Hybrid Modelling for Enhanced Bioreactor Performance

Thaysen, Mads; Jørgensen, Sten Bay

Publication date:
2006

Document Version
Publisher's PDF, also known as Version of record

Hybrid Modeling for Enhanced Bioreactor Performance

Mads Thaysen

2005

Computer Aided Process Engineering Center
Department of Chemical Engineering
Technical University of Denmark

ISBN 87-9143-40-4
Printed by Nørhaven Book A/S, Copenhagen, Denmark

Preface

This thesis was prepared at the Computer Aided Process Engineering Center (CAPEC) at the Department of Chemical Engineering, Technical University of Denmark (DTU) in collaboration with BioProcess Laboratories, Novo Nordisk A/S, in partial fulfillment of the requirements for receiving the Ph.D. degree and the Industrial Ph.D. degree. The work was jointly financing by the Danish Academy of Technical Sciences and Novo Nordisk A/S and was carried out as a Danish Industrial Ph.D. study programme from May 2001 to December 2004.

A large number of people have contributed to the work presented in this thesis, providing either scientific, technical and/or social support, help and advice. Only a too short list of names is mentioned here; for those names missing ... your contribution has been highly appreciated.

I would like to thank my three supervisors: Professor Sten Bay Jørgensen for his guidance and inspiration, presenting me to academic research and for giving me the opportunity to carry out this Industrial Ph.D. study. Scientist Jørgen Petersen for introducing me to industrial research and teaching me the ropes for maneuvering in the corporate jungle. Scientist Erik Steffensen for providing "the floodlights of production" to my conclusions and adding the spice of reality to the abstract world of modeling. Only through your joint effort has this thesis been possible.

For the financial support I thank the Danish Academy of Technical Sciences and Novo Nordisk A/S.

At DTU the research groups at CAPEC and Biocentrum provided support, advice and social support to cultivate me from a young, green research student to a mature researcher. Amongst others, I thank my old colleagues at CAPEC: Morten Skov Hansen, Dennis Bonné, Krist Gernaey, Niels Rode Kristensen, Jan Kamyno Rasmussen, Rasmus Leth-Miller, Irene Papaeconomue, John Bagterp Jørgensen, Lars Gregersen and Peter Harper. At Biocentrum the help of Lisbeth Olsson and Anna Eliason along with their laboratory technicians was much appreciated. My two short visits to CPACT at University of Newcastle upon Tyne, provided me with an alternative view to academic research. My thanks goes out to Julian Morris, Elaine Martin, Angela Bott, Ewan Mercer and the rest of the group for a memorable time.

At Novo Nordisk I wish to acknowledge the help and support from the people at the old BioProcess departments: 201.01-03. The friendship, willingness to help and family-like atmosphere in the laboratory group will long be remembered. Saqib Akhtar, thanks for colonizing 8A with me and providing help and high spirits during long lab hours. The professionalism and sincere interest of Stine L. Hansen and Jens G. Poulsen were the cornerstones of much of the work presented here. Also the wizardry of Jakob Helsø in taming Delta-V was without comparison. The discussions, opinions, ideas and visions on Process Analytical Technologies of Erik Skibsted, Jesper Wagner, Søren Antoft, Martijn Wiertz and Casper Leuenhagen were a couple of years ahead of schedule but fruitful for this thesis, thanks.

Sincere thanks to my friends, family and parents-in-law that listened with patience

to my ups and downs. My uncle Claus provided great support and scientific comments to my work, thanks. A big thanks to my parents, Hanne and Kjeld, and my brother, Jacob, for their love and support and believing in me.

Finally, I wish to thank my wife Karen for her love and understanding. I believe that this thesis would not have been possible without your kind support, providing an environment for me to extend my horizon and for this I am deeply grateful. And to my son Marko, your arrival opened my eyes to what is truly precious.

Kalundborg, September 2005.
Mads Thaysen

Contents

Appendices

Summary

The topic of this thesis is the modeling of industrial process data from bioreactors for analysis, monitoring, control and optimization of cultivations for production of insulin precursors using genetically modified strains of *Saccharomyces cerevisiae*. The aim of this work is to develop models in order to facilitate and support the analysis of production data from industrial cultivation processes, which can be applied in both offline and online process analysis.

Soft sensors are first developed for obtaining quantitative information from an existing industrial cultivation process. Two similar models are developed for monitoring of the biomass and product concentrations using first principle engineering modeling (FPEM). Application of the two simple soft sensors on industrial data provide reasonable descriptions of the general biomass and product concentration trajectories. Implementation and use of the soft sensors will enable a very simple yet highly attractive way of providing online information of the two key process variables. Process Chemometrics provide an alternative approach for monitoring the product concentration. A multiway projection to latent structures (MPLS) regression model is presented, providing both one-step ahead and end point predictions of the product concentration within 5-10 % of analytical offline measurements. Comparison of the two soft sensors for describing the product concentration indicates that the MPLS-predictor for the one-step ahead prediction gives a slightly better description of the variations in the product concentration.

Attempts to use the two FPEM models mentioned above on the cultivation of a similar recombinant strain of *S. cerevisiae* failed, since unanticipated production of acetate indicates a different metabolic response to the growth conditions. Literature reviews on the transport and effect of organic acids in cultivation processes as well as the genetic engineering performed on the strain are presented. For modeling using mass balances the elemental composition of the biomass is required and since it has not been reported if the elemental composition of the recombinant strain has been influenced by the genetic engineering, this is investigated using macroscopic mass balances. The elemental composition of the biomass is found to be similar to what is reported in the literature.

An indicator of the onset of oxidoreductive growth is presented. The indicator is based on comparing the measured ammonia flow rate to a modelbased estimate of the ammonia flow rate during oxidative growth. The model of the ammonia flow rate takes into account the effects from changes in the volume of the culture broth, effluent flow rate and glucose syrup feed rate. The indicator not only facilitates the monitoring of the cultivation process for onset of oxidoreductive growth. Together with the modelbased residual between measured and estimated ammonia flow rate, the indicator provides the foundation for extended modeling of the cultivation process.

Using the information provided by the indicator of the onset of oxidoreductive growth, a soft sensor is constructed for the estimation of the conversion rate of

acetate as well as the production rate of biomass. The soft sensor is based on a combination of a proton balance and a carbon balance, and using these rates in dynamic mass balances online estimations of acetate and biomass concentrations are provided as well. Applying the model on data from a number of cultivations, provides a surprising observation namely that acetate is being produced in large amounts 1-2 hours before formation of ethanol occurs. The reason for the onset of the acetate formation has not yet been determined.

A small metabolic flux model is proposed using calculated and estimated conversion rates of substrate, biomass and key metabolites combined with physiological parameters reported in the open literature on another strain of *S. cerevisiae*. The model is used to illustrate and discuss observations from cultivations showing both normal and abnormal process behavior. The model illustrates how acetate is produced prior to ethanol formation. The model also shows how the activity of the oxidative phosphorylation changes extensively as ethanol formation starts and as ethanol consumption ends, which is interpreted as effects from repression/derepression of the oxidative phosphorylation. It has not been possible to explain what mechanism is responsible for this control of the oxidative phosphorylation, although it is discussed that it could not be a fixed limitation in the capacity of the oxidative phosphorylation, since an experiment using closed loop control of the ethanol concentration in the offgas shows even higher activity of the oxidative phosphorylation than are seen in similar open loop experiments.

Finally a simple model is proposed to describe the specific productivity of the product. The description is based on a first order model expression for the dependence of production rate and biomass synthesis rate, with a time constant proportional to the specific glucose uptake rate provided by the metabolic flux model presented above. The model gives a reasonable description of the observed trajectories of product concentration in a normal cultivation and by a small extension of the model, it is also able to provide a reasonable estimation of the product concentration profile during process upset in the form of acetate formation.

Resume

Emnet for denne afhandling er modellering af industrielle procesdata fra bioreaktorer med henblik på analyse, monitering, regulering og optimering af gæringer til produktion af insulin precursors ved brug af genmodiferede stammer af *Saccharomyces cerevisiae*. Formålet med dette arbejde er at udvikle modeller til at lette og understøtte analyse af produktionsdata fra industrielle gæringsprocesser og som kan anvendes ved både offline og online analyse af processen.

Soft sensorer er først udviklet for at opnå kvantitativ information fra en eksisterende industriel gæringsproces. To sammenlignelige modeller udvikles til monitering af biomasse- og produktkoncentrationen ved brug af *first principle engineering modeling* (FPEM). Anvendelsen af de to simple soft sensorer på industrielle data resulterer i en fornuftig beskrivelse af de generelle biomasse- og produktkoncentrationsprofiler. Implementering og brug af soft sensorerne vil muliggøre en meget simple men meget attraktiv måde at skaffe online information af de to nøglevariable for processen på. Proceskemometri bidrager med en alternative tilgang til monitering af produktkoncentrationen. En *multiway projection to latent structures regression model* (MPLS) præsenteres og denne bibringer 1-skridts og slutpunkts prædiktion af produktkoncencentrationen som er inden for 5-10 % af offline analyserne. Sammenligningen af de to soft sensorer som beskriver produktkoncentration indikerer, at MPLS-prædiktoren for 1-skridts prædiktion giver en smule bedre beskrivelse af variationerne i produktkoncentrationen.

Forsøg på at anvende de to FPEM modeller nævnt ovenfor på gæringer af en tilsvarende rekombinant stamme af *S. cerevisiae* slog fejl, da uventet produktion af acetat antydede et anderledes metabolisk respons på vækstbetingelserne. Litteraturstudier af transport og indflydelse af organiske syrer i gæringsprocesser præsenteres, såvel som de genetiske modifikationer udført på stammen. Til modellering der anvender massebalancer skal elementarsammensætningen af biomassen gerne kendes. Da der ikke er rapporteret, hvorvidt stammens elementarsammensætning er påvirket af de genetiske modifikationer af stammen, undersøges dette ved brug af makroskopiske massebalancer. Undersøgelsen viser, at stammens elementarsammensætning er sammenlignelig med det som rapporteres i litteraturen.

En indikator for begyndelse af oxidoreduktiv vækst bliver præsenteret. Indikatoren er baseret på en sammenligning af den målte flowhastighed af ammoniak med et modelbaseret estimat af samme flowhastighed under oxidativ vækst. Modellen af flowhastigheden af ammoniak tager hensyn til effekter fra ændringer i volumen af kulturvæsken, væskeflow fra bioreaktoren og doseringshastigheden af glucose sirup. Indikatoren letter ikke kun monitering af gæringer for begyndelse af oxidoreduktiv vækst. Sammen med det modelbaserede residual mellem den målte og estimerede flowhastighed af ammoniak, bidrager indikatoren med et grundlag for yderligere modellering af gæringsprocesser.

Med information fra indikatoren for begyndelse på oxidoreduktiv vækst, konstrueres en soft sensor til estimering af omdannelseshastigheden af acetat og produk-

tionshastigheden af biomasse. Soft sensoren er baseret på en kombination af en protonbalance og en kulstofbalance, og ved at benytte disse i dynamiske massebalancer kan man ydermere opnå online estimering af acetat- og biomassekoncentrationer. Anvendelse af modellen på data fra et antal gæringer resulterer i den overraskende observation at acetat dannes i store mængder 1-2 timer før ethanoldannelsen starter. Forklaring for mekanismen bag denne observation er endnu ikke bestemt.

En lille metabolisk fluksmodel er præsenteret, som anvender beregnede og estimerede omdannelseshastigheder af substrat, biomasse og nøglemetabolitter i kombination med fysiologiske parametre rapporteret i den åbne litteratur om andre stammer af *S. cerevisiae*. Modellen benyttes til at illustrere og diskutere observationer fra gæringer der viser både normal og unormal opførsel. Modellen illustrerer, hvorledes acetat produceres før ethanoldannelse. Modellen viser ydermere, at aktiviteten af den oxidative fosforylering ændres meget, når ethanoldannelse påbegyndes og afsluttes, hvilket fortolkes som en effekt af repression/derepression af den oxidative fosforylering. Det har ikke været muligt at forklare, hvilken mekanisme der er ansvarlig for denne regulering af den oxidative fosforylering. Det diskuteres, at det ikke kan være på grund af en fast begrænsning i kapaciteten af den oxidative fosforylering, da et eksperiment, som anvender lukketsløjfe regulering af ethanolkoncentrationen i afgangsgassen, viser endnu højere aktivitet af den oxidative fosforylering end hvad der ses i eksperimenter med åbensløjfe regulering.

Til slut præsenteres en simpel model til beskrivelse af den specifikke produktivitet af insulin precursoren. Beskrivelsen er baseret på et første ordens modelleringsudtryk for afhængigheden af produktionshastigheden og biomassesyntesehastigheden med en tidskonstant som er proportional med den specifikke glucoseoptagelseshastighed, som leveres af den metaboliske fluksmodel nævnt ovenfor. Modellen giver en tilfresstillende beskrivelse af de observerede produktkoncentrationsprofiler for normale gæringer. En lille udvidelse af modellen gør, at denne også giver en tilfresstillende beskrivelse af produktkoncentrationsprofilen ved procesforstyrrelse såsom acetatdannelse.

1

General Introduction

1.1 Background

The construction of, introduction of and use of genetically modified organisms (GMO's) in biotechnological and pharmaceutical production processes have extended the potential of these industries to a stage, where nothing seems impossible or unachievable. The fund of knowledge within Microbiology and Biochemistry is nowadays so extensive that microorganisms can literarily be tailor-made to serve a desired purpose *e.g.* produce antibiotics or remove hevay metals from industrial waste streams.

There is still a long way from successful genetically modifying a microorganism in the laboratory to obtaining optimal productivity and effectiveness in large scale industrial cultivations of that microorganism. The path of process development between mL to m^3 scale requires development of the cultivation method and medium composition, screening of different suitable strains for highest productivity during anticipated production conditions, scale-up from lab to pilot plant and further to production scale with investigations and tests at each level, and finally optimization of the process and process economics under industrial production conditions.

A pivotal aspect of the investigations, decision making and scale up of any production process is the ability of monitoring the process state *i.e.* to follow the most important process parameters and variables during the process. In cultivation processes monitoring is often restricted to the measurement of macroscopic entities such as flows, pressures, temperatures and volumes. Sometimes the off-gas from the bioreactor is analysed online for its primary constituents, but in general the monitoring of the biological entities in the cultivation broth is limited to a couple of easily measured species and is carried out as off-line analyses often with a significant time delay between sampling and availability of the analysis result.

The purchase and introdution of more advanced process analytical instruments could be used to provide information on the process by measuring more of the biological entities and species in the culture broth. However the information the instruments provide must be compared to their influcence on the process economics and educational requirements of the operating personel. The pharmaceutical industry has tradionally been very reluctant to introduce new process analytical technologies, since it was not clear how regulating authorities would challenge the use of such technologies. With new guidelines from the powerful Federal Drug Administration in USA (Guidance for Industry on Process Analytical Technologies (PAT) USFDA (2004b)), the regulating authorities are now endorsing the integrated use of measurements, experience and knowledge in order to improve product and process quality.

Most of the primary focus of the PAT discussions have concerned measurement technologies and their use and application. However an important aspect of the PAT guidelines concerns the use of information by integrating data from different sources. In this thesis this aspect is approached by combining knowledge and experience of cultivating microorganisms with engineering principles such as mass and elemental balances. This combination of different expertise is able to provide additional information on the dynamics of the cultivation process, information that could potentially be made available for online monitoring of the process. Even in offline analysis of cultivation data it is desirable to ensure enhanced use of information in order to analyse the process dynamics and improve the understanding of the process.

1.1.1 Novo Nordisk A/S

For almost 2 decades Novo Nordisk A/S have used genetically modified organisms in the industrial production of insulin and insulin analogues and the company is one of the biggest producers with a market share around 45 % by volume. Insulin is used for treatment of diabetes, a metabolic disorder in which glucose is not consumed normally by cells, due to low concentrations of insulin. It has been estimated that the number of diabetic patients will grow to more than 200 millions by 2010, mainly due to the world wide development into Western life style with too little exercise combined with too much fatty food in the diet.

At Novo Nordisk A/S a continuous improvement of existing production processes as well as investigation and development of new processes for new products is carried out. The cultivation of the microorganism is only one of the many steps in the production line and each of the steps needs to be developed and improved in order to ensure robustness and high yields. The aim of the cultivation process is to obtain high productivity of the product a task that is intrinsicly related to the growth of the microorganism, due to the strategy used for expression of the recombinant product.

Although many biological variables are being measured during a typical cultivation at Novo Nordisk A/S most of these variables are measured offline at a frequency that does not allow for detailed analysis of the state of the metabolism, and definitely not online, since the time delay between sampling and analysis is more likely days than hours. The cultivations are mostly evaluated based on the rate of formation of the desired product and the biomass with only few attempts to provide a physiological explanation for variations in the cultivation data and offline measurements.

The present industrial Ph.D. project was established on the desire to use more of the information available in the cultivation process in order to monitor the state of the process and detect disturbances (small variations to the process recipe occurring frequently) and upsets (large and severe variations to the process recipe, less frequently). The project activities were to be carried out in the laboratory and pilot plant facilities of the company. Unfortunately major organizational rearrangements in the company made it impossible to setup a dedicated bioreactor for almost 2 years of the project. In this time period work was instead carried out on modeling of data from the production site.

1.2 Aim and Hypotheses

The aim of this work has been to develop models in order to facilitate and support the analysis of production data from industrial cultivation processes. The models should 1) support process engineers in documenting the course of cultivations *e.g.* as batch documentation required by regulatory authorities in the pharmaceutical industry, 2) support the evaluation of the process performance and analysis of process dynamics and 3) support the online analysis of the cultivation by monitoring the trajectories of essential cultivation states.

It was a further aim of the project to prove that more information and knowledge could be extracted from the existing process data. It was desired to construct models such that their application did not require purchasing and installation of new instruments or changes to the sampling and analysis of the culture broth.

The hypotheses of this thesis are therefore:

- The construction of simple models of a cultivation process will enhance the use of existing process instrumentation both in offline and online applications thereby facilitating improved modelbased monitoring of key process variables.
- The estimation and monitoring of key process variables will support the modeling and subsequent offline and online estimation of the production rate and concentration of the desired product, insulin precursors.

Such methods may be used for process analysis, monitoring, control and optimization.

1.3 Outline of Thesis

The thesis has been divided in two parts. The first part (chapters 2, 3, 4 and 5) provides background information on some topics that are relevant prerequisites for the chapters in the second part (chapters 6, 7, 8 and 9), describing the cultivation data and the construction of models for improved analysis and interpretatation of these data.

Chapter 2 provides an introduction to the generel concepts behind modeling. The aim of this chapter is to provide an idea of why the use of models is attractive. A number of modeling approaches is mentioned and it is argued that modeling is purpose-driven. The chapter ends with a section on the vision behind the use of models in a pharmaceutical company like Novo Nordisk A/S and a section on how soft sensors may be used to facilitate realization of such a vision for improved monitoring, control and optimization.

In **chapter 3** the most important metabolic pathways for aerobic growth on glucose as substrate are reviewed with focus on the catabolic reactions.

Surprisingly high concentrations of acetate were observed by offline analysis of culture broth during process disturbances. **Chapter 4** provides a review of the literature on the effects of weak organic acids on the growth dynamics of *Saccharomyces cerevisiae*.

In this thesis data from cultivations of two recombinant strains of *Saccharomyces cerevisiae* have been investigated. **Chapter 5** presents a detailed description of the

genetic modifications on the strain used in chapters 7, 8 and 9.

