optimization of batch processes:: II. Role of measurements in handling uncertainty.* Computers & chemical engineering, 2003. **27**(1): p. 27-44.
185. Joshi, A. and B.O. Palsson, *Escherichia-Coli Growth Dynamics - a 3-Pool Biochemically Based Description.* Biotechnology and Bioengineering, 1988. **31**(2): p. 102-116.
186. Lahtvee, P.J., K. Adamberg, L. Arike, R. Nahku, K. Aller, and R. Vilu, *Multi-omics approach to study the growth efficiency and amino acid metabolism in Lactococcus lactis at various specific growth rates.* Microb Cell Fact, 2011. **10**(12): p. 4158-4168.
187. Kumari, S., R. Tishel, M. Eisenbach, and A.J. Wolfe, *Cloning, characterization, and functional expression of acs, the gene which encodes acetyl coenzyme A synthetase in Escherichia coli.* Journal of bacteriology, 1995. **177**(10): p. 2878.
188. Lin, H., N.M. Castro, G.N. Bennett, and K.Y. San, *Acetyl-CoA synthetase overexpression in Escherichia coli demonstrates more efficient acetate assimilation and lower acetate accumulation: a potential tool in metabolic engineering.* Applied microbiology and biotechnology, 2006. **71**(6): p. 870-874.
189. Cruz Bournazou, M.N., H. Arellano-Garcia, J.C. Schöneberger, S. Junne, P. Neubauer, and G. Wozny, *A Novel Approach to Mechanism Recognition in Escherichia Coli Fed-Batch Fermentations.* Computer Aided Chemical Engineering, 2009. **27**: p. 651-656.
190. Veloso, A.C.A., I. Rocha, and E.C. Ferreira, *Monitoring of fed-batch E. coli fermentations with software sensors.* Bioprocess and biosystems engineering, 2009. **32**(3): p. 381-388.
191. Shimizu, H., R.A. Field, S.W. Homans, and A. Donohue-Rolfe, *Solution structure of the complex between the B-subunit homopentamer of verotoxin VT-1 from Escherichia coli and the trisaccharide moiety of globotriaosylceramide.* Biochemistry, 1998. **37**(31): p. 11078-11082.
192. Jensen, E.B. and S. Carlsen, *Production of recombinant human growth hormone in Escherichia coli: expression of different precursors and physiological effects of glucose, acetate, and salts.* Biotechnology and bioengineering, 1990. **36**(1): p. 1-11.
193. MacDonald, H.L. and J.O. Neway, *Effects of medium quality on the expression of human interleukin-2 at high cell density in fermentor cultures of Escherichia coli K-12.* Applied and environmental microbiology, 1990. **56**(3): p. 640.
194. Holms, W.H., *The central metabolic pathways of Escherichia coli: relationship between flux and control at a branch point, efficiency of conversion to biomass, and excretion of acetate.* Current topics in cellular regulation, 1986. **28**: p. 69-105.
195. Majewski, R.A. and M.M. Domach, *Simple constrained-optimization view of acetate overflow in E. coli.* Biotechnology and bioengineering, 1990. **35**(7): p. 732-738.
196. Bauer, K.A., A. Ben-Bassat, M. Dawson, V.T. De La Puente, and J.O. Neway, *Improved expression of human interleukin-2 in high-cell-density fermentor cultures of Escherichia coli K-12 by a phosphotransacetylase mutant.* Applied and environmental microbiology, 1990. **56**(5): p. 1296.
197. Chou, C.H., G.N. Bennett, and K.Y. San, *Effect of modified glucose uptake using genetic engineering techniques on high-level recombinant protein production in escherichia coli dense cultures.* Biotechnology and bioengineering, 1994. **44**(8): p. 952-960.
198. Chou, C.H., G.N. Bennett, and K.Y. San, *Effect of modulated glucose uptake on high-level recombinant protein production in a dense Escherichia coli culture.* Biotechnology progress, 1994. **10**(6): p. 644-647.
199. Jung, H., R. Räbenhagen, S. Tebbe, K. Leifker, N. Tholema, M. Quick, and R. Schmid, *Topology of the Na+/Proline Transporter ofEscherichia coli.* Journal of Biological Chemistry, 1998. **273**(41): p. 26400.

200. Konstantinov, K., M. Kishimoto, T. Seki, and T. Yoshida, *A balanced DO-stat and its application to the control of acetic acid excretion by recombinant Escherichia coli.* Biotechnology and bioengineering, 1990. **36**(7): p. 750-758.
201. San Martin, M.C., N.P.J. Stamford, N. Dammerova, N.E. Dixon, and J.M. Carazo, *A structural model for the Escherichia coli DnaB helicase based on electron microscopy data.* Journal of structural biology, 1995. **114**(3): p. 167-176.
202. Lara, A.R., L. Caspeta, G. Gosset, F. Bolivar, and O.T. Ramirez, *Utility of an Escherichia coli strain engineered in the substrate uptake system for improved culture performance at high glucose and cell concentrations: An alternative to fedâ€ batch cultures.* Biotechnology and bioengineering, 2008. **99**(4): p. 893-901.
203. Lemoine, A., *Design of Experiments and parameter Estimation of a kinetic model for E. coli cultivations.* Studienarbeit, 2011.
204. Angersbach, A. and V. Bunin, *Continuous On-line Electrooptical Monitoring of Physiological State of Bacterial Cultures in Biotechnological Processes.*
205. DIN_EN_ISO_11213, *modifizierte Stärke-Betimmung des Acetylgehaltes-Enzymatisches Verfahren.* 1995.
206. Nisselbaum, J.S. and S. Green, *A simple ultramicro method for determination of pyridine nucleotides in tissues 1.* Analytical biochemistry, 1969. **27**(2): p. 212-217.
207. Bernofsky, C. and M. Swan, *An improved cycling assay for nicotinamide adenine dinucleotide.* Analytical biochemistry, 1973. **53**(2): p. 452-458.
208. San, K.Y., G.N. Bennett, S.J. Berrios-Rivera, R.V. Vadali, Y.T. Yang, E. Horton, F.B. Rudolph, B. Sariyar, and K. Blackwood, *Metabolic engineering through cofactor manipulation and its effects on metabolic flux redistribution in Escherichia coli.* Metabolic engineering, 2002. **4**(2): p. 182-192.
209. Parsopoulos, K.E. and M.N. Vrahatis, *Recent approaches to global optimization problems through particle swarm optimization.* Natural computing, 2002. **1**(2): p. 235-306.
210. Usuda, Y., Y. Nishio, S. Iwatani, S.J. Van Dien, A. Imaizumi, K. Shimbo, N. Kageyama, D. Iwahata, H. Miyano, and K. Matsui, *Dynamic modeling of Escherichia coli metabolic and regulatory systems for amino-acid production.* Journal of Biotechnology, 2010. **147**(1): p. 17-30.
211. Singh, V.K. and I. Ghosh, *Kinetic modeling of tricarboxylic acid cycle and glyoxylate bypass in Mycobacterium tuberculosis, and its application to assessment of drug targets.* Theor Biol Med Model, 2006. **3**: p. 27.