Chapter 6 presents the modeling of the production rate and concentration of an insulin precursor. Two different modeling approaches have been used, and the resulting models have been constructed with the aim of online application. First principle engineering modeling based on knowledge of the physiology and genetic modifications to the recombinant strain provides models for estimating the biomass and product concentration. The construction of a model using Multiway Partial Least Squares Regression is entirely driven by data. The identified model is limited to a fixed time period consisting of the fed-batch operation and some hours of continuous operation of a cultivation, referred to as the end point. The model is able to provide an online estimate of the product concentration as well as continuous prediction of the end point concentration.

In **chapter 7** the elemental composition of biomass of the strain presented in chapter 5 is investigated. Macroscopic mass balances are used to determine the elemental composition and the result is compared to the result obtained by elemental analysis of the biomass with respect to carbon, hydrogen and nitrogen. The two methods provide similar estimates of the elemental composition. The investigation also proves that the mass balances close properly under stationary conditions during continuous operation.

Cultivation data from experiments carried out in pilot plant showed that onset of oxidoreductive growth was accompanied by elevated ammonia flow rates in order to maintain a constant culture broth pH. An indictor of the onset of oxidoreductive growth was constructed by identifying periods of elevated ammonia flow rates and quantifying the extent and magnitude of the variation. A model of the ammonia demand during oxidative growth in fed-batch and continuous cultivations was constructed and the output of this model was compared to measurements of the ammonia flow rate. **Chapter 8** describes model building and verification. The chapter also illustrates the performance of the model during oxido-reductive growth, and it is described how the monitoring of the residual between measured and estimated ammonia flow rates can be used to determine the onset of oxido-reductive growth.

Chapter 9 presents the construction and performance of 3 simple mathematical models formulated as soft sensors. The first soft sensor provides estimates of biomass and acetate conversion rate and concentration and uses the model presented in chapter 8 to determine the onset of oxido-reductive growth. It is illustrated how significant acetate production occurs prior to ethanol formation at the onset of oxido-reductive growth. The second soft sensor provides estimates of the intrinsic metabolic flux distribution in a simple metabolic network. Use of the soft sensor is illustrated by investigating oscillations due to synchronized growth and analysis of the critical dilution rate by a slowly increasing ramp in the dilution rate. The third soft sensor provides estimates of the production rate and concentration of the product, an insulin precursor. All three soft sensors have been constructed for use both in off-line analysis of cultivation data and for online monitoring of a running cultivation in fed- batch and continuous operation.

Chapter 10 presents the overall conclusions for the work reported in this thesis along with suggested directions of future work.

1.4 List of Publications

During the industrial Ph.D. project the following conference contributions and articles have been produced:

- Mads Thaysen and Sten Bay Jørgensen; *On-line Estimation and Prediction of Biomass Concentration in Yeast Fermentations Using a Software Sensor*; Annual Meeting 2002 of AIChE, Indianapolis, USA. Oral presentation of conference article.

- Mads Thaysen and Sten Bay Jørgensen; *On-line Estimation and Prediction of Biomass Concentration in Yeast Fermentation using a Software Sensor*; ESBES-4 2002 in Delft, Holland. Poster presentation of conference abstract.

- Mads Thaysen and Sten Bay Jørgensen; *Process Software Sensor for Plant Optimization*; PSE 2003, Kunming, China. Poster presentation of conference article.

- Mads Thaysen and Sten Bay Jørgensen; *Application of Software Sensors for Monitoring and Prediction in Fermentation Processes*; AdChem Conference 2003, Hong Kong, China. Oral presentation of conference article.

- Mads Thaysen, Dennis Bonné and Sten Bay Jørgensen; *Modeling industrial fermentation data using Grid of Linear Models (GoLM)*; CAB-9 Conference 2004, Nancy, France. Poster presentation of conference article.

- Mads Thaysen, Dennis Bonné and Sten Bay Jørgensen; *Modeling industrial fermentation data using Grid of Linear Models (GoLM)*; BatchPro Final Meeting 2004, Poros, Greece. Poster presentation of conference article.

- Mads Thaysen and Sten Bay Jørgensen; *Formation of Acetate and Glycerol in High-density Cultivations of* Saccharomyces cerevisiae; Annual Meeting 2004 of AIChE, Houston, USA. Oral presentation of conference abstract.

- Mads Thaysen and Sten Bay Jørgensen; *Application of Software Sensors for Monitoring and Prediction in Fermentation Processes*; Submitted to Chemical Process Engineering Journal and correponds to chapter 6 in the present work.

Part I

Introductory Chapters

2

Modelbased Analysis, Monitoring and Control in Biotechnology and Biopharmaceutical Industries

Mathematical models are used in sciences to *capture* experience and knowledge in order to pass it on to others for many different purposes. One important application of models is testing or validation of hypotheses. By analyzing discrepancies between proposed models and the systems they represent, it is possible to propose model extensions, or a new model, and then validate the extended or new model using data of the system either obtained by planned experiments or from previous measurements of system behavior.

The purpose of this chapter is to give a brief and general introduction to modeling. Concepts of academic mathematical modeling of bioprocesses can often seem very different from the problems the professional is facing at a production plant when attempting to develop models of a process. Reasons for this apparent discrepancy could for example be that complex substrates are being used or that the constraints or dynamics of the utility systems providing aeration, steam, cooling water, stirring or dosing of feed streams affect the bioprocess. By introducing scientists and engineers to some of the concepts of modeling for application in complex industrial environments, the information, models and results presented in later chapters of this thesis can be understood, appreciated and serve as a starting point for modelbased analysis, monitoring and control in other areas of the bioprocess industry. Key points to be noted in the following are that 1) modeling is purpose driven, 2) information and control hierarchies affect not only a given plant but should rather be approached from a holistic point of view and 3) different kinds of model types exist depending both on how well the mechanisms of the process is understood and described and on the actual purpose of the modeling. The aim of this chapter is to argue that models are attractive tools for the understanding, monitoring, control and optimization of cultivation processes and production lines in general.

The first section argues that modeling is purpose-driven and categorizes the application of models depending on the purpose. A model of the control hierarchy of a production line is presented and used to illustrate how different models can be used at different levels in an organization to interpret the same process data for different purposes. Section 2.2 presents different types of mathematical model structures and shortly describes their potential applications. As experience and knowledge of a given process is extended it is possible to construct more complex and accurate process models. Section 2.3 provides a vision of how model based plant and/or business-wide monitoring, control and optimization can be envisoned in a

pharmaceutical company.

2.1 Purpose-driven Modeling

Before modeling of a given system is attempted it is important to clearly define what the purpose of the modeling is. This is needed in order to limit and focus the attention in the model construction and guiding the introduction of assumptions to obtain the best possible performance within the intended operating regime. When evaluating the model performance and assumptions during validation, it is likewise important that this is done within the intended operating regime and with the original purpose of the model in mind. When a model has been applied and validated for a given purpose in a given operating regime it can then serve as a starting point in order to develop an extended model for application in neighboring operating regimes. Here it is important to keep in mind that the original purpose and assumptions can be violated and therefore the model might need modifications to serve the altered purpose.

In the following, five different modeling categories for process performance optimization are presented. The concepts of information and control hierarchies are then presented and discussed in order to provide a framework for understanding how models shaped as observers or soft sensors can provide refined information at different levels in the two hierarchies and how modelbased controllers can be applied in the control hierarchy. In the present work only soft sensors were developed as will be shown in chapters 6 and 9.

2.1.1 Categories

The purpose of mathematical models in relation to process performance optimization can be divided into several categories:

- Analysis of steady state behavior
- Analysis of process dynamics
- Monitoring of key process variables
- Prediction of performance in possible future scenarios
- Control of the process

Models constructed for analysis of process data is often the first step on the path to understanding process dynamics and subsequently the construction of models for monitoring and control. The modeling cycle is repeated as models for monitoring and control are constructed, validated and implemented. Further improvement of process performance can be identified provided suitable methods are available This leads to construction of new models for analysis of the monitored and controlled process. This modeling cycle is illustrated in figure 2.1. In the different categories models are constructed for different specific purposes.

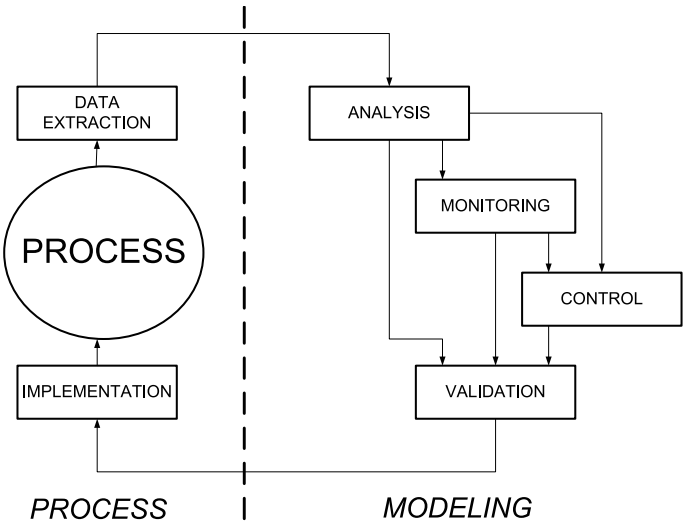


Figure 2.1. Schematic illustration of the modeling cycle. It is indicated how analysis of process data can lead to an understanding of the process dynamics. The acquired knowledge can then be used to change the operation of the process or be used in modelbased monitoring of the process performance. The information from the modelbased monitoring can also be used for modelbased control of the process.

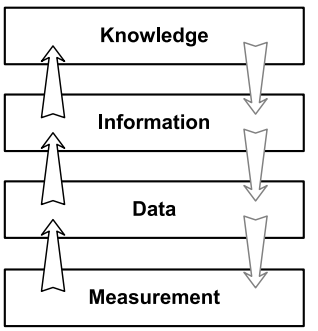


Figure 2.2. Illustration of the information hierarchy at a production plant. The arrows pointing upwards illustrate the flow of information. Process measurements are turned into data. Analysis of data produces information on the system, which can be condensed into knowledge in the form of justified true belief (Nonaka, 1994). The arrows pointing downwards illustrate how increased knowledge of the system can have an influence on the measurements, either by fine tuning of existing equipment, introduction of new measurements and control loops or construction of state observers/ soft sensors (see section 2.1.2.1).

2.1.2 Information Hierarchy

The key to supervision and control of production processes is to have data and information available to analyze and evaluate the process conditions. The transformation of raw process measurements to process knowledge is schematically illustrated through an information hierarchy as shown in figure 2.2. Here the raw measurements from the sensors in the process are translated to physical units *e.g.* pH, flows and temperatures, and then referred to as data. Analyzing the vast amounts of data from different sensors produces information on how the system reacts under differ-

ent circumstances. As *necessary* and *sufficient* information is gathered to capture and describe underlying principles responsible for the process behavior, an understanding of the behavior can be shaped and process knowledge can be developed. Increased knowledge and information on the system behavior can lead to new uses of process measurements or introduction of new derived variables or instruments in order to improve and optimize the monitoring and control of the productivity in a given process.

2.1.2.1 Observers

In biopharmaceutical cultivation processes it is often not possible to measure the desired variables in-line, on-line or at-line. Having information on the important product concentration profiles available after the production steps have been terminated is adequate for quality assurance and regulatory purposes (*e.g.* required by the authorities), however for process monitoring, control and optimization it is desired to have an on-line evaluation of the state of the process.

Using mathematical models describing the known or assumed mechanisms and dynamics of the production so-called state observers can help to infer information about unmeasured variables (outputs) by using available information from other measured variables (inputs). In this work the synonym soft sensor is used instead of state observer to highlight the similarities to the hardware sensors already present, installed and used in the process. Different frameworks can be used for formulating the mathematical models for the development of a soft sensor depending on the nature of the problem. The choice of which variables to relate can be based on analysis of knowledge, experience and insight into a given process, but can also be attempted through data driven approaches, where correlations between variables is evaluated. The nature of inputs and outputs for a soft sensor ranges from physical variables (e.g. flows and concentrations) to indicatory variables (e.g. operational/process phases and signal quality/noise level). It should, however, be emphasized that even though the soft sensor is mathematically formulated to measure a certain process state, the underlying assumptions for the model might fail, thus rendering the signal from the soft sensor erroneous. If possible, evaluation of assumption validity should be wrapped around a process soft sensor e.g. in an algorithm framework for monitoring application of the soft sensor.

In relation to the information hierarchy shown in figure 2.2, soft sensors are positioned between the data and information layers, providing additional information on unmeasured or infrequently measured variables. Introduction of process soft sensors constitutes an effective way to improve the level and quality of available process information. Essential in this discussion is the availability of specific information at many decision levels in the production organization. Production floor personnel can use the extended amount of information for improving the performance of a single unit, while the production management can use the information for improving planning and scheduling of the production line.

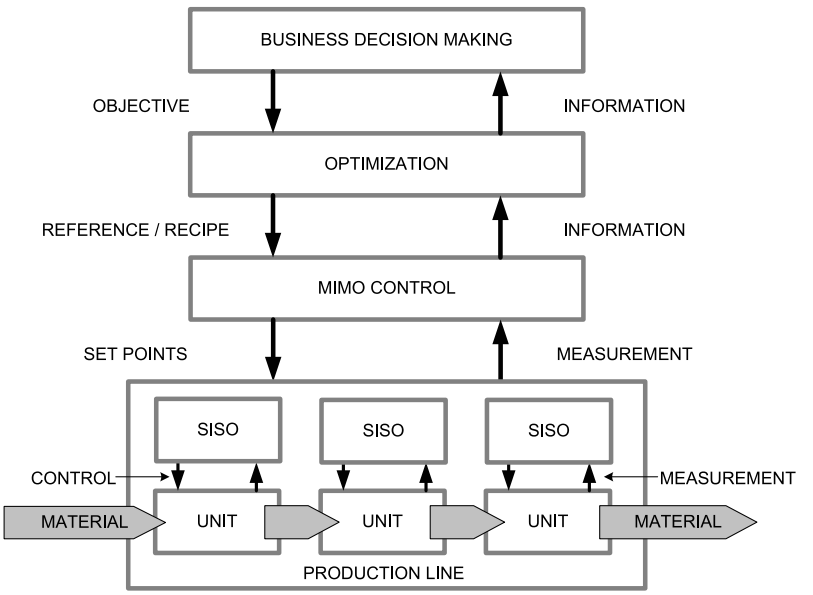


Figure 2.3. Overview of the control hierarchy in a process plant. The box at the bottom represents the plant floor where single input/single output (SISO) control loops support the process in each given process unit. This control action could both be situated in local programmable logic controllers (PLC's) or in a central process control system. Above the SISO control layer is a multiple input/multiple output (MIMO) layer, where more sophisticated multivariate process models are applied for the monitoring and control of the production line. Process optimization is carried out in the third layer, where the evaluation of different process scenarios can be analyzed and tested for optimal performance. At the top of the control hierarchy is the business decision making layer, where the production strategies are (re-)evaluated and decided. The type of information exchanged between layers is also illustrated. The characteristic time constants for the control action increase from the bottom (seconds) to the top layer (weeks or months), while the model specificity and detail decreases.

2.1.3 Control Hierarchy

It can be relevant to construct a number of similar process models from the same process data, although the use of the information from the process can be rather different. Figure 2.3 illustrates a schematic overview of the control hierarchy in a production plant, highlighting different control objectives at the different control layers *e.g.* optimization of productivity and maintaining process conditions with respect to infections. The bottom control layer is in close contact with the process units in the production line. Single input/single output (SISO) control ensures that the production plans are carried out according to the set points provided by the multiple input/multiple output (MIMO) layer. The MIMO layer contains the more sophisticated control algorithms using multivariable information and ensures that the information from the SISO layer is used to carry out the production plans as scheduled in the production recipe. As a response to possible discrepancies between desired and realized process trajectory, new set points are provided to the SISO controllers to guide the system between desired phases and within these. Whereas the control algorithms of the SISO controllers can be positioned both in decentralized

programmable logic controllers (PLC's) at the local process unit or in the overall process control system (PCS), the MIMO control algorithms are mostly situated in the central PCS. Above the MIMO layer in figure 2.3 an optimization layer is positioned, which serves the function of identifying and evaluating optimal process performance based on the constraints/objectives provided by the business decision making layer and information on current or historical process dynamics and performance. The function of the business decision making layer is to analyze the market for business opportunities and threats and combining this with knowledge of the capabilities of the production process to formulate a strategy in the form of optimization objectives that can guide the production process.

In all the layers of the control hierarchy models are used in order to analyze the correlation between inputs and outputs, monitor the variation of process variables and attempt to control the output variables by manipulating relevant input variables. The characteristic time constants for the control action increase from the bottom (seconds) to the top layer (weeks, months or years), while the model specificity and detail decreases in the same direction. With changing control objectives, the information requirement also changes as does the involved process measurements. Most often it is not the raw process online measurements *e.g.* stirring rates, pressures and temperatures that are used in the upper layers of the control hierarchy, but rather composite measurements such as energy consumption (power, heating, cooling), productivity, quality and yield (unit product per unit raw material). Some of the variables needed in the calculation of composite measurements might only be sparsely available due to cumbersome offline analysis which only occurs with significant time delay (days) between sampling and measurement. Especially large time delays can be a limitation for the responsiveness of the control hierarchy to maintain the process at or close to optimal process performance.

At the different levels of the control hierarchy models of different detail are applied for specific purposes. Using the same data as input different models provide different outputs depending on the purpose of the model. Having the control hierarchy and the categorization of process models in mind will aid to focus on a specific purpose during model construction in order to obtain the desirable (model) performance.

2.2 Different Model Types

The previous section presented the information and control hierarchies and discussed how models shaped as observers and predictors can serve the purpose of refining and communicating information between the layers in the two hierarchies. It was also discussed how models are used in the layers of the control hierarchy in order to control and guide the process towards a desired objective.

In this section the focus is turned to the different model types which are relevant in this thesis, and that can be used in the aforementioned information and control hierarchies. A short description is provided for the main model categories. Traditionally model types can be separated into three groups:

- White-box models
- Black-box models

- Grey-box models

This separation is illustrated in figure 2.4. The color coding refers to the degree in which the true mechanisms of the system are reflected in the model equations. In other words, a black-box model is constructed without including the known mechanisms of the system, whereas a white-box model reflects the present knowledge of the mechanisms behind the system and their dynamics. In this sense grey-box models contains all the modeling types between the two extreme groups.

2.2.1 White-box Models

Examples of white-box model types are first principles engineering models which are based on conservation laws regarding mass and energy and characterized by being knowledge driven *i.e.* based on knowledge and experience of how the mechanisms of a given system are understood and interrelated. Included in this group are among others:

- Algebraic equations
- Differential equations (ordinary, non-linear, partial)
- Difference equations (ordinary, non-linear)

Even without data of a given system it is possible to construct knowledge driven models. For this reason, knowledge driven models are preferred as educational tools and in other knowledge sharing activities. When applied in processes the performance of white-box models can deteriorate if noisy or uncertain data is used as input to the models. Therefore the construction of suitable filters to reduce or attenuate noise is a prerequisite before nosiy or uncertain process data can be used for modeling, monitoring, control and optimization.

2.2.2 Black-box Models

Complementary to the knowledge driven model types are the entirely data driven model types *e.g.* neural networks and other empirical datafitters. Here the term *black box models* should be understood in a broad sense rather than the use of the term in relation to modeling of physiological pathways *e.g.* Nielsen and Villadsen (1994).

Application of this group of models is to a large extent dictated by the information in the available measurement data. Clearly it is not possible to build a model of a given system without data of how the system behaves and performs. The data driven models are attractive for a first analysis. Under these circumstances the data driven modeling approach is much simpler, although it can still be quite a challenge to construct a model suitable to fulfill the desired purpose.

Examples of data driven models include among others:

- Neural networks (a number of different topologies (structures) exist)
- Time series models (ARMA(X), AR(X), CVA ...)[1]

[1] ARMAX: AutoRegressive Moving Average with eXogenous input. ARX: AutoRegressive with eXogenous input. CVA: Canonical Variate Analysis.

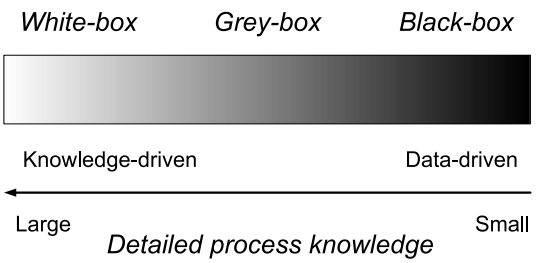


Figure 2.4. Model type categories: White-, grey- and black-box models, depending on the amount of mechanistic or intrinsic knowledge used in the model construction. The color code refers to the transparency of the modeling *box*.

- Chemometrical models (PCA, PCR, PLS ...)[2]

2.2.3 Grey-box Models

The need for grey-box models emerge, where some knowledge of mechanisms and dynamics of a system is available, yet not sufficient to fulfill the desired purpose of the modeling. Under such conditions, it is desirable to combine knowledge and data such that mechanistic models are used to describe the known relations, while data driven structures are used to describe the unknown relations in the mass and energy balances.

An example of grey-box models could be the use of a neural network to use process data in order to estimate unmeasured process variables, which in turn is used as input to a first principles engineering model comprised of differential equations. Another example of grey-box modeling is the use of an modeling framework which incorporate terms that can handle the stochastic nature of process data. One such approach is state space models build on stochastic differential equations (SDE's) to form grey box stochastic models (Kristensen, 2003).

2.2.4 Summary

The review in the above section has provided a very short introduction to how different modeltypes can be used to describe a given problem depending on the level of knowledge and information available on the mechanisms and dynamics of the system.

2.3 A Vision for Model-based Monitoring and Control at Novo Nordisk

In a biopharmaceutical company such as Novo Nordisk surprisingly few soft sensors are used in the analysis, monitoring, control and optimization of the processes along the production lines. The reason for this can possibly be found in the role quality control and quality assurance play in the development and production activities. At

[2]PCA: Principal Component Analysis. PCR: Principle Component Regression. PLS: Partial Least Squares.

a first glance this might seem contradictory, since monitoring and control is intrinsically related to quality assurance and quality control. The explanation for this possible contradiction is found in the activities and documentation that are related to the involvement of the two areas. Rather than attempting to obtain high product quality through robust design, monitoring, control and operation of the process, accepting multivariable process dynamics, the quality of a product is often evaluated solely on the measurement of product purity and concentration. Therefore in a market, where first-to-market is associated with large profits and a dominating position in the market, the focus is on fast development of a reasonable process, where product is released from quarantine by satisfying offline quality control measurements; there is seldom believed to be time available for using modeling in such a market, to develop online quality control indicators such as soft sensors.

However a new approach has been proposed by the U.S. Federal Drug Administration (USFDA) to regulation and supervision of the pharmaceutical industry (USFDA, 2004a). With their vision of a risk-based approach to complement current Good Manufacturing Practices (cGMP's) for the pharmaceutical industry the USFDA will encourage the industry to embrace modern manufacturing tools provided by advances in manufacturing sciences, quality management systems and risk management, to facilitate more robust manufacturing hereby enabling and ensuring production of high quality pharmaceuticals and support continuous process improvement.

One aspect of the risk-based approach to cGMP is the use of process analytical technologies (PAT's) for monitoring and control of industrial processes, where process understanding and process modeling is an integral part of the PAT framework. The underlying incentive for PAT is that: *"quality can not be tested into products; it should be built-in or should be by design"*(USFDA, 2004b) ... of the product and of its production process.

In its guidance for industry on PAT (USFDA, 2004b) the USFDA recognises that the PAT framework is not limited to the purchase and installation of new advanced measurement equipment and instruments, but also includes the use of multivariate modeling methods for design, data acquisition and data analysis as well as for monitoring and control of the processes using the new information provided by process analyzers as well as the traditional process measurements (*e.g.* material flows, temperature, pressure *etc.*). For the continuous improvement of the process understanding and knowledge management, USFDA furthermore mentions the importance of developing tools able to support the data mining of multifactorial relationships as well as evaluate performance of these relationships in different possible scenarios. In summary, the use of models is a central part in the development of a PAT strategy and the PAT framework furthermore facilitates, supports and encourages the extended use of measurements and modeling in the pharmaceutical industry to ensure improved prodcut quality control.

2.3.1 Software Sensors Used in Production Plants

Application of process soft sensors is evident for process monitoring and control. Primarily the information from a the soft sensor can be used in models for control, tuning and optimizing the process. However in many cases the generated information

from a soft sensor can be made available to other parts of the production line through IT-systems such as manufacturing execution/information systems (MES/MIS). In this sense process soft sensors can provide information for scheduling of subsequent unit operations and thereby provide a possibility for tuning and optimizing not only of the single unit, but ideally of the whole production line.

To illustrate the possible application of process sensors in an industrial production framework, a case from the pharmaceutical industry is presented. Figure 2.5 shows a schematic illustration of a production line, consisting of: upstream, downstream and formulation of an active pharmaceutical ingredient (API). From the figure, it can be seen that a number of separate yet sequentially interconnected processes serves a common purpose, namely timely and cost-effective production of high quality API. Unfortunately it is common that relevant information is not exchanged timely and efficiently between these production plants as illustrated in figure 2.6. An example of such lack of communication and exchange of information could be reporting of minor process disturbances that has occurred upstream with possible minor effects on the downstream processing or characteristics of the materials in the process line *e.g.* concentration, composition *etc.*

Why is this so? Modern production and process information systems allow access to historical data bases through company intranet, rendering it possible for personnel at different production plants to analyze not only their own process data, but also the data from other production plants, which in figure 2.6 can be interpreted as intercommunication between the *data* layer of the information hierarchy. Communication also exists at the *information* and *knowledge* layers *e.g.* meetings between personnel from different plants for the exchange of experience and knowledge of the processes in the production line. The reason for the horizontal spacing within the layers in figure 2.6 is to highlight that despite the potential availability of data from other plants, very seldom work is carried out to analyze, let alone implement monitoring and/or control, how data from other plants can be used to generate extended

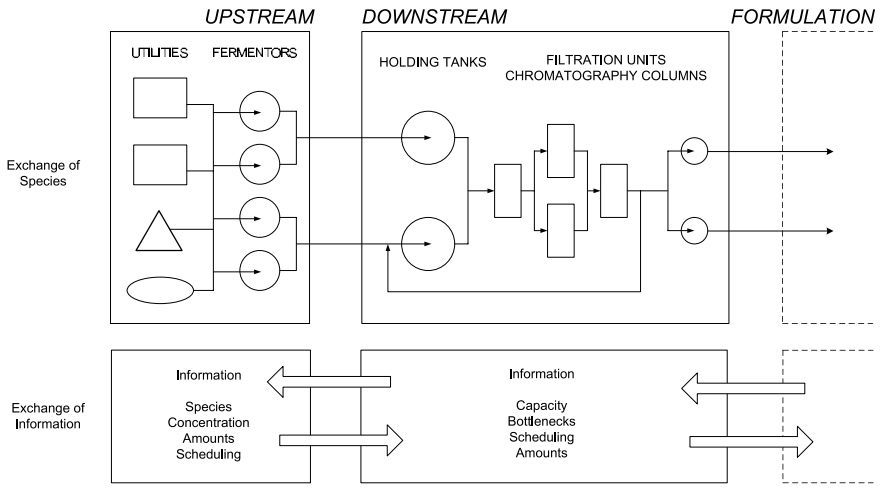


Figure 2.5. Schematic illustration of the exchange of species and information between upstream and downstream processes at a pharmaceutical production site. Examples of information that could flow upstream are capacity status and demands for minimal product concentration due to processing problems.

information and knowledge of the behavior of the production line as a whole. To sum up the discussion above, information exchange should ideally occur at every level of the information hierarchy. Different plants should be able to obtain relevant process information, measurements and data at their desire in order to develop their own soft sensors and tools to monitor, control and optimize the production line. Knowledge and experience from other plants should also be included in the model development, maintaining a holistic approach to improve the performance of the production line as a whole.

Whereas the use of soft sensors for vertical communication at individual plants should be obvious, it is believed that the biggest challenge and also the biggest potential for improving the production line as a whole, will be found when addressing horizontal communication between plants. Building model-based soft sensors seems to be an obvious starting point in an effort to construct plant-wide monitoring and control. Software sensors can facilitate horizontal communication by *e.g.* providing tools to monitor a given production batch while it is still physically being processed at an upstream plant.

The **vision** for Novo Nordisk is that by addressing the development and use of modelbased soft sensors for exchange of information and control of the production line, the company will be able to:

- respond quicker to process disturbances, thereby minimizing their effects on downstream processing
- document and demonstrate detailed process understanding, which can be a prerequisite for parametric release[3] of pharmaceutical products
- improve process performance by continuously developing, maintaining and improving process models for analysis, monitoring, control and optimization
- attain a holistic approach to process optimization through information and knowledge sharing between different plants

The model development activities should not be limited to the production sites, but should already be initiated and addressed during process development and pilot plant trials and seen as an activity that occurs in parallel with continuing process development.

2.3.2 Obstacles to the Implementation of Model-based Monitoring, Control and Optimization

The tight regulation imposed on production of APIs by goverment authorities calls for substantial validation and documentation of the consequences related to a change

[3]Parametric release is defined by the Working Party on Control of Medicines and Inspections as: *"A system of release that gives the assurance that the product is of the intended quality based on information collected during the manufacturing process and on the ecompliance with specific GMP requirements related to Parametric Release"* (Working Party on Control of Medicines and Inspections, 2001). In other words, if a given comprehensive set of in-process tests and control of critical process parameters are within specifications, this can provide greater assurance that the finished product meets specifications than cumbersome testing of the finished product quality.

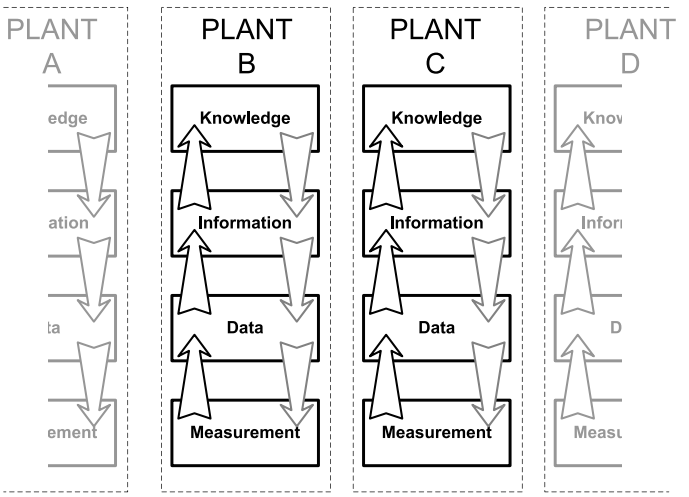


Figure 2.6. Illustration of how information hierarchies exist at different production plants (A, B, C, D). (For details see text to figure 2.2). The boundries between plants exemplified by dashed lines are results of organizational as well as physical boundries. Horizontal exhange is highly relevant and should be carried out through communication between personnel at the different plants, but more importantly by providing access to data bases and documented experience *e.g.* reports, memos *etc.*

.

in operation of a process unit[4]. This severely hinders the implementation of soft sensors for other purposes than monitoring in existing plants, since the expected benefits from a change in operation must exceed the costs of documentation of consequences, possibly resulting in a revalidation [5] of the process, while not adding further constraints on the operation. However it is still through development of soft sensors based on information from existing plants that experience and process insight can be gathered for use in designing new plants and operations. The development of sensors should be an integrated part of establishing a new process, starting in the laboratory, through pilot plant to the final production scale validation of the process. The PAT initiative by USFDA provides the incentive for the biopharmaceutical industry to embrace continuing development through application of hardware and soft sensors both in current production facilities, but especially during development of new production process for novel pharmaceuticals.

A number of key technological aspects must be addressed in order to design process soft sensors meeting the requirements for corrective, preventive and optimizing action of the pharmaceutical industry. A major challenge is to build a mathematical model which exhibits reliable predictive capabilities in a large region of operating

[4]The term *validated process* is used to describe a process for the production of a given product, where the regulatory authorities have received extensive documentation of the process design, performance and ability to produce the product at a given quality, including acceptance ranges for all critical process parameters and variables.

[5]When large changes to a validated process is introduced *e.g.* as a consequence of process optimization, regulatory authorities can require a revalidation of the process in order to verify that the proposed changes have not had any influence on the product quality. A revalidation is considered to be a very time and ressource consuming activity since extensive documentation is required on product quality, new accepantance ranges and robust process performance.

conditions. Furthermore it is required that the reliability of the sensor output most be improved and combined with robust performance with respect to handling of constraints. In this framework filtering of input data to the soft sensor should occur as should a validation of the input data. Such tools should be integrated into the soft sensor along with methods for fault detection, diagnosis and handling. Finally strategies for continuous evaluation and recalibration of model parameters should be available and integrated with a strategy for model improvement when needed.

It is important to stress that the understanding of process dynamics and operation is still maintained with the production floor personnel but the overall effects of changes in operation conditions will become more transparent and more quantitative for business decision makers with the introduction of appropriate soft sensors and process models. Process soft sensors support both these control levels by primarily providing better and more specific information of variables in a process unit, and secondarily by providing quantified information for the business decision makers thus promoting communications between the two levels in the control hierarchy closer together.

2.4 Summary

The aim of this chapter has been to provide a brief and general overview of the concepts behind modeling and behind typical model types. The concept of a soft sensor was introduced as a tool to provide online information on unmeasured or seldomly measured process variables, possibly replacing slow and cumbersome offline chemical analysis and provide quicker and cheaper estimates. Both an information hierarchy and a control hierarchy was presented in order to explain how the intended application should be kept in mind during model construction, *i.e.* the modeling must be purpose-driven.

A vision for the application of models for model-based monitoring and control at Novo Nordisk was presented with the construction and use of soft sensors as a tool to support quicker response to process disturbances, document and demonstrate detailed process understanding, improve process performance and attain a holistic approach to process optimization through information and knowledge sharing between sequential process units and between different plants. It was stressed that the design and implementation of soft sensors is not trivial, however it is still through development of soft sensors based on information from existing plants that experience and process insight can be gathered for use in designing new plants and operations.

In chapters 6, 8 and 9 a number of models are constructed for use in soft sensors.

3

Physiology of *Saccharomyces cerevisiae*

For the construction of mechanistic models of a cultivation process it is necessary to have an understanding of the interplay between the numerous and intertwined reactions responsible for making it possible for cells to survive and reproduce *i.e.* the physiology of the microorganism. Identifying the most important aspects of the physiology of *Saccharomyces cerevisiae* with respect to cultivation under the desired process condition and modeling these aspects using mechanistic models have been a cornerstone in the analysis and understanding of growth dynamics of the yeast (Sonnleitner and Käppeli (1986); Sonnleitner and Hahnemann (1994); Nielsen and Villadsen (1994); Stephanopoulos *et al.* (1998); Duboc *et al.* (1998); Pham *et al.* (1998)).

The purpose of this chapter is to provide an overview of the physiology of *S. cerevisiae* that is relevant for understanding the investigations and subsequent modeling in later chapters on cultivations conducted for this thesis. Since glucose has been the primary carbon and energy source, and since aerobic conditions have been maintained throughout the cultivations, the emphasis is mainly on the metabolic pathways that are active during aerobic growth on glucose and ethanol. Special attention will be given to the overflow metabolism referred to as the *Crabtree* effect (Pronk *et al.* (1996); Postma *et al.* (1989a)), since significant acetate formation has been observed in some of the cultivations.

The first section will provide a very brief description of the cytology of *S. cerevisiae* *i.e.* the architecture of the microorganism. Then follows an overall description of the purpose of the metabolism of *S. cerevisiae*. The metabolism can roughly be divided into two parts: the anabolism described in section 3.3 and the catabolism described in section 3.4. Only a short description of the anabolism is given here, while a deeper description is given in chapter 5. More details of the catabolism is presented. In section 3.4.1 are the topics of transport and phosphorylation of substrates presented. In section 3.4.2 the glycolysis is described as a source of metabolic energy and of precursors for the biosynthesis *i.e.* construction of biomass. Sections 3.4.3 and 3.4.4 present the tricarboxylic acid cycle and the oxidative phosphorylation, respectively. The fermentative pathway responsible for the production of ethanol is described in section 3.4.5, while the pathway for production of glycerol is presented in section 3.4.6. Two further pathways are presented, namely the gluconeogenesis in section 3.4.7 and the glyoxylate cycle in section 3.4.8. The chapter ends with a brief summary.

3.1 Cytology of *Saccharomyces cerevisiae*

To fulfill the tasks of survival, growth and reproduction under different environmental conditions, the *S. cerevisiae* has evolved over millions of years and has formed separate compartments and organelles within the boundaries of the cell wall to facilitate this adaptation. Figure 3.1 provides a non-exhaustive schematic overview of some of the most important subcellular compartments in *S. cerevisiae*.

The purpose of this section is to provide a brief overview of the cytology of *S. cerevisiae*, where the different subcellular compartments are introduced and their known functions explained. The main source of information for this chapter is Walker (1998).

3.1.1 The Barrier

As all living organisms, the yeast is surrounded by a barrier against the environment. For *S. cerevisiae* this barrier contains three layers: the cell wall, the periplasm and the plasma membrane. The cell wall is the outer barrier and is primarily composed of the polysaccharides glucan and mannan. The purpose of the cell wall is cell protection, shape maintenance, cellular interactions *i.e.* haploid cell fusion to form a diploid cell and specialized enzymatic activities. The periplasm between the cell wall and the inner barrier, the plasma membrane, is the space where secreted (but not excreted) proteins and enzymes reside, whose purpose it is to catalyze the hydrolysis of substrates that can not otherwise cross the plasma membrane. An example of such an enzyme is invertase. The primary function of the plasma membrane in *S. cerevisiae* is to control the passage of hydrophilic molecules between the environment and the cell. The plasma membrane is a lipid bilayer mainly comprised of phosphorlipids this membrane changes both structure and functionality depending on the growth conditions.

Reproduction is mostly carried out by budding, where the cell wall is weakened at a certain spot as the mother cell reaches a critical size. The bud is expanded by extrusion of cytoplasm and after migration of the new bud nucleus (formed by DNA synthesis and mitosis) and other organelles, the bud is release from the mother cell to become a daughter cell.

3.1.2 The Cytoplasm

The cytoplasm is the aqueous fluid that makes up the most part of the contents of a yeast cell. A part from large organelles and the nucleous (described below), the cytoplasm contains low and intermediate molecular weight components, dissolved proteins and other macromolecules. The ribosomes, formed by ribosomal RNA, are the place for translation of messenger RNA for protein biosynthesis, a vital step in the growth and adaptation of the cell. Amongst a number of other membrane-delimited microbodies in the cytoplasm are the perixomes, which server as a site for oxidase activity, most importantly catalase.

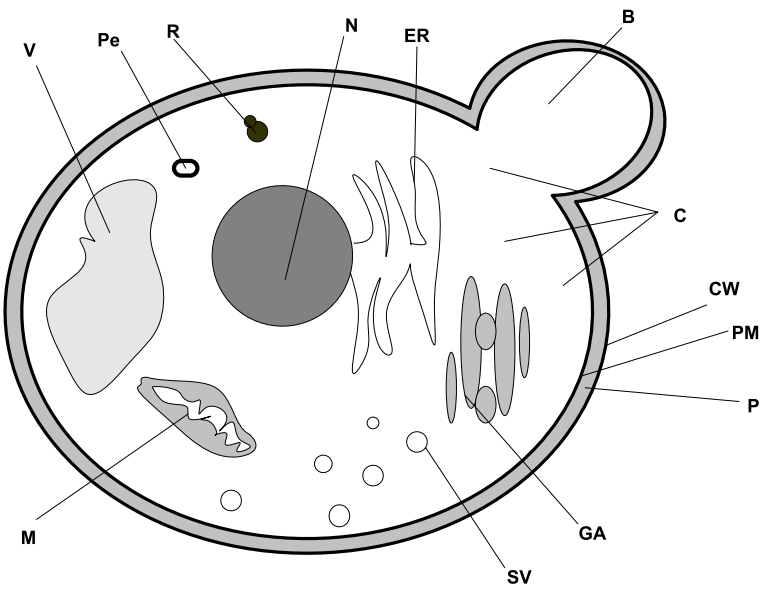


Figure 3.1. Schematic illustration of a *Saccharomyces cerevisiae* cell, see text for details. **B**: bud, **C**: cytoplasm, **CW**: cell wall, **ER**: endoplasmic reticulum, **GA**: Golgi apparatus, , **M**: mitochondrion, **N**: nucleus, **P**: periplasm, **Pe**: perixome, **PM**: plasma membrane, **R**: ribosome, **SV**: secretory vesicle and **V**: vacuole.

3.1.3 The Nucleus

The nucleoplasm is separated from the cytoplasm by a porous double membrane and contains DNA, RNA and basic proteins along with extrachromosomal elements *e.g.* the 2μm-DNA plasmid, as opposed to chromosomal DNA. The nucleus is the site for transcription of DNA which forms ribosomal, messenger and transport RNA. During cell division new DNA is synthesized to create a copy of the chromosomes of the mother cell to be used by the daughter cell.

3.1.4 The Mitochondria

The primary function of mitochondria during aerobic, glucose limited cultivations of *S. cerevisiae* is to synthesize ATP during respiration. The inner of the mitochondria, the matrix, is separated from the cytoplasm by a double membrane. The outer membrane contains enzymes involved in lipid synthesis, while the inner membrane is the site for the cytochromes and the respiratory chain. In the matrix, enzymes of the tricarboxylic acid cycle are found.

3.1.5 The Secretory Pathway

A number of membrane-delimited organelles are involved in the maturation and transport of synthesized proteins to the cell wall. In the endoplasmic reticulum protein folding and glycosylation occur as well as proteolytic activities to remove signal peptides. In the Golgi apparatus further reshaping (folding and enzymatic cutting) of the proteins occur by addition of carbohydrate side chains or removal of parts of the amino acid chain(s). The mature proteins are transported to the

cell wall by secretory vesicles. The vacuoles are the sites for non-specific proteolysis and also serve the function as storage reserve for amino acids and inorganic acids. Chapter 5 provides a more detailed description of the role of the secretory pathway for the production of a recombinant precursor for insulin.

3.2 Purpose of the Metabolism

Simple microorganisms incorporate a vast and intrinsically complicated network of reactions (Stryer (1995); Walker (1998)). Despite this complexity the purpose of metabolism can be formulated quite simply: provide the means to survive, grow and reproduce in the environment.

As mentioned above the focus of this chapter is primarily on the use of glucose as primary carbon and energy source under aerobic conditions. However it is necessary to include descriptions of growth on both acetate and ethanol, since growth on these carbon sources occurs under some process conditions *e.g.* in batch operation following growth on glucose and sometimes may occur following a process disturbance.

The utilization of glucose is illustrated in figure 3.2. Assimilatory processes lead to the formation of biomass, while dissimilation of glucose provides the energy for the formation of biomass. For the assimilatory processes NADPH is needed, and the middle reaction illustrates this special need (Verduyn, 1991). An important message provided by figure 3.2 is the coupling of the different processes by energy carriers (ATP) and redox carriers (NADH/NADPH) and that balancing between the

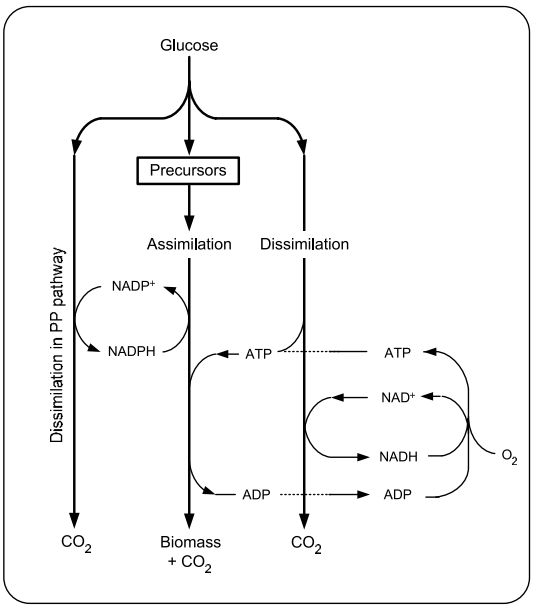


Figure 3.2. Schematic illustration of the central metabolic processes involved in the utilization of glucose as primary carbon and energy source. Metabolism ("dissimilation") of glucose yields NADH which (in respiration) is used to provide large amounts of (free) energy. Another reduction equivalent NADPH is also required for biomass production and produced by conversion of glucose to CO_2. PP: Pentose Phosphate. Adapted from Verduyn (1991).

production and consumption, reduction and oxidation respectively, are important for the cell to survive and grow/reproduce.

Although as being illustrated as separate, the three processes involving the use of glucose, depicted in figure 3.2, are tightly coupled. The more traditional approach of subdividing the growth metabolism into anabolism and catabolism is presented and roughly sketched in figure 3.3 and described below.

3.3 The Anabolism

The purpose of anabolic reactions in the metabolism of *S. cerevisiae* is to supply and support all the essential components and compartments, enabling the survival and growth of the microorganism in the surrounding environment. More specific the anabolism accounts for the production of structural proteins, enzymes, structural polysaccharides, lipids, storage carbohydrates, nucleic acids and cytochromes as well as the assembling of these to form the organelles of the cell (Walker, 1998).

For many industrial applications producing recombinant proteins, it is of paramount importance that the activity of the post-transcriptional modifications are neither limiting growth nor productivity of the desired protein product. Therefore post-translational modifications to the protein precursor, *i.e.* cutting off a leader sequence, folding of the peptide *etc*, have received significantly more attention than the primary metabolism. The modifications during folding, transport and excretion of the protein precursor occurring in the endoplasmic recticulum, the Golgi apparatus and in transport vesicles *i.e.* the secretory pathway, define the possibility of manipulating the precursor into the desired product, in this case an insulin precursor, in the downstream processing, and at the same time avoid or at least minimize by-product formation. A more detailed description of the secretory pathway of excreted proteins is given in chapter 5. To support high and robust productivity of the desired product significant effort has also been devoted to investigate the impacts of medium composition and growth conditions *e.g.* temperature, pH and aeration.

In this thesis modeling of the growth dynamics represented by the primary metabolism *i.e.* the catabolism, has been given most attention. As understanding is developed on how the microorganism directs its flow of carbon from glucose to energy and precursor metabolites, further investigations of the fate of these precursor metabolites can be pursued.

The precursor metabolites referred to in figure 3.3 are listed in table 3.1 along with the building blocks and macromolecules they in turn lead to. The drainage of these metabolites from pools in the primary metabolism calls for mechanisms replenishing the pools, which is carried out by the anaplerotic pathways *e.g.* the glyoxylate cycle (see section 3.4.8).

3.4 The Catabolism

The purpose of the catabolic reactions is to provide metabolic energy and precursors for the anabolism. In the following the major catabolic pathways will be presented and their influences on the overall metabolism of the microorganism discussed. Although the combined transport and phosphorylation of glucose during uptake is

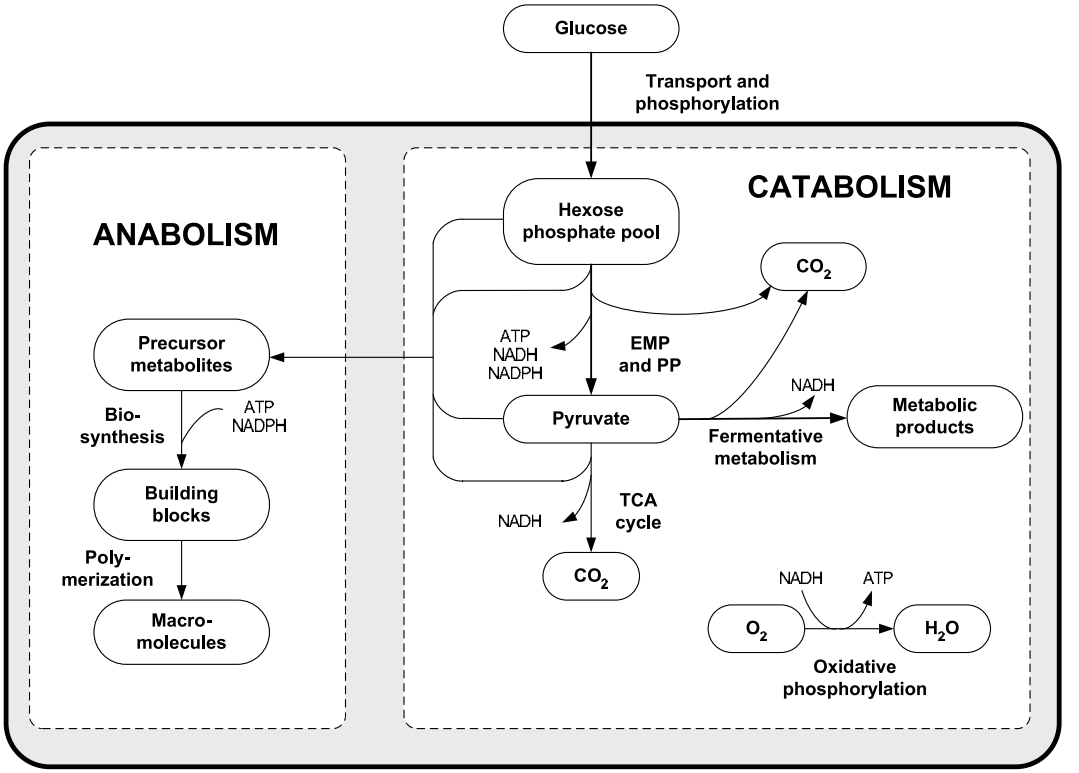


Figure 3.3. Overall structure of cell synthesis from glucose. The glucose is transported through the cell wall, and phosphorylated to enter the hexose phosphate pool. Through reactions in the Embden-Meyerhof-Parnas (EMP) or the Pentose Phosphate (PP) pathway the hexose phosphates are either converted to pyruvate or used in the biosynthesis of carbohydrates. Pyruvate can be converted to carbon dioxide in the respiratory tricarboxylic acid (TCA) cycle, or to metabolic products through fermentative pathways. Reducing equivalents in the form of NADH generated in the EMP and TCA cycle can be oxidized in oxidative phosphorylation. A number of intermediates from the EMP, PP and TCA cycle pathways serve as precursor metabolites for biosynthesis of building blocks, which through polymerization form macromolecules *e.g.* DNA, RNA, proteins, lipids and carbohydrates. ATP and NADPH formed by catabolic reactions are used for anabolic reactions. Adapted from Stephanopoulos *et al.* (1998).

not directly considered as being part of the catabolism, the mechanism will also be described.

3.4.1 Transport and Phosphorylation

Walsh *et al.* (1994) proposed the existence of a multicomponent glucose uptake system with a variable affinity for glucose in *S. cerevisiae*. In general the system can be formulated as two interdependent contributions as described by Postma *et al.* (1989a), namely a high-affinity transporter ($K_{m,h} \approx 1$ mM) and a low-affinity transport system ($K_{m,l} \approx 20$ mM). The uptake is known to be assisted by a phosphorylation of glucose leading to the entry point of the glycolysis, namely glucose 6-phosphate as shown in figure 3.4.

3.4.2 Glycolysis

The term glycolysis is used to describe the main pathways dissimilating glucose 6-phosphate, and is here represented by the Embden-Meyerhof-Parnas (EMP) and Pentose Phosphate (PP) pathways (see figures 3.4 and 3.5 respectively). The EMP pathway is generally considered the primary route for dissimilation of glucose to pyruvate during aerobic growth on a glucose-limited substrate. The overall stoichiometry of this pathway can be written as:

$$\text{Glucose} + 2\ \text{ADP} + 2\ \text{P}_i + 2\ \text{NAD}^+ \rightarrow 2\ \text{Pyruvate} + 2\ \text{ATP} + 2\ \text{NADH} + 2\ \text{H}^+ \quad (3.1)$$

From this reaction scheme it is seen that the sequence of reactions supply metabolic energy to the microorganism in the form of ATP at the expense of the oxidized form of the cofactor NAD^+ and glucose. Furthermore some of the products of the intermediate reactions serve as precursor metabolites as mentioned above.

The reaction scheme of the PP pathway is shown in figure 3.5. The PP pathway has been recognized to serve both an oxidative[1] (equation 3.2) as well as an anaplerotic function (equation 3.3) (Stephanopoulos *et al.*, 1998):

$$\text{Glucose 6-phosphate} + 12\ \text{NADP}^+ \rightarrow 6\ \text{CO}_2 + 12\ \text{NADPH} + 12\ \text{H}^+ \quad (3.2)$$

$$\text{Glucose 6-phosphate} + 2\ \text{NADP}^+ \rightarrow \text{Ribose 5-phosphate} + \text{CO}_2 + 2\ \text{NADPH} + 2\ \text{H}^+ \quad (3.3)$$

Especially the oxidative function (equation 3.2) is interesting seen in the perspective of figures 3.2 and 3.3, since this is a major source for NADPH reducing equivalents needed in the biosynthesis. Equation 3.3 illustrates the generation of ribose sugars, which are important for the synthesis of nucleotide precursors for nucleic acids, RNA and DNA as well as nucleotide coenzymes such as NAD^+, $NADP^+$ and FAD (Walker, 1998).

[1] Glucose 6-phosphate is oxidized to CO_2.

Table 3.1. 12 precursor metabolites, their origin in the primary metabolism, as well as their role in the synthesis of biomass. Adapted from Stephanopoulos *et al.* (1998) and Walker (1998). Abbreviations are listed at the end of the thesis.

Precursor	Origin	Building block/Macromolecules
Glucose 6-phosphate	EMP	Polysaccharides, Inositol
Fructose 6-phosphate	EMP	UDP-NAG, UDP-NAM
Glyceraldehyde 3-phosphate	EMP	Lipids
3-phosphoglycerate	EMP	Amino acids
Phosphoenolpyruvate	EMP	Amino acids
Pyruvate	EMP	Amino acids
Erythrose-4-phosphate	PP	Amino acids
Ribose-5-phosphate	PP	Amino acids
Acetyl-CoA	EMP and TCA	Lipids, Sterols
α-ketoglutarate	TCA	Amino acids
Oxaloacetate	TCA *etc.*	Amino acids
Succinyl-CoA	TCA	Hemes

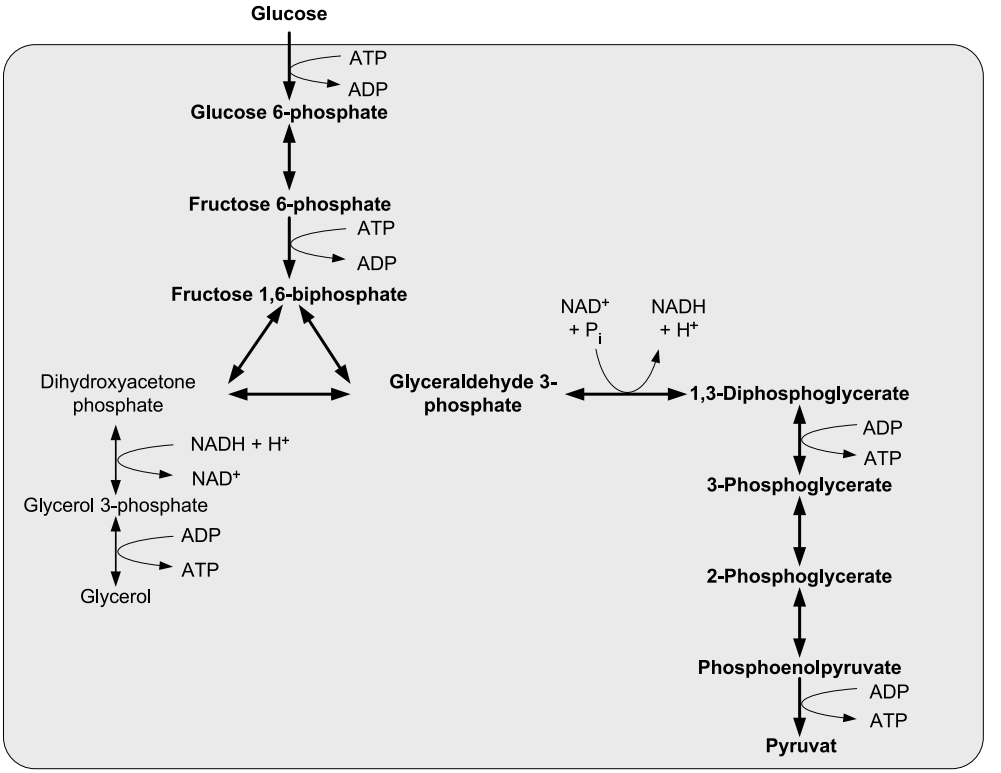


Figure 3.4. The Embden-Meyerhof-Parnas (EMP) pathway including the glycerol pathway. The frame signifies the cell wall, where the uptake of glucose is assisted by phosphorylation to form glucose 6-phosphate. The arrows indicate reactions all catalysed by different enzymes. Some of the reactions are accompanied by consumption of metabolic energy in the form of ATP, while others produce metabolic energy by adding an additional high-energy phosphate group to ADP to form ATP. Two of the reactions involve the coenzyme NADH/NAD^+. During balanced oxidative growth of glucose the reaction path is indicated by the bold arrows, converting glucose to pyruvate. Note that all reactions except two are reversible as illustrated by the arrows in the figure. Irreversible reactions are found between fructose 6-phosphate and fructose 1,6-biphosphate, and between phosphoenolpyruvate and pyruvate.

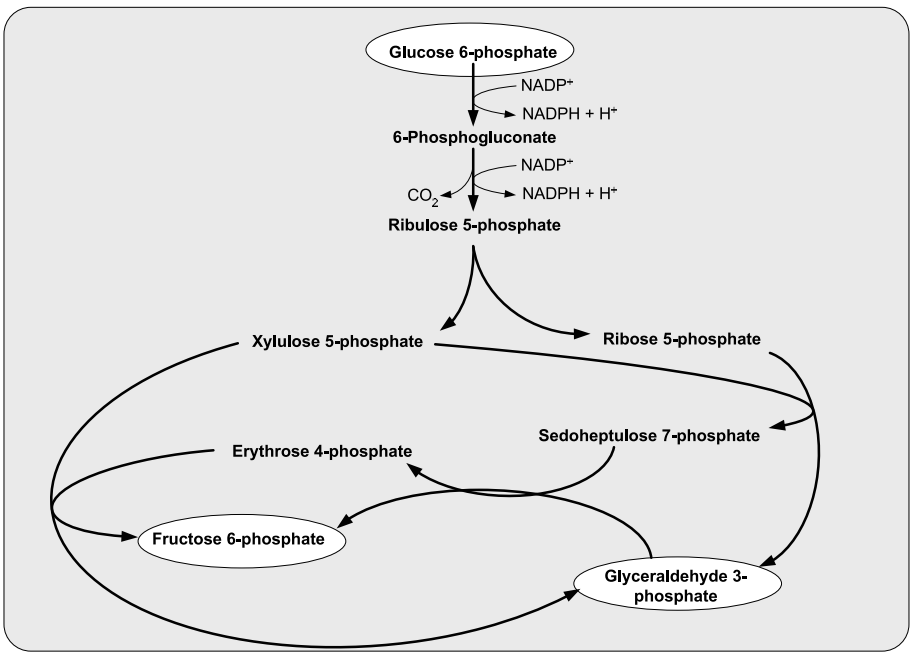


Figure 3.5. The Pentose Phosphate (PP) pathway. The ellipses illustrate that these components are also part of the EMP pathway and form the links between the two pathways. The reduction equivalent NADPH is formed by reduction of $NADP^+$. As illustrated in figure 3.2, NADPH is involved in anabolic reactions for the formation of biomass.

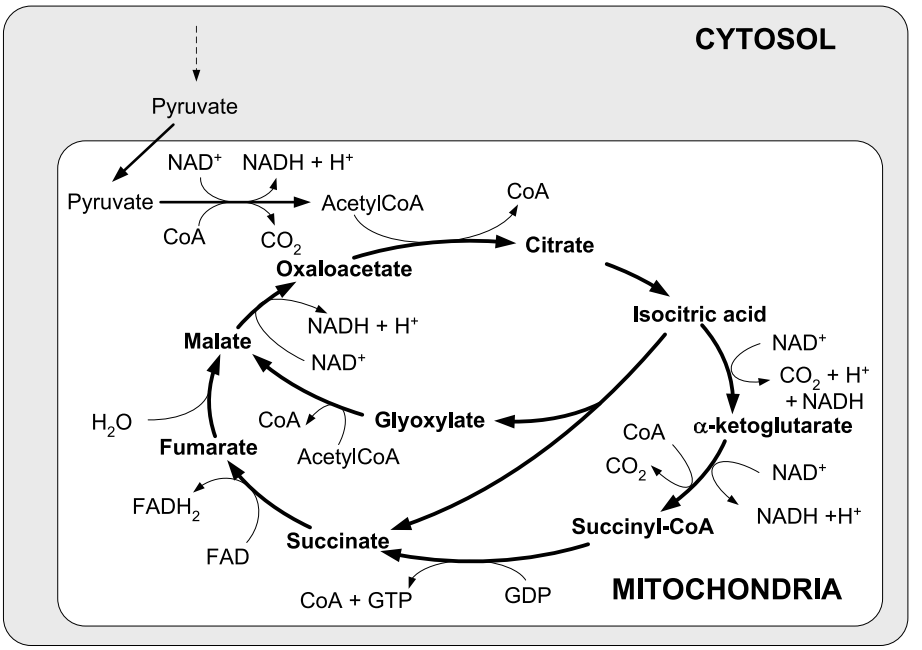


Figure 3.6. The Tricarboxylic Acid Cycle (TCA) and the glyoxylate cycle. The activities of the TCA cycle occur exclusively in the mitochondria. Intermediates of the TCA are used as precursor metabolites (see table 3.1). The glyoxylate cycle is essential for growth on ethanol and acetate and is furthermore used to replenish intermediates of the TCA cycle used for biosynthesis of amino acids and nucleotides to ensure continued operation of the TCA cycle.

3.4.3 Tricarboxylic Acid Cycle

The TCA cycle results essentially in the total dissimilation of pyruvate to CO_2 as illustrated in figure 3.6 and reflected by equation 3.4:

$$\text{Pyruvate} + \text{GDP} + \text{FAD} + 4\ \text{NAD}^+ + \text{P}_i \rightarrow 3\ \text{CO}_2 + \text{GTP} + 4\ \text{NADH} + 4\ \text{H}^+ + \text{FADH}_2 \tag{3.4}$$

From the equation it can be seen that a large number of NADH and $FADH_2$ reducing equivalents are generated when the TCA cycle is active.

As mentioned in section 3.3 four of the metabolites in the TCA cycle also act as precursor metabolites for biosynthesis. Therefore additional reactions are necessary to replenish the TCA cycle. These are referred to as anaplerotic pathways, and include among others the carboxylation of pyruvate to oxaloacetate by the enzyme pyruvate decarboxylase (see figure 3.7), the carboxylation of phosphoenolpyruvate to oxaloacetate by the enzyme phosphoenolpyruvate carboxylase and the glyoxylate shunt (see section 3.4.8).

3.4.4 Oxidative Phosphorylation

Both the EMP pathway (equation 3.1) and the TCA cycle (equation 3.4) result in a net production of reducing equivalents NADH and $FADH_2$. In the presence of oxygen the reoxidation of the two co-factors can occur in the mitochondria in combination with the addition of a high-energy phosphate group to ADP, thus forming ATP.

The theoretical oxidation of NADH and $FADH_2$ is:

$$\text{NADH} + 1/2\ \text{O}_2 + 3\ \text{ADP} \rightarrow \text{NAD}^+ + \text{H}_2\text{O} + 3\ \text{ATP} \tag{3.5}$$

$$\text{FADH}_2 + 1/2\ \text{O}_2 + 2\ \text{ADP} \rightarrow \text{FAD} + \text{H}_2\text{O} + 2\ \text{ATP} \tag{3.6}$$

The number of moles of ATP formed for each oxygen atom used in the oxidative phosphorylation is referred to as the P/O ratio. Since the ratio of FAD_2 to NADH is not constant at different operating conditions the P/O ratio is not constant. The NADH formed in the cytoplasm (*e.g.* in the EMP pathway) can not pass the mitochondrial membrane, and the oxidation to NAD^+ is coupled to the reduction of FAD to FAD_2. Furthermore, incomplete coupling between the oxidation and the phosphyrolation in the respiratory chain, renders the operational P/O ratio to be different from the theoretical P/O ratio (Nielsen and Villadsen, 1994). van Gulik and Heijnen (1995) report the operational P/O ratio to be 1.2-1.3.

3.4.5 Fermentative Pathway

Both mitochondrial and cytosolic reoxidation of NADH can also be achieved by the action of alcohol dehydrogenase converting acetaldehyde to ethanol. This reaction is however not assisted by the formation of ATP as is the case in oxidative phosphorylation. Acetaldehyde is the product of a decarboxylation of pyruvate as illustrated in figure 3.7. It can be converted to acetate by acetaldehyde dehydrogenase, involving either NAD^+ or $NADP^+$ for the oxidation, depending on the compartment in which the reaction occurs. Acetyl-CoA can not be transported across the inner

mitochondrial membrane, unless through the so-called carnitine shuttle (Swiegers *et al.* (2001); Kispal *et al.* (1991)).

Due to the central positioning of pyruvate at the point linking glycolysis, the TCA cycle, the fermentative pathway and the gluconeogenesis (see section 3.4.7), a lot of work has been carried out to elucidate the dynamics of the reactions and the enzyme systems controlling the distribution of fluxes out of the pyruvate pool (see Pronk *et al.* (1996) for a review). Although referred to as a pool, pyruvate does not appear exclusively in one compartment, but appears both in the cytosol and in the mitochondrial matrix. Linked to the distribution of fluxes at the pyruvate branch point is the *Crabtree* effect, which describes the occurrence of alcohol fermentation under aerobic conditions (Barford and Hall (1979); Postma *et al.* (1989a); Pronk *et al.* (1996)). Two versions of the Crabtree effect are often referred to (Barford and Hall (1979); Pronk *et al.* (1996); Walker (1998)). The long-term Crabtree effect accounts for aerobic alcoholic fermentation during high growth rates, both for growth under sugar-limitation and at sugar excess. The short-term Crabtree effect accounts for transient aerobic alcoholic fermentation observed at a transition from sugar-limited growth to a sugar excess.

The overall reaction from pyruvate to ethanol becomes:

$$\text{Pyruvate} + \text{NADH} + \text{H}^+ \rightarrow \text{Ethanol} + \text{NAD}^+ + \text{CO}_2 \tag{3.7}$$

and to acetate by:

$$\text{Pyruvate} + \text{NAD(P)}^+ \rightarrow \text{Acetate} + \text{NAD(P)H} + \text{H}^+ + \text{CO}_2 \tag{3.8}$$

where NAD(P) indicates that two isoenzymes can catalyze the reaction from acetaldehyde to acetate using either NAD^+ or NADP^+ as discussed above.

3.4.6 Glycerol Production

As illustrated in figure 3.4 production of glycerol in a pathway that branches off from the EMP pathway. Under anaerobic conditions production of glycerol serves the function of reoxidizing NADH formed by activities of the EMP pathway concurrently with ethanol formation, since oxidative phosphorylation can not reoxidize the NADH in the absence of oxygen (Schulze, 1995). As can be seen from equation 3.9 the formation of 1 mole glycerol from glucose is assisted by the net hydrolysis of 1 mole ATP. In other words, glycerol serves as a redox sink but also as an energy sink.

$$\text{Glucose} + 2\ \text{ATP} + 2\ \text{NADH} + 2\ \text{H}^+ \rightarrow 2\text{Glycerol} + 2\ \text{ADP} + 2\ \text{P}_i + 2\ \text{NAD}^+ \tag{3.9}$$

3.4.7 Gluconeogenesis

The term gluconeogenesis is used to describe the metabolic processes enabling conversion of pyruvate to glucose (Walker, 1998). This is relevant during (aerobic) growth on non-carbohydrate substrates *i.e.* ethanol, acetate and glycerol. As mentioned above a number of the intermediates in glycolysis serve as precursor metabolites for biosynthesis, and these also need to be produced during growth on other substrates. Since two of the reactions in glycolysis are irreversible, gluconeogenic

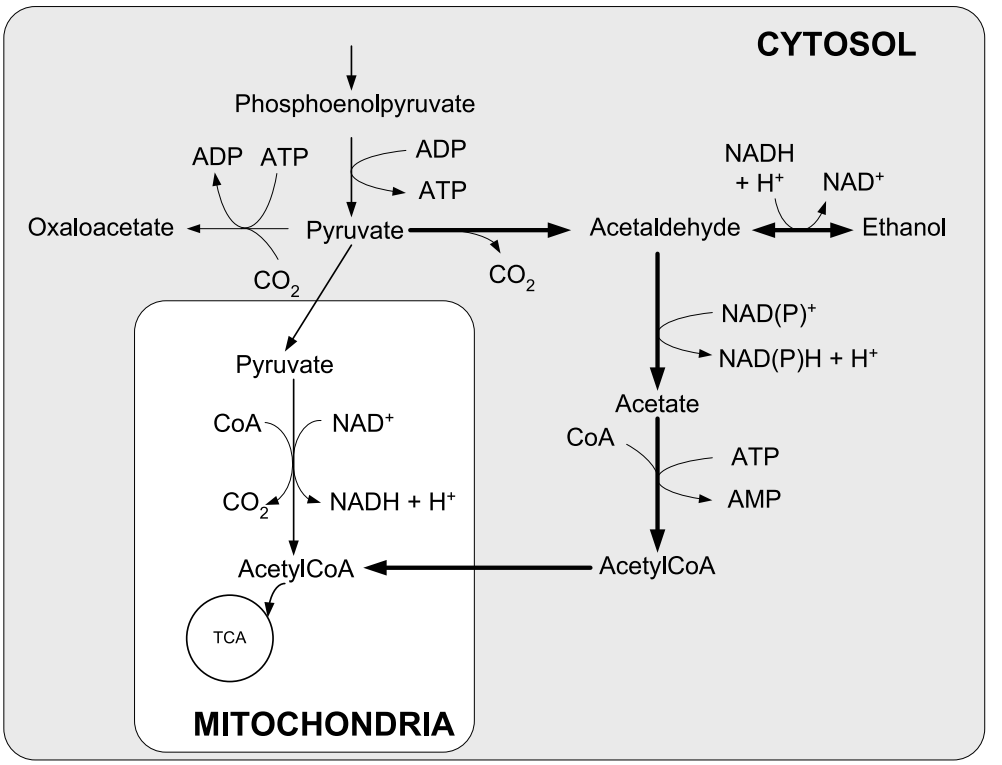


Figure 3.7. The pathways surrounding the pyruvate branch point. Pyruvate can either be converted to oxaloacetate by pyruvate carboxylase, to acetaldehyde by pyruvate decarboxylase or to acetyl-CoA by pyruvate dehydrogenase. Several of the reactions occur in more compartments than shown on the figure (Remize *et al.* (2000); Walker (1998)). The oxidation of acetaldehyde to acetate can be accompanied by either formation of NADPH or NADH depending on the compartment in which the reaction occur.

alternative reactions are present. One of these is the energy-demanding decarboxylation of oxaloacetate to phosphoenolpyruvate:

$$\text{Oxaloacetate} + \text{ATP} \rightarrow \text{Phosphoenolpyruvate} + \text{CO}_2 + \text{ADP} + \text{P}_i \tag{3.10}$$

Another gluconeogenic alternative is the conversion of fructose 1,6-biphosphate to fructose 6-phosphate.

$$\text{Fructose 1,6-biphosphate} \rightarrow \text{Fructose 6-phosphate} + \text{P}_i \tag{3.11}$$

The overall reaction from pyruvate to glucose 6-phosphate becomes:

$$2\ \text{Pyruvate} + 2\ \text{ATP} + 2\ \text{GTP} + 2\ \text{NADH} + 2\ \text{H}^+ \tag{3.12}$$

$$\rightarrow\ \text{Glucose 6-phosphate} + 2\ \text{ADP} + 2\ \text{P}_i + 2\ \text{NAD}^+ \tag{3.13}$$

3.4.8 Glyoxylate cycle

As mentioned above and shown in figure 3.6, the function of the glyoxylate cycle is primarily anaplerotic and is essential for growth on C_2 components, such as ethanol and acetate, resulting in a net formation of C_4 dicarboxylic acids, namely malate and succinate. The steps in the glyoxylate cycle are:

$$\text{Isocitrate} \rightarrow \text{Succinate} + \text{Glyoxylate} \tag{3.14}$$

$$\text{Glyoxylate} + \text{Acetyl CoA} \rightarrow \text{Malate} + \text{CoA} \tag{3.15}$$

Malate can then be converted into oxaloacetate, the starting point for gluconeogenesis reactions (see equation 3.10).

3.4.9 Summary

It has been the aim to provide a brief description of the catabolism in *S. cerevisiae* for formulating simple mass balances and calculating flux distributions. The metabolic pathways of *S. cerevisiae* are in general far more complex and intertwined than has been portrayed here, especially due to the existence of isoenzymes and the distribution of these among different organelles, which complicates the understanding of the flux distributions in the cell. Additional complexity is added when the behavior of industrial strains are investigated; here only limited information on the physiology of the organism and on the concerted or isolated kinetics of the metabolic reactions are available.

In the following chapter the presence of weak organic acids in the culture broth and their effects on the growth of *S. cerevisiae* will be illustrated through a literature review. The aim is to provide the background for interpreting the observations made in some of the cultivations, to be presented in later chapters. Following the chapter on weak organic acids is a review on the genetic modifications that have been carried out on one of the two recombinant strains of *S. cerevisiae* that have been cultivated in the work behind this thesis.

4

Transport and Effects of Organic Acids

The previous chapter on the general physiology of *S. cerevisiae* briefly mentioned that a number of weak organic acids are part of the metabolism *e.g.* pyruvic acid, succinic acid, malic acid, fumaric acid, acetic acid and citric acid. The transport and effects of organic acids will be shown in later chapters to have a significant effect on the growth dynamics of one of the recombinant strains investigated in this thesis. The presence of organic acids in the culture broth can have a large effect on the growth of *S. cerevisiae*. The organic acids can either be produced by the organism itself or be present in the environment. Normally during aerobic growth on glucose as limiting substrate in a chemostat with the aim of producing products related to the biomass production, large amounts of organic acids in the culture broth are not anticipated, and therefore seldom measured/monitored. In some of the experiments carried out in this work, surprising process behavior was observed during continuous operation, which was suspected to be a consequence of organic acids being produced by the yeast. Therefore the purpose of this chapter is to provide an overview of the influence the presence of weak organic acids in the culture broth can have on the growth energetics of *S. cerevisiae*.

In order to limit the extent of this chapter, most of the detailed information on cultivation conditions and strains used in the cited literature is not presented here; please refer to the original publications.

Most of the literature is based on the behavior of *S. cerevisiae* at glucose feed concentrations up to 30 g/L and only few detailed descriptions are available on the metabolic profiles of high-density cultivations, which is the focus of this study. In this context the term *high-density* is defined as cultivations with a biomass concentration above 25 g dry weight per L, which during aerobic conditions corresponds to a substrate containing more than 50 g glucose/L. As the glucose concentration of the feed is increased, so are the by-product concentration in the culture broth. During transients occuring at these elevated feed concentrations levels, changes in the distribution of metabolic fluxes can lead to the production of by-products to an extent that activates effects which normally are not observed during transients at lower cell densities.

Normal by-products during aerobic glucose-limited chemostat cultivations are ethanol (mM level) and organic acids such as pyruvate, succinate and acetate at μM levels. At high concentrations these metabolites have been shown to influence growth of *S. cerevisiae* in several ways. Benzoate has been used to study the influence of weak acids on the physiology of *S. cerevisiae*. Benzoate is a weak acid that can not be metabolized by the microorganism and is therefore suitable for the study of

transport of weak acids across the plasma membrane and their effect on the growth energetics of *S. cerevisiae*. These effects are the topic of the following subsections which review effects of acetate, benzoate and ethanol on growth of *S. cerevisiae*, as well as observations made in a number of dynamic experiments reported in literature involving this organism.

4.1 Effects of Acetate

A number of factors influence the transport of acetate across the plasma membrane of *S. cerevisiae*. With acetate (0.5 vol%) as the sole energy and carbon source, Makuc *et al.* (2001) observed a mediated transport system for acetate uptake involving proton symport with $K_m = 0.17 \pm 0.14$ mM and a $V_{max} = 2.27 \pm 0.61$ nmole/s/(mg dry weight) at pH 5.0. These findings correspond to the observations reported by Casal *et al.* (1996), where acetate transport across the plasma membrane during growth on acetate (0.5 vol%) was carried out by mediated transport with $K_m = 0.21 \pm 0.039$ mM and $V_{max} = 0.6$ nmole/s/(mg dry weight) at pH 6.0. They further observed that this transport system was repressed for cells growing on glucose, where initial uptake rates of acetate were below 0.05 nmole/s/(mg dry weight) for concentrations up to 0.1 mM undissociated acetic acid.

Passive diffusion of undissociated acetic acid across the plasma membrane occurs due to a difference in pH between the cytosol (pH 6.8 - 7.2) and the culture broth (pH $\approx$ 5.5-6.0) (Casal *et al.* (1996); Verduyn *et al.* (1990)). Undissociated acetic acid diffusing inward across the plasma membrane is quickly dissociated (pK_a = 4.75) due to the higher pH in the cytosol, resulting in a net transport of acetate across the membrane along with an acidification of the cytosol by the accompanying protons. This acidification is then balanced through proton extrusion by ATPase in the plasma membrane at the expense of ATP (Verduyn, 1991).

When measuring acetate in the culture broth it is therefore important to keep in mind that the primary transport during growth on glucose is diffusion of the undissociated acid. With an intracellular pH $\approx$ 7.0, a pK_a = 4.75 for acetic acid and a culture broth pH = 5.9, the accumulation factor ([acid$_{in}$] / [acid$_{out}$]) is 12 (see details of calculation in (Verduyn *et al.*, 1990)). This means that the theoretical intracellular concentration of acetate is 12 times higher than the extracellular concentration at equilibrium.

Pampulha and Loureiro-Dias (2000) have investigated how batch cultivations of a respiratory deficient mutant of *S. cerevisiae* are influenced by acetic acid concentrations between 0-170 mM. Their results show that the maximum specific growth rate (μ_{max}) and the yield coefficient of biomass on glucose (Y_{sx}) decreased, while the specific glucose uptake rate increased with increasing concentration of acetic acid in the broth, indicating that ATP was being used by ATPase to maintain the intracellular pH by excretion of protons. Similar observations of reduced Y_{sx} in the presence of acetic acid were reported for chemostat cultures with a respiratory-positive strain by Postma *et al.* (1989a). The acetic acid was produced metabolically by the yeast at constant concentration of 0.6 - 2 mM. The drop in Y_{sx} was seen for dilution rates that were up to 25 % below the critical dilution rate (D_{crit}).

4.2 Effects of Benzoate

This section will provide an overview of investigations in the open literature on the effect of the weak acid, benzoic acid, on chemostat experiments with *S. cerevisiae*.

Verduyn *et al.* (1992) and coworkers investigated the general effects of weak acids on growth by using non-metabolizable benzoic acid. Glucose-limited chemostat cultivations were conducted and the transport of benzoate across the plasma membrane could only occur by passive diffusion of the undissociated acid. By increasing the concentration of benzoate in the feed from 0 to 10 mM, maintaining a constant dilution rate of 0.10 hr^{-1}, the yield of biomass on glucose dropped from $Y_{sx} = 51$ w% to $Y_{sx} = 15$ w% without formation of ethanol. The calculated glycolytic flux increased from 1.1 to 3.9 mmole/g/hr, while the specific oxygen consumption rate increased from 2.5 to 19.5 mmole/g/hr. The results were similar for experiments conducted under different concentrations of glucose in the feed, 56 and 111 mM (10 and 20 g/L). At benzoate concentrations above 10 mM, ethanol formation sets in with a constant specific oxygen uptake at 13 mmole/g/hr. These results indicate that in the presence of a weak acid additional energy is needed to maintain the intracellular pH. This energy is provided by increased respiration.

Using the addition of benzoate to manipulate the activity of the oxidative phosphorylation, the highest attainable specific oxygen uptake rate qO_2^{max} (19.5 mmole/g/hr at 0.10 hr^{-1}) depends on the dilution rate. Verduyn and coworkers demonstrated that at lower dilution rates[1] (below 0.20 L/L/hr), qO_2^{max} was approximately 20 mmole/g/hr, while above 0.20 L/L/hr the value dropped from 15 to 13 mmole/g/hr as the dilution rate was increased from 0.25 to 0.38 L/L/hr. The authors conclude that it was the specific glycolytic flux (qC) that determined the qO_2^{max} obtained. As qO_2 equals qO_2^{max} the dilution rate where ethanol formation occurs has been reached.

Schulze (1995) investigated the effects of adding benzoate (0-8.5 mM) to glucose-limited anaerobic chemostat cultivations of *S. cerevisiae* with a feed concentration of 25 g/L glucose (139 mM). It was observed that the protein content of the cells increased from 46 to just below 60 w% and that the carbohydrates (glycogen and trehalose) decreased from 40 to 23 w%. At the same time the specific glycolytic flux and the specific rate of ethanol formation increased linearly with increasing benzoate concentration, which was also the case for the specific rate of ATP production. The uncoupling effect of benzoate was calculated to be 4.2 mmole ATP/g/hr/mM benzoate.

4.3 Effects of Ethanol

On a respiratory deficient mutant of *S. cerevisiae* Pampulha and Loureiro (1989) found that there were combined effects on the fermentation rate from ethanol and acetic acid present in the culture broth. The added effect in the presence of ethanol is explained by Casal *et al.* (1998) as a result of ethanol increasing the permeability of the plasma membrane for undissociated organic acids during growth on glucose.

[1] For this strain in the absence of benzoate, ethanol formation is observed at dilution rates above 0.39 L/L/hr. Glucose concentration in the feed was 83 mM (15 g/L) (Postma *et al.*, 1989b)

4.4 Dynamic Experiments

This section will provide an overview of dynamic experiments reported in the open literature with *S. cerevisiae* in presence of weak organic acids.

Verduyn *et al.* (1992) investigated the effects of pulse wise injection of sodium benzoate to obtain a concentration of 10 mM benzoate in a glucose-limited chemostat cultivation. The effects were quite surprising, since the investigations in chemostat cultivations with constant benzoate levels in the feed had indicated that an increased specific oxygen uptake rate, qO_2, could be expected along with a decrease of the biomass concentration. The results showed that biomass synthesis came to a full stop, while glucose and acetate accumulated in the broth for the first 10-15 min after the injection of sodium benzoate. At this point biomass synthesis slowly started again, accompanied by ethanol production at the expense of glucose and acetate. Surprisingly the qO_2 decreased for the first 2-4 min after the injection to 50 % of the level prior to the injection. From 5-20 min the qO_2 increased to a level of 300 % of the initial level. Accompanying these observations were measurements of intracellular ATP that increased to 200 % of the initial value after 7 min. and then dropped down close to the initial value after 10-15 min *i.e.* as biosynthesis and ethanol formation sets in.

Several observations are surprising in this pulse experiment. From the steady state chemostat experiments it was expected that the ATP level would decrease, since ATPase-facilitated excretion of protons, accompanying the passive inward diffusion of undissociated benzoic acid would be required to maintain a constant intracellular pH. This would then be followed by an increase of qO_2 to reestablish the ATP balance.

This was not observed. Firstly the sudden exposure to benzoate led to an increase in the intracellular ATP combined with a decrease in both qO_2 and glucose uptake. The importance of ATP as a regulator of the glycolytic flux has been discussed by Larsson *et al.* (2000), concluding that elevation of the intracellular ATP resulted in a lowering of the glycolytic flux, where the flux control seemed to be distributed along the glycolysis as a whole rather than at a single enzyme serving as a bottleneck. The glycolysis was far from stopped in the studies of Verduyn and coworkers, since both pyruvate and acetate were seen to accumulate during the first 10 min following the pulse. The reason for the rise in ATP could be the sudden halt of biosynthesis and its accompanying energy demanding processes leading to accumulation of ATP.

Secondly the rapid rise in acetate concentration without significant accumulation of ethanol was surprising. The high acetate concentrations (up to 23 mM) might have intensified the effects of the benzoate pulse, since acetate is also a weak acid and therefore the transport of acetate across the plasma membrane will influence the ATP balance, as mentioned earlier. This is an interesting observation opening up for an alternative explanation for the dynamic behavior, since it could be that the metabolic response to the pulse triggers a large production of acetate (up to 23 mM), which then is responsible for the observed fluctuations rather than the benzoic acid (only 10 mM injected). This hypothesis would not be out of line with the steady state observations discussed above; however no discussion on this subject has been found in the literature.

Pons *et al.* (1986) investigated the influence of metabolically produced acetate

during batch cultivation on glucose subsequently followed by growth on ethanol. At low initial glucose concentration (19 mM or 3.5 g/L), glucose uptake for biosynthesis was accompanied by ethanol and acetate production. As glucose was completely exhausted, continued biomass formation was based on ethanol and acetate. As glucose became depleted, acetate produced during glucose consumption was initially consumed, followed by consumption of ethanol. Just before the ethanol was exhausted a short period of acetate formation was observed, and this acetate was then consumed following the depletion of ethanol. The acetate concentration was below 3 mM throughout the batch.

The behavior changed when a high initial concentration of glucose (244 mM or 44 g/L) was used. During the first part of growth on glucose the pattern reported at low initial glucose concentration was repeated. As the acetate concentration exceeded 8 mM (0.48 g/L) a strong reduction of the growth rate was observed, whereas glucose continued to be consumed, but it was primarily converted to ethanol. Studies of fed-batch cultivations indicated that biomass yields on glucose and ethanol were already affected at acetate levels of 3 mM (0.18 g/L).

Lei *et al.* (2004) investigated the effects of pulsing 67 mM (4 g/L) of acetic acid to a chemostat cultivation running at 0.24 hr^{-1} with 167 mM (30 g/L) glucose in the inlet for the strain CEN.PK113-7D. Immediately following the pulse, the specific oxygen uptake rate qO_2 dropped to 30 % of its value prior to the acetate pulse. At the same time a large drop in pH was also observed (pH 5.0 to pH 3.3), however restored within 10 min. Lei concluded that it was not possible to decide if the drop in qO_2 was due to the change in pH or a direct effect from increased levels of acetate. Comparing with the observations of Verduyn *et al.* (1992) discussed above, who do not mention a drop in pH, a similar drop in qO_2 was also seen when adding benzoate. Lei reported that acetate was continuously produced up to 1.5 hours after the pulse, after which acetate was taken up by the cells and consumed.

Herwig and von Stockar (2002) built a small metabolic flux model to elucidate the distribution of intracellular fluxes in transients following a step change in the feed rate of glucose in a chemostat culture of *S. cerevisiae*. They reported similar, however slightly different, production of metabolites for the strains CEN.PK113-7D and ENY.WA-1C as a consequence of a step change in the feed rate. Whereas the results from the first strain were similar to those reported and discussed by Lei *et al.* (2004), the second strain showed a surprising response to a shift up from a dilution rate $D_0 = 0.074$ L/L/hr to $D_1 = 0.204$ L/L/hr, where the critical dilution rate for this strain was reported to be $D_{crit} = 0.19$ L/L/hr. Immediately following the step change, acetate is observed in the broth up to 17 mM (1 g/L) and is continuously being produced for the next 8 hours; a steady state level at 2 mM is reached after 20 hours. A small amount of ethanol is formed immediately after the step change (2 mM), whereas significant ethanol production is not observed until 2 hours after the step change. The ethanol concentration rapidly increases to almost 65 mM (3 g/L) and settles at 57 mM after 20 hours. Herwig and von Stockar explain their observations by a transient saturation of the oxidative catabolism, shunting the glycolytic flux into the fermentative metabolism. No saturation of the oxidative phosphorylation was seen and therefore it was not necessary to obtain extra capacity for re-oxidation of NADH by production of ethanol, but rather to direct the fermentative flux towards acetic acid and a production of NADH. Also a

significant rise in the anabolic flux was seen after the step change, increasing more rapidly than the glycolytic flux. The oxidative phosphorylation capacity obtained its maximum after 2 hours followed by a decrease accompanied by formation of ethanol. The authors concluded that the question of how the metabolic regulations were controlled remains unsolved.

4.5 Discussion and Summary

From the above results it has become clear that the effects of weak acids on the growth of *S. cerevisiae* are very complex. Most of the studies presented above have only addressed the effects of adding weak acids to the abiotic phase *i.e.* either directly into the broth or as a part of the feed. The studies by Verduyn *et al.* (1992) on the transient effects of adding a pulse of benzoate to a chemostat culture revealed that the dynamics and effects resulting from this type of experiment were quite different from the observations made in a number of chemostat cultures with benzoate added in the feed. Similar observations were reported by Lei *et al.* (2004) adding acetic acid. Herwig and von Stockar (2002) described how acetate was formed following a step change in the dilution rate to a value above D_{crit}, whereas ethanol was not produced in significant amounts until 2 hours after the step change. These investigations were performed on a strain of *S. cerevisiae* with a relatively low D_{crit} = 0.19 L/L/hr. No systematic investigation on the effects of metabolically produced acetate in high density (> 50 gDW/L) glucose-limited chemostat cultures has been found.

From the literature review it can be concluded that a quantification of the amount of acetate produced is essential, when attempting to construct mathematical models. This is important since the presence of this metabolite is linked to the intrinsic flux distribution, but can also have a severe effect on the growth energetics when appearing in the abiotic phase. Passive diffusion of the undissociated acid into the cell and subsequent active transport of the anion into the abiotic phase, requires use of metabolic energy therefore decreasing the metabolic energy available for growth. In high-density cultivations this latter effect would be expected to become very important since minor process upsets can lead to short term production of acetate and accumulation of significant amounts of acetate in the culture broth. If corrective actions as a response to the minor process upset do not take the risk of acetate accumulation into account (often the acetate concentration is not measured), the process can enter into an undesired spiral of decreasing biomass production and increasing acetate production, in turn leading to another and more severe process upset.

5

Genetics

The industrial strains of *Saccharomyces cerevisiae* studied in this work have been genetically engineered in order to produce an insulin precursor that can be converted into human insulin. The genetic modifications can potentially have an influence on the growth dynamics of the recombinant strain and are therefore highly relevant to present and discuss. The purpose of this chapter is to provide an overview of issues related to the genetic modifications that are needed in order to construct and optimize a genetically modified organism that can efficiently produce and express the desired recombinant protein in its correct form. The genetics of the host strain of *S. cerevisiae* in which the modifications have been introduced are also highly relevant to discuss in relation to growth and expression of a recombinant protein, however in the present work this will not be presented and discussed.

Reasons for choosing yeast and more specifically *S. cerevisiae* as a host organism for expression of heterologeous genes are many and plentiful. A non-exhaustive list of such reasons is provided by Walker (1998) and summarized below in table 5.1.

Table 5.1. Selected attributes of *S. cerevisiae* as a host organism for heterologeous protein production. Adapted from Walker (1998).

Historical	Yeast and yeast products have always had a general public acceptance *e.g.* beer and bakers yeast
	S. cerevisiae is non-pathological and generally regarded as safe (GRAS)
	S. cerevisiae is the most studied simple eucaryote in terms of biochemistry and genetics
Technological	Cultivation and downstream processes are well understood and developed
	A wide range of carbon sources for growth and cultivation exist
Genetical	Yeast can efficiently express heterologeous eucaryotic genes
	Contain *natural* plasmids (*e.g.* 2μ-DNA) but no virus to kill cells in large scale propagation
Molecular biological	Post-translational modifications (*e.g.* glycosylation) occurs together with proteolytic maturation and multimeric particle assembly
	Protein secretion is quite efficient and controllable using endogenous signal sequences
	Yeast RNA polymerase recognize many animal promoters

From an industrial and more specifically a pharmaceutical point of view, the status of *S. cerevisiae* being a GRAS organism is important as this simplifies the downstream processing, *i.e.* cumbersome and expensive process steps to ensure removal of pyrogens and virus can be avoided and regulatory authorities are already familiar with the microorganism due to its long and extended use in food and drug pro-

duction. Another highly attractive attribute of *S. cerevisiae* is its ability to utilize inexpensive, complex carbon sources *e.g.* enzymatically degraded polysaccharides. This latter attribute has a large positive impact on the process economy and furthermore reduces the dependency on suppliers.

The data used in this thesis have been produced by cultivation of two different, yet highly similar, industrial recombinant strains of *S. cerevisiae* both genetically modified for the production of insulin precursors; the two resulting insulin precursors are also slightly different. Of the two strains only one, strain (A), will be discussed in detail in this chapter. This strain has been used for experiments in pilot plant, and cultivations with this strain has provided the data for the work and modeling reported in chapters 7, 8 and 9. The other strain, strain (B), is used in chapter 6, which presents the modeling of cultivation data from production. The approach in the modeling of the data from production is much more data driven and therefore less mechanistic knowledge is needed for model construction adn interpretation. The only difference in the genetic modifications between the two strains is the engineering of the gene that leads to the excretion of the two different insulin precursors.

The chapter is structured as follows: First the full expression route of a recombinant protein will be described from transcription to secretion, followed by a presentation of the genetic modifications applied to the genome of the parent strain to construct the industrial production strain. The strategy behind the genetic modifications will then be presented with a more detailed description of the expression route of the recombinant protein until secretion into the extracellular medium. The chapter finishes with a discussion on the possible effects on the physiology of the recombinant strain as a result of the genetic modifications and the expected influence on the metabolism. The main sources of information for this chapter are the doctoral dissertation by Thomas Kjeldsen (Kjeldsen, 2000) and the Ph.d. thesis by Kirsten Væver Jochumsen (Jochumsen, 1995).

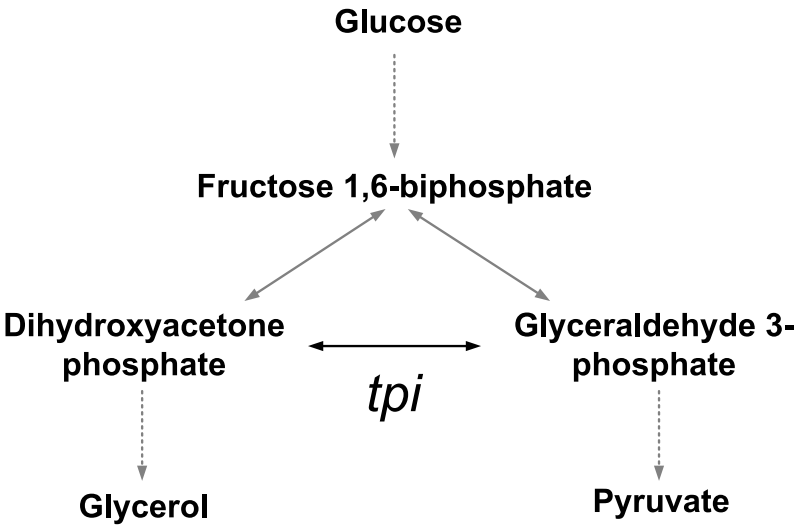


Figure 5.1. Schematic illustration of the position of the *tpi* catalyzed reaction between glyceraldehyde-3-phosphate and dihydroxyacetone-3-phosphate in the glycolysis.

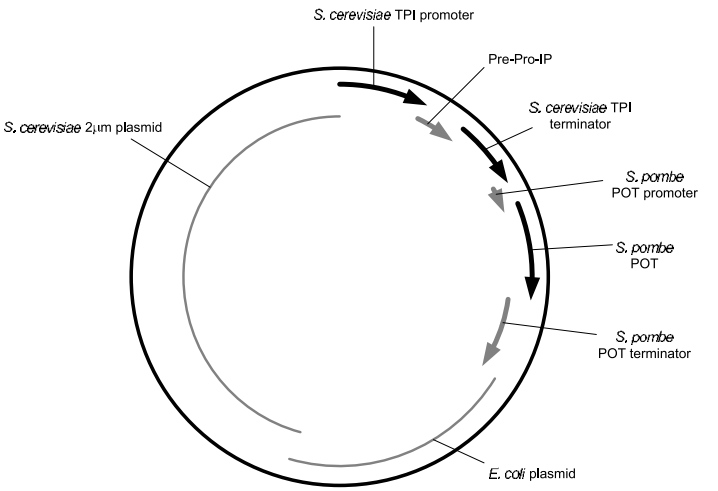


Figure 5.2. Schematic illustration of the genetic modifications to the 2μm plasmid of *S. cerevisiae* to facilitate production of the insulin precursor (IP). In the figure it can be seen that the expression plasmid consists of parts of the *S. cerevisiae* 2μm plasmid (top left hand side) as well as *E. coli* plasmid (bottom right hand side). In addition the expression cassette consisting of the *S. cerevisiae TPI1* promoter, the gene encoding the *Pre-Pro-IP* protein and the *S. cerevisiae TPI1* terminator can be seen in the upper right hand side, while the *S. pombe POT* promoter, *S. pombe POT* and *S. pombe POT* terminator jointly responsible for the expression of *tpi* can be on the right hand side (middle). The directions of the arrows indicate the reading direction during transcription of the genes. Adapted with modification from Kjeldsen *et al.* (2001).

5.1 The 2μm-DNA Plasmid

In order to express the desired insulin precursor (IP) in *S. cerevisiae* genetic modifications have been used. To facilitate the transport and correct folding of the insulin precursor protein through the secretory pathway and subsequent expression, two extensions to the insulin precursor gene sequence have been added for the *Pre-Pro*-insulin precursor gene[1]. The gene encoding the *Pre-Pro-IP* has been inserted into the *POT* plasmid, which is a hybrid shuttle plasmid based on the *S. cerevisiae* 2μm plasmid and the *E. coli* pBR322 plasmid (Kjeldsen *et al.* (2001); Kawasaki and Bell (1999)). The *POT* expression plasmid contains the *Schizosaccharomyces pompe* triose phosphate isomerase (*tpi*), encoded by the *POT* gene, the enzyme catalyzing the interconversion between glyceraldehyde-3-phosphate and dihydroxyacetone-3-phosphate in the glycolysis (see figure 5.1 and 3.4 in chapter 3). By introducing the *POT* 2μm plasmid into a host organism such as the strain MT663 (*MATa/MATα pep4-3/pep4-3 HIS4/his4 tpi1::LEU2/tpi1::LEU2 cir*$^+$), which is carrying a deletion in the native *TPI1* gene, a selection of the *POT* plasmid is obtained as only cells carrying this plasmid have the ability to grow on complex medium requiring glycolytic and therefore *tpi* activity to enzymatically degrade polysaccharides (Kjeldsen *et al.*, 2001). The expression plasmid is illustrated in figure 5.2.

It should be noted that in the strain MT663, the *PEP4* gene has also been deleted.

[1]This gene is different for the two strains used in this thesis

The *PEP4* gene product (proteinase A) is related to the activation of vacuolar proteases, leading to an attenuation of the proteolytic system in the vacuoles (Diers *et al.*, 1991).

As the *S. pombe POT* gene is only weakly expressed in *S. cerevisiae*, a high copy number of the *POT* plasmid is required to ensure sufficient gene product for growth on glucose. In similar strains Egel-Mitani *et al.* (1988) and Jochumsen (1995) estimated plasmid copy numbers of approximately 20 per cell. By using the strong constitutive native *S. cerevisiae TPI1* gene promoter in the expression casette of the *Pre-Pro-IP* gene, in combination with the high copy number of the *POT* plasmid a high productivity of the *Pre-Pro-IP* gene product is ensured.

5.2 The Secretory Pathway

Figure 5.3 presents a schematic illustration of the secretory pathway for an insulin precursor expressed in *S. cerevisiae*. The figure highlights some of the challenges that need to be addressed when attempting to genetically modify microorganisms to produce recombinant proteins. The challenge is not limited to the introduction of the gene encoding for the desired protein in a suitable vector *e.g.* 2μm-DNA plasmid. Stabilization of the protein and correct transport through the organelles are a prerequisite for obtaining an intermediate product that can be converted into the correct and active form of the desired recombinant protein. Further adding to the complexity are the optimization issues of utilizing the capabilities of the *S. cerevisiae* expression system to form disulfide bridges, avoid unwanted glycosylations and excrete the insulin precursor in a configuration that simplifies the operations needed to obtain the desired final form in a purity suitable for pharmaceutical applications.

As seen in figure 5.3 the polypeptide sequence that enters into the endoplasmic reticulum contains both a *Pre-* and *Pro*-extension to the insulin precursor as well as a spacer peptide mentioned in the caption of figure 5.3. It is the engineering of these peptides that have played a major role in improving the transport through organelles and concurrent folding of the insulin precursor as well as facilitating an easy downstream processing to obtain the desired product (Kjeldsen, 2000).

A newly formed polypeptide chain needs to be directed through the secretory pathway in order to ensure correct folding, translocation and excretion of the recombinant protein. The *Pre*-peptide of the *Pre-Pro-IP* gene product is a signal peptide such as the *Yap3* (yeast aspartyl protease 3) endoprotease signal peptide and the function of *Yap3* is to guide the nascent polypeptide chain into the endoplasmic reticulum (ER) for post-translational translocation. In the ER the signal peptide is removed by a signal peptidase, while disulfide bonds and tertiary structure of the *Pro-IP* protein is formed, facilitated by the *Pro*-peptide of the *Pro-IP* protein. A large amount of research has been carried out on how to synthesize a *Pro*-peptide that results in high yield of the correctly folded insulin precursor (Kjeldsen, 2000). Within the ER a *quality control system* consisting of chaperones and enzymes ensures that the protein is correctly folded. By a vesicular system, mediated transport of the protein to the Golgi apparatus occurs.

In the Golgi apparatus the *Kex2* endoprotease removes the *Pro*-peptide, while vesicles transport the *IP* protein to the cellular plasma membrane for excretion into

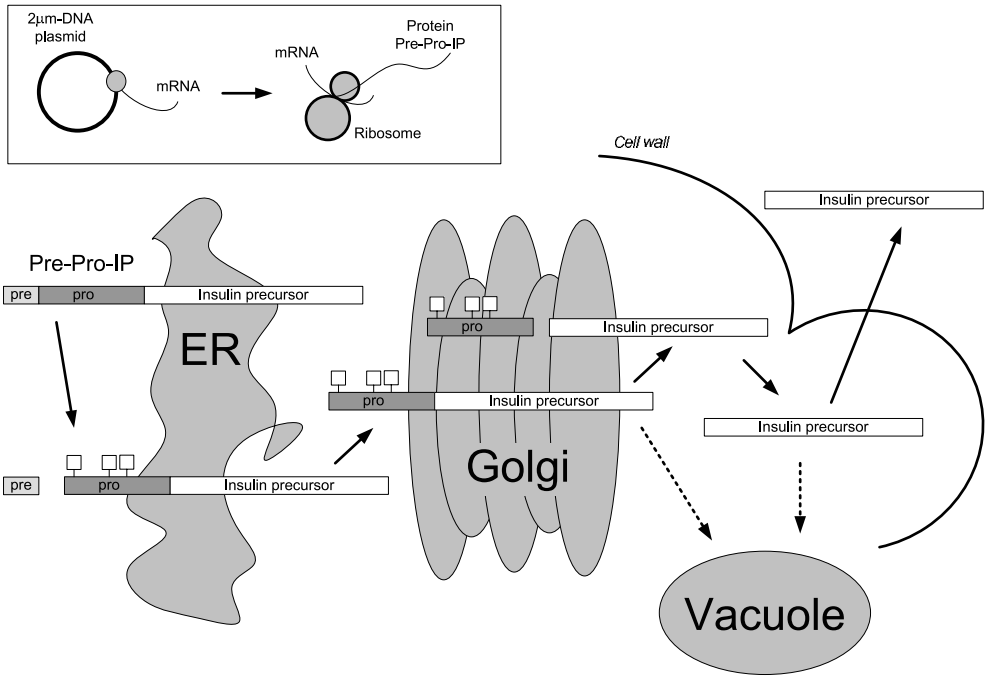


Figure 5.3. The secretory pathway of insulin precursor expressed in *S. cerevisiae*. The insert shows how the gene encoding for the *Pre-Pro*-insulin precursor is positioned on the 2 μm-DNA plasmid located in the nucleoplasm of the yeast cell. The DNA sequence is transcribed into a mRNA sequence, which in a cytoplasmic ribosome is translated into a polypeptide forming the *Pre-Pro*-Insulin precursor protein. In the endoplasmic reticulum (ER), signal peptidase removes the *Pre*-part, while primary oligosaccharides (illustrated as square boxes) are attached to the of the remaining protein. A vecsicle system mediates the transport to the Golgi apparatus where the *Pro*-part of the *Pro*-insulin precursor protein is removed by *kex2* endoprotease. The insulin precursor is then transported to the cell wall for excretion or to a vacuole. Not shown on the figure is a small spacer peptide positioned at the N-terminal of the insulin precursor. Following purification of the insulin precursor, an enzymatic conversion of the insulin precursor leads to the desired human insulin protein. The figure has been adapted with modifications from Kjeldsen (2000).

the extracellular medium.

Kjeldsen *et al.* (2001) suggested a correlation between the *in vitro* folding stability and the cultivation yield and suggested that this reflects a selective adaptation of the plasmid copy number. This selective adaptation is partially a consequence of the constitutively expressed insulin precursor's folding properties and partially the need for complementation of the deleted *TPI1* by *POT*. In other words, the plasmid copy number is adjusted based on the folding stability and the need for complementation of the deleted *TPI1* by *POT*, and therefore cultivation yield can not be increased solely by increasing the plasmid copy number.

5.2.1 Expression Rate

Kjeldsen *et al.* (1999) performed so-called pulse chase experiments on the *Pre-Pro-IP*

protein using metabolic labelling with [^{35}S]cysteine for 2.5 min whereafter unlabelled cysteine medium was used. Cells for the pulse chase experiment were grown in a batch cultivation and exposed to the pulse during exponential growth. Samples were taken for isolation of the secreted IP. The first labelled IP appeared in the supernatant 2-4 min after the pulse, and the majority of the secreted IP appeared in the supernatant within 15 min after the pulse, leading to a $t_{1/2}$ in the range of 5-10 min.

Analysis of the intracellular retention of the IP showed that 30 min after a metabolic labelling for 2.5 min, approximately 30% of the labelled IP was still present as processed intracellular IP and was not secreted into the culture broth. This indicated that two different intracellular routes for the IP was present in the late secretory pathway, and that secretion may reflect saturation of a sorting mechanism due to over-expression of the IP or that secretion occurs in competition with intracellular retention. It was found that the retained IP was correctly processed indicating that the retention occurred after the cleaving by the *Kex2* endoprotease positioned in the late Golgi apparatus. As proteins can be routed to a vacuole from the Golgi apparatus, it seemed likely that the IP had been sorted to a vacuole. Normally a vacuole is a site for proteolytic degradation of proteins, but this does only occur to some extend in this stain, since the *PEP4* gene has been deleted, leading to attenuation of the proteolytic system in the vacuoles (Diers *et al.*, 1991).

5.3 Effects on Physiology

Genetic manipulations in a recombinant strain can have a number of effects on the physiology of the recombinant strain. Jochumsen (1995) investigated expression of the gene product of *PEP4* (proteinase A) using different promoters in two recombinant strains similar to the one used in this study, also based on an *POT* expression plasmid. Furthermore Jochumsen conducted experiments with a dummy strain similar to the other investigated strains, containing the same *POT* expression plasmid however lacking the *PEP4* gene. The observed critical dilution rates (D_{crit}) and maximal growth rates (μ_{max}) of the two types of promoters used by Jochumsen (1995) as well as a dummy strain are listed in table 5.2 together with results reported by Postma *et al.* (1989b) for a non-recombinant strain of *S. cerevisiae*.

Table 5.2. Observed critical dilution rates (D_{crit}) and maximal growth rates (μ_{max}) for two recombinant strains with different promotors and a dummy strain of *S. cerevisiae* (Jochumsen, 1995) as well as the values reported for a non-recombinant strain of *S. cerevisiae* (Postma *et al.*, 1989b).

Strain	Precursor promoter	Type	D_{crit} [L/L/hr]	μ_{max} [hr^{-1}]
JG176	*S. cerevisiae TPI1*	Constitutive	0.22	0.27
JG180	*S. cerevisiae PEP4*	Inductive	0.16	0.29
MT888	Dummy strain	-	0.29	0.48
CBS8066	None (non-recombinant strain)	-	0.38	0.48

Jochumsen concludes that both presence of plasmid DNA and production of pro-

teinase A affect the growth rate and that the type of promoter (constitutive vs. inductive) has an influence on the growth energetics.

5.3.1 Regulation of the *TPI1* Promoter During Oxido-reductive Growth

Kjeldsen *et al.* (2001) mentions that the *TPI1* promoter might not be strictly constitutive and may exhibit regulation in response to the glucose concentration, resulting in a sub-optimal gene expression pattern during fermentation. Jochumsen (1995) found that presence of ethanol in the culture broth during chemostat cultivations on glucose *i.e.* during oxido-reductive growth, did not seem to have an effect on the specific productivity of *pep4* when the native *TPI1* promoter was used in a *POT* expression plasmid similar to the one used by Kjeldsen *et al.* (2001)[2]. Jochumsen (1995) also noted that high glucose concentrations coincided with the highest levels of specific productivity of the product, indicating that the production and expression of the product was not subjected to substrate inhibition. Although no direct measurement of the *tpi* was reported, the influence on the expression of *pep4* did not seem to be affected by the presence of neither ethanol nor glucose. van Hoek *et al.* (2000) studied the levels of glycolytic enzymes in two different strains of *S. cerevisiae* (CEN.PK113-7D and DS28911, an industrial recombinant strain for bakers yeast production[3]). They observed two different activity profiles of *tpi* as a function of the dilution rate in chemostat cultivations of the two strains. The two strains had almost identical critical dilution rates (D_{crit}). The activity profile for CEN.PK113-7D was constant for dilution rates up to D_{crit} and then increased almost proportional with the dilution rate up to the highest dilution rate studied; ethanol was present above D_{crit}. For DS28911 the activity profile of *tpi* was constantly decreasing from the lowest to the highest dilution rate studied. At the lowest dilution rate the *tpi* activity was 2.5 times higher in DS28911 compared to CEN.PK113-7D. At the critical dilution rate the *tpi* activity was 1.5 times higher in DS28911 compared to CEN.PK113-7D, while at the highest dilution rate studied the *tpi* activity was 3 times lower in DS28911 compared to CEN.PK113-7D. Again ethanol was present above D_{crit}. From the data it can be concluded that the activity of *tpi* does not seem to exhibit regulation in response to the ethanol concentration.

From the observations cited above it does not seem that the *TPI1* promoter is regulated in response to the glucose concentration. The sub-optimal gene expression pattern during fermentation mentioned by Kjeldsen *et al.* (2001), is therefore probably due to other effects.

5.3.2 Plasmid Copy Number

An explanation for the difference in critical dilution rates of the two recombinant strains of Jochumsen (1995) can also be attributed to an observed difference in plasmid copy numbers. In the strain containing *S. cerevisiae TPI1* promoter and

[2]The difference between the strains investigated by Jochumsen (1995) and Kjeldsen *et al.* (2001) is the recombinant gene product, *pep4* and an insulin precursor respectively. The size of the insulin precursor is approximately 6 kDa, while the size of *pep4* is approximately 42 kDa.

[3]No information on the genetic modifications are available.

terminator (D_{crit} = 0.22 hr^{-1}) a copy number of 52 is reported at D = 0.260 hr^{-1}, while the strain containing a *S. cerevisiae PEP4* promoter and terminator (D_{crit} = 0.16 hr^{-1}) a copy number of 16 is reported at D = 0.188 hr^{-1}. The combination of a higher copy number and a strong selection of *POT* expression plasmids, suggests that the *tpi* activity would be higher in the strain containing a *S. cerevisiae TPI1* promoter and terminator leading to an increase in the D_{crit}, but interestingly not in μ_{max}.

Attempting to explain the differences in D_{crit} observed by Jochumsen solely by the difference in plasmid copy numbers does not seem to be correct as Jochumsen also notes. The reason for this is that the observed increase in D_{crit} by an increase in *tpi* activity would indicate that the rate limiting step in the glycolysis in the strain containing *S. cerevisiae PEP4* should be a limiting capacity of *tpi*. A consequence of this would be an overflow metabolism at dihydroxyacetone-3-phosphate leading to production of glycerol rather than ethanol as was observed and reported by Jochumsen (1995). An explanation for the observed differences in D_{crit} is more likely linked to the energetics related to the expression of the recombinant proteins than linked to the plasmid copy number and therefore the expression of *tpi*.

The difference in the plasmid copy numbers could be explained by an observation by Egel-Mitani *et al.* (1988). Attempting to increase the expression of insulin in *S. cerevisiae*, also using a *POT* expression plasmid similar to the one used by Jochumsen, they investigated several approaches to expressing two insulin precursor expression units per plasmid. The result was that the average plasmid copy number in the strains containing two insulin precursor units per plasmid was half (9.4) the plasmid copy number of similar strains containing a single insulin precursor unit per plasmid (18.8).

These findings indicated that the plasmid copy number is regulated in order to keep the total number of insulin precursor units constant in a recombinant strain, suggesting a maximal level at which the yeast system can accommodate expression of the recombinant gene product. This was also mentioned by Kjeldsen *et al.* (2001).

5.3.3 Specific Productivity

Jochumsen (1995) reports further that the specific productivity of proteinase A in chemostat cultivations of the two recombinant strains are very different. The strain containing a *S. cerevisiae PEP4* inductive promoter and terminator had a local maximum of the specific productivity at approximately 0.70 mg/g/hr just around and below the critical dilution rate of D_{crit} = 0.16 L/L/hr resulting in a maximum volumetric productivity of 5.5 mg/L/hr in this range. At dilution rates D > 0.18 L/L/hr the specific productivity drops with a factor of approximately 5 to 0.15 mg/g/hr at D= 0.18 L/L/hr, and slowly increases with further increasing dilution rate to a level of 0.9 mg/g/hr at the maximum growth rate of μ_{max} = 0.29 hr^{-1} resulting in a volumetric productivity of 1.4 mg/L/hr.

For the strain containing a *S. cerevisiae TPI1* constitutive promoter and terminator no local maximum is found for the specific productivity at the critical dilution rate of D_{crit} = 0.22 hr $^{-1}$. Instead the specific productivity was slowly increasing with values of 0.1 mg/g/hr at D = 0.14 L/L/hr and 0.4 mg/g/hr at D = 0.24 L/L/hr, reaching a maximum specific productivity of 0.65 mg/g/hr at the maximum growth

rate of $\mu_{max} = 0.27$ hr^{-1}. The maximum volumetric productivity of 2.1 mg/L/hr is obtained at $D = 0.18$ L/L/hr.

From these numbers it is clearly seen that the dilution rate at which the maximum specific productivity is obtained does not necessarily coincide with the dilution rate at which the maximum volumetric productivity occurs. Comparing the results with the observations of plasmid copy numbers suggest that a higher plasmid copy number does ensure a higher specific productivity, although the data material behind this conclusion is rather small. Table 5.3 summarizes this data.

Table 5.3. Comparision of plasmid copy number (PCN) and specific productivity (q_p) of the two strains investigated by Jochumsen (1995). Ratio shows the specific productivity divided by the plasmid copy number.

Precursor promoter	D [L/L/hr]	PCN	q_p [mg/g/hr]	Ratio $1000 \cdot q_p$/PCN
S. cerevisiae TPI1	0.18	16	0.15	9.4
S. cerevisiae PEP4	0.26	52	0.65	12.5

5.3.4 Glycerol Production in a *tpi* Deficient Mutant

Compagno *et al.* (1996) describes how anaerobic batch cultivations with a *S. cerevisiae* strain with a deletion in the *TPI1* gene product, resulting in a Δtpi strain, yielded primarily glycerol as the major by-product during growth on glucose. Cells collected during the exponential growth phase on a rich medium containing 20 g/L ethanol + 1 g/L glucose media were resuspended at 100 g/L in a solution containing 50 g/L glucose as well as 7 g/L NaH_2PO_4, pH 6. The glucose consumption rate was 25% slower than was the case for the wild-type strain (wt), and the glucose was exhausted after 22 hours (wt: 4 hours). Only small amounts of ethanol were produced by the Δtpi strain, and the maximum value at 2 g/L was obtained after just 3 hours (wt: 12 g/L) where furthermore a glycerol concentration of 6 g/L was observed (wt: 0 g/L). As the glucose was exhausted after 22 hours, the glycerol concentration had risen to 17 g/L (wt: 1 g/L) and also an acetate concentration of 1.8 g/L (wt: 0.5 g/L) was reported. A further experiment was conducted in a rich medium containing a yeast nitrogen base (Difco), 100 g/L glucose and 50 g/L NaH_2PO_4 and pH 6. After 50 hours the glucose was not fully exhausted and the culture broth contained 20 g/L glucose, 25 g/L glycerol and 7 g/L acetate.

The results above are interesting, since they indicate that an inefficiency in the expression of *tpi* could lead to production of glycerol with simultaneous reoxidation of NADH. The observed production of acetate rather than ethanol could be explained by an overflow into the fermentative metabolism which then yields acetate since the oxidative state of the cells are maintained by the reoxidation of NADH by the glycerol production.

5.4 Discussion

Jochumsen (1995) concluded that the dilution rate at which the maximum specific productivity was obtained did not necessarily coincide with the dilution rate at which

the maximum volumetric productivity appeared. This was explained by the presence of plasmid RNA and production of a recombinant protein affecting the growth rate, since metabolic energy is needed for maintenance of the plasmid DNA and expression of the recombinant protein. This was concluded since a comparison of the observed critical dilution rates and maximal growth rates for two recombinant strains were significantly different from a a dummy strain of *S. cerevisiae*, with the only difference being the presence of a recombinant gene product and its promoter/terminator. Furthermore Jochumsen (1995) showed that the type of promoter (constitutive vs. inductive) also had an influence on the growth energetics due to slight differences in critical dilution rates and maximal growth rates for two recombinant strains.

From analysis of the data provided in Jochumsen (1995), it was also argued that weak indications suggested that a higher plasmid copy number did lead to a higher specific productivity. Combining this with the findings of Egel-Mitani *et al.* (1988), that the plasmid copy number is regulated in order to keep a maximal expression level of a recombinant gene product, suggests that the specific productivity is a function of the amount of metabolic energy that the cell can provide in addition to the energy needed for maintenance and growth. It could be speculated that low concentrations of key amino acids for anabolic reactions could also be responsible for the regulation of the plasmid copy number, however no indications of this have been reported. In the future it would be interesting to carry out further investigations to elucidate this topic.

From these speculations an interesting dilemma appears. Could the down regulation of the plasmid copy number during conditions of shortage in metabolic energy for expression of the insulin precursor, influence the expression of *tpi e.g.* lowering the activity of *tpi*, whereby a bottleneck at *tpi* appears? The results of Jochumsen (1995) did not provide an answer to this question, since high specific productivity of *pep4* were seen for both strains investigated at the maximal dilution rate.

5.5 Conclusion

Only a couple of the many factors influencing construction of an efficient expression system for production of recombinant proteins have been presented here. Using *S. cerevisiae* as an expression host for recombinant proteins, the cultivation yield is rather low with a maximum of 1-5% of the total protein synthesis (Kjeldsen *et al.*, 2001). Still the replication and transcription of a multicopy plasmid require large amounts of energy, which in turn can have significant effects on the metabolism of the yeast, with large differences in critical dilution rates and maximal growth rates between wild type and recombinant strains as presented above.

It is concluded that the plasmid copy number is regulated in order to keep a maximal expression level of the recombinant gene product. This maximum expression level, and therefore the plasmid copy number, is dependent on the metabolic energy being produced and the energy requirements for maintenance and growth. Furthermore a weak indication suggested that a higher plasmid copy number leads to a higher specific productivity.

The general conclusion of this chapter is, that despite the large efforts to investigate and elucidate the consequences of genetic modifications in microorganisms, it is very

difficult to understand the full impact of the modifications regarding the effects on dynamics of growth. This is a significant dilemma, as the drive towards improvement and optimization of process performance will attempt to push the capacity of the recombinant microorganism to its limits. This can lead to surprising results as will be shown in later chapters, which were not anticipated during the design of the genetic modifications.

Part II

Main Thesis Chapters

6

Use of Soft Sensors for Monitoring and Prediction in Cultivation Processes

The development of modelbased process soft sensors for monitoring of biomass and product concentration in fed-batch and continuous yeast cultivations is presented, followed by sensor validation using data from industrial cultivations. Alternatively, using multiway projection to latent structures (MPLS) algorithm, a model for prediction of one-step ahead and end point product concentrations is developed and demonstrated on industrial process data. The one-step ahead MPLS-predictor is compared to the model based product concentration soft sensor. Both sensors show good performance in estimating the product concentration.

To improve monitoring and control of industrial cultivation processes it is desirable to include interpreted information of dynamic responses of relevant biological and chemical species to changes in process conditions.

Fulfilling this objective is however not trivial, since measurements of relevant species are often difficult to achieve and often impossible to obtain at the desired rate. An alternative approach to the direct measurement of species is the development of process soft sensors based on mathematical models correlating measurable variables to the desired variables. The purpose of this work is to investigate the development of two different types of soft sensors, one based upon first principles engineering modelling and another upon Chemometrics, more specifically Projection to Latent Structures regression. First the process and the basis for the developed soft sensors are presented in sections 6.1 and 6.2. The performance of the knowledge and the data based sensors are compared in section 6.3 using data from an industrial yeast cultivation. Section 6.4 presents a discussion and conclusion of this study.

6.1 Introduction

The different soft sensors that will be presented below are developed in order to improve monitoring and control of an industrial cultivation process for the production of a recombinant strain of *Saccharomyces cerevisiae* genetically modified for the production of an insulin precursor. The process is conducted in industrial scale bioreactors and before the main cultivation is carried out the culture has been pregrown on agar in a Fernbach flask and in a seed bioreactor operated in batch mode on a complex medium with glucose as the primary carbon source. The main biore-

actor, containing similar complex medium as the seed bioreactor, is inoculated with the contents of the seed bioreactor. For the first period of time the cultivation is operated in batch mode. As the culture broth is depleted of carbon sources, feeding of two complex substrates at a fixed ratio is initiated, with glucose as the primary carbon source. Continuous operation is undertaken as a predetermined level is reached. The effluent is led to recovery and further downstream processing. The continuous operation is carried out for a predetermined length of time after which the process is terminated.

Offline samples are taken throughout the process in order to monitor the evolution of the cultivation. Both of the analytical methods used for determining the biomass and the insulin precursor concentrations require a long processing time. Often the results from laboratory analysis are only available to process operators or engineers with delays of more than 24 hours. This time delay limits the usefulness of the measurements as tools to monitor and control the cultivation, and the primary use of the measurements are limited to batch documentation and data analysis after the cultivation has been terminated.

The proposed soft sensors are constructed for slightly different scenarios. The soft sensor built upon a first principles engineering model (FPEM) can be used throughout fed-batch and continuous operation, whereas the MPLS based soft sensor is limited to the use in fed-batch and the initial period of the continuous operation. The reduced use of the MPLS based soft sensor is due to process specific operating procedures requiring cleaning of the effluent pipe every 24 hours[1]. For this reason comparison of the two soft sensors has been reduced to the operating time range of the MPLS based soft sensor.

6.2 Process Software Sensors

First principles engineering models (FPEM) can form the foundation for soft sensors. The models infer information of unmeasured entities by using available information from other measured entities. Different frameworks can be used for the model development. First soft sensors using FPEMs will be developed and investigated for the prediction of biomass and product (insulin precursor) concentration in a culture broth. Subsequently a chemometric model is used for developing a soft sensor for product estimation. Finally the two types of product concentration sensors are compared.

To evaluate the performance of the soft sensors the root mean square error of prediction is used:

$$RMSEP = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(\hat{y}_k - y_k\right)^2} \tag{6.1}$$

where K is the number of data points.

[1]The pipes are cleaned frequently in order to avoid fouling.

6.2.1 FPEM based Sensor for Biomass Concentration

During the last decade a number of research groups have contributed to the understanding of the mechanisms behind the proton balance in microbes during growth (Castrillo *et al.* (1995), Siano (1995), and Lei (2001)), leading to the observation that a relationship exists between the alkali addition (KOH/NaOH) and the biomass production rate. The inclusion of effects to the alkali addition rate due to a difference in pH between the substrates and the broth however has not been reported in the literature, where Castrillo *et al.* (1995) and Pham *et al.* (1999) conducted their experiments with similar pH in the substrate and medium, while Lei (2001) found that a pH difference of 0.35 only required an increase in the base flow of 1% and therefore could be neglected.

Lei (2001) demonstrated that it was possible to use a component mass balance on the proton production or consumption rate in a high performance laboratory setup to obtain a simple online estimation of the biomass concentration in batch, fed-batch and continuous cultivation of *Saccharomyces cerevisiae*.

A simplified illustration of the contributions to the proton balance in a bioreactor is shown in figure 6.1. A component mass balance for the proton concentration $[H^+]$ in the extracellular medium yields:

$$V\frac{d[H^+]}{dt} = F_s[H^+]_{s,in} - F_e[H^+]_{e,out} + F_{H^+,gen} - F_{NH_3} \tag{6.2}$$

where the dual role of NH_3 is i) to maintain a constant pH-level in the medium and ii) to act as the primary nitrogen source for biomass production.

The following assumption is used for simplification of the mass balance expression:

- Constant pH-level in the bioreactor

In the original work by Castrillo *et al.* (1995) the pH of the feed was adjusted to the pH of the medium. In the present it has been estimated that approximately

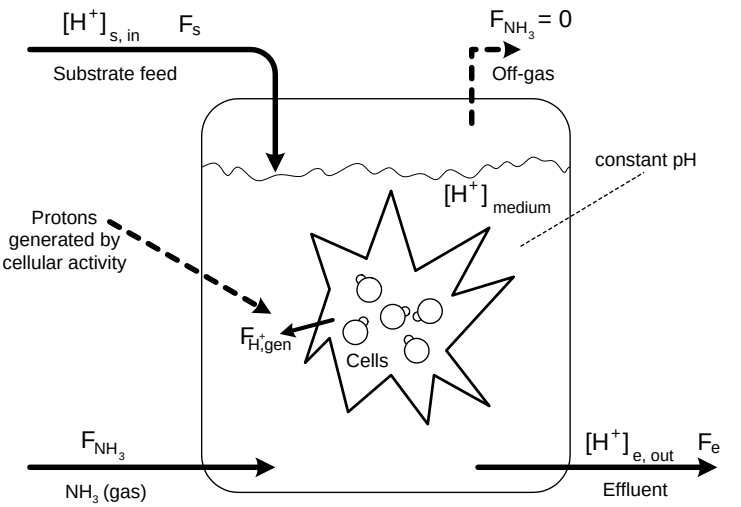


Figure 6.1. Simplified schematic illustration of flows and factors that influences the extracellular proton concentration balance in the culture broth.

10% of the added NH_3 is used for neutralizing the effect of a pH difference (of pH 2.3) between the culture broth and the complex substrate. Since the substrate and NH_3 flows are not always balanced, a continuous evaluation of the amount of NH_3 needed for neutralization has to be calculated. Hence the mole balance yields:

$$0 = F_{H^+,gen} - F_{NH_3} + F_s[H^+]_{s,in} - F_e[H^+]_{e,out} \tag{6.3}$$

By assuming that $F_s \approx F_e$ in continuous operation and $F_s \ll V$ in the fed batch phase, equation 6.3 can be simplified to:

$$0 = F_{H^+,gen} - F_{NH_3} + F_s q_{s,H^+} \tag{6.4}$$

where q_{s,H^+} is the number of proton equivalents needed for changing the pH of the substrate to the pH of the medium. The volumetric proton production rate can now be calculated as:

$$r_{H^+} = \frac{F_{H^+,gen}}{V} = \frac{F_{NH_3} - F_s q_{s,H^+}}{V} \tag{6.5}$$

Studies of the buffer capacity of the substrate indicate an experimental value of $q_{s,H^+} = 36$ mmole H^+eq/L [2].

The following assumptions have been made concerning possible sources contributing to the proton production rate from cellular activities during aerobic growth on a complex medium:

- Uptake of NH_4^+ as primary nitrogen source
- Negligible production or consumption of carboxylic acids
- Negligible consumption of amino acids from complex medium
- No acidification of the medium due to production of CO_2
- Balanced growth

During aerobic growth on glucose as substrate only negligible amounts of carboxylic acids are produced; CO_2 and biomass being the primary carbon-containing products formed. Contribution to the proton balance by the solution and dissociation of CO_2 (H_2CO_3) to carbonate can be disregarded when the pH-level is significantly below pH 7. Combined with the assumption that only negligible organic N-sources from the complex substrate are consumed during aerobic growth [3], the above assumptions leave the uptake of NH_4^+ as the sole contributor to the proton production rate and the only significant nitrogen source. Castrillo *et al.* (1995) observed a 1:1 ratio between proton production rate and the NH_4^+ uptake rate (using $(NH_4)_2SO_4$ as N-source) indicating that the biomass production rate is proportional to the proton production rate, when the nitrogen content of the biomass can be assumed constant during balanced growth.

[2] This value was determined by titration of the substrate by 1N NaOH.

[3] It has yet to be investigated if this assumption holds, since it is possible that a contribution to the overall N-balance is provided by organic N from amino acids in the yeast extract.

Based on the above comments and assumptions the volumetric biomass production rate, $r_x^{H^+}$ [unit: C-mole biomass/kg broth/hr], can be calculated from the volumetric NH_3 addition rate:

$$r_x^{H^+} = \frac{M_{DW} \cdot r_{H^+}}{Y_{xH}} \tag{6.6}$$

where M_{DW} is the formular weight of C-mole dry weight biomass and Y_{xH} is the yield coefficient of mole protons produced per C-mole biomass produced, and is equal to the molar content of nitrogen in biomass based on the overall growth stoichiometry:

$$CH_xO_yN_z + \mathrm{a}CO_2 + \mathrm{Y}_{xH}H^+ - \mathrm{b}CH_2O - \mathrm{c}NH_4^+ - \mathrm{d}O_2 = 0 \tag{6.7}$$

From the stoichiometric equation it can be seen that Y_{xH} is constant, since NH_4^+ is the only proton source and $z = c\ (= Y_{xH})$ since NH_4^+ is the only nitrogen source. Combination of the above expressions with a dynamic mass balance for biomass (x):

$$\frac{dx}{dt} = r_x^{H^+} - Dx \tag{6.8}$$

yields a simple biomass estimator:

$$x_{k+1} = x_k \cdot exp\left(\left(\frac{M_{DW} \cdot (F_{NH_3,k} - F_{s,k}q_{s,H^+})}{V_k \cdot Y_{xH} \cdot x_k} - D_k\right)(t_{k+1} - t_k)\right) \tag{6.9}$$

where D_k is the dilution rate at time point t_k, and the appropriate expression for $r_x^{H^+}$ in equation 6.6 is used.

The above model has been developed assuming ideal conditions in the bioreactor. Both for small and large scale cultivations with high cell densities this assumption is unlikely to be valid, for example due to changes to culture broth characteristics *e.g.* viscosity. In large scale bioreactors formation of zones and concentration gradients occur due to the physical distances between inlets of substrate streams and the extremes of the bioreactor *e.g.* walls, bottom or top. An example of this could be an inlet stream being positioned at the top of the bioreactor feeding glucose to the process and cells positioned at the bottom of the bioreactor meters away with several tons of high cell density culture broth inbetween.

To account for these variations the model has been modified as follows:

$$x_{k+1} = x_k \cdot exp\left(\left(\frac{\alpha \cdot (F_{NH_3,k} - F_{s,k}q_{s,H^+})}{V_k \cdot x_k} - D_k\right)(t_{k+1} - t_k)\right) \tag{6.10}$$

where $\alpha = f(t, M_{DW}, Y_{xH}, \text{vessel properties})$ is determined for the individual bioreactor. In this work it has been assumed that α is independent of time variations and that the contributions from physiology (M_{DW}, Y_{xH}) and vessel properties can be modelled as two constant values in a multiplicative model:

$$\alpha = \alpha_p \cdot \alpha_v \tag{6.11}$$

where α_p represents the physiological contribution and α_v represents the contribution from vessel properties *e.g.* vessel size, propeller configuration *etc.*

6.2.2 Parameter Estimation

The estimation of parameters for the two FPEM models in sections 6.2.1 and 6.2.3 is gathered in this section since the approaches are almost identical due to the similarities in the underlaying models; therefore only the estimation of α will be described in detail.

The contribution from the vessel properties (α_v) are first determined, since α_p is expected to be constant and independent of the vessel properties. The analysis takes its onset from equation 6.8. By assuming stationary conditions one obtains:

$$\frac{dx}{dt} = 0 = r_x^{H^+} - Dx \tag{6.12}$$

$$\Leftrightarrow Dx = r_x^{H^+} \tag{6.13}$$

$$Dx = \alpha_p \cdot \alpha_v \cdot r_{H^+} \tag{6.14}$$

6.2.2.1 Determining α_v

Plotting r_{H^+} versus $D \cdot x$ for a number of cultivations carried out in different vessels provides an illustration of the extent of the tank to tank variation. Figure 6.2 provides such an illustration, showing the data points of cultivations carried out in two different bioreactors of the same size. Bioreactor 1 has an average value of the normalized volumetric proton production rate of $1.56{\cdot}10^{-3}$moles/(kg hr), while the average value of bioreactor 2 is $1.90{\cdot}10^{-3}$moles/(kg hr) at comparable levels of normalized volumetric biomass production rate, approximatly 5.3 g/(kg hr).

For all the bioreactors investigated the picture is the same as illustrated in figure 6.2, where different average values of the normalized volumetric proton consumption rate ($\bar{r}_i^{H^+}$) have been found for each bioreactor (i). In order to isolate the tank variation, α_v has been defined as the ratio:

$$\alpha_v = \frac{\bar{r}_i^{H^+}}{\max_i \bar{r}_i^{H^+}} \tag{6.15}$$

so that $\alpha_v \in [0, 1]$.

6.2.2.2 Determining α_p

After determining α_v for all bioreactors, attention can now be turned to the determination of α_p. The approach is similar to the one used for determining α_v. Figure 6.3 shows the distribution of data points from a number of batches conducted in a number of different bioreactors. Based on this data, α_p is calculated by least squares regression. This leads to a value of $\alpha_p = 220$ ($\pm$ 1) g biomass / N-mole. The theoretical value of $\alpha_p = M_{DW}/Y_{xH}$, however neither of these values are known for the strain in question. Using $M_{DW}/Y_{xH} = 27$ g/C-mole / 0.14 N-mole/C-mole, a theoretical value of $\alpha_p = 193$ g biomass / N-mole is formed, which is 12 % lowered than the experimentally determined value. It has not been possible to identify an explanation for the observed discrepancy, however possible explanations could be: incorrect estimates of M_{DW} and Y_{xH} and non-ideal mixing properties in large scale bioreactors.

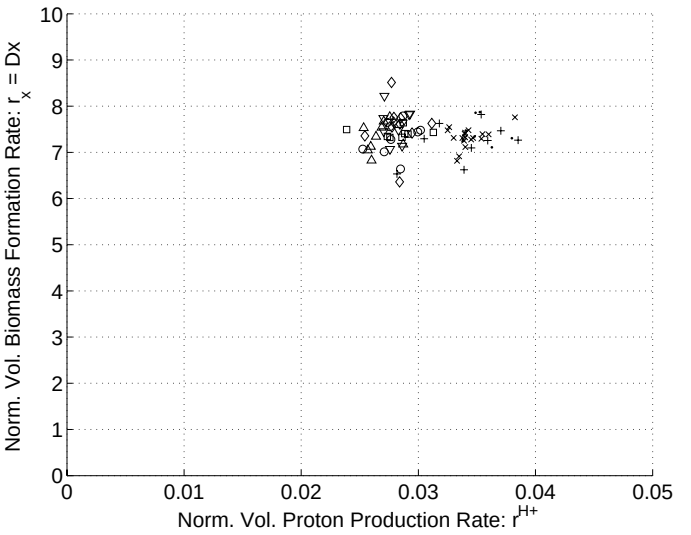


Figure 6.2. Illustration of vessel differences between bioreactor 1 (legends: x, $\cdot$, $+$) and bioreactor 2 (legends: $\square$, $\triangle$, $\triangledown$, $\Diamond$, $\circ$). The figure shows how the average ratio between the normalized volumetric proton consumption rate and the normalized biomass production rate can be different due to vessel properties. A number of batches are shown for each vessel, each represented by different legends.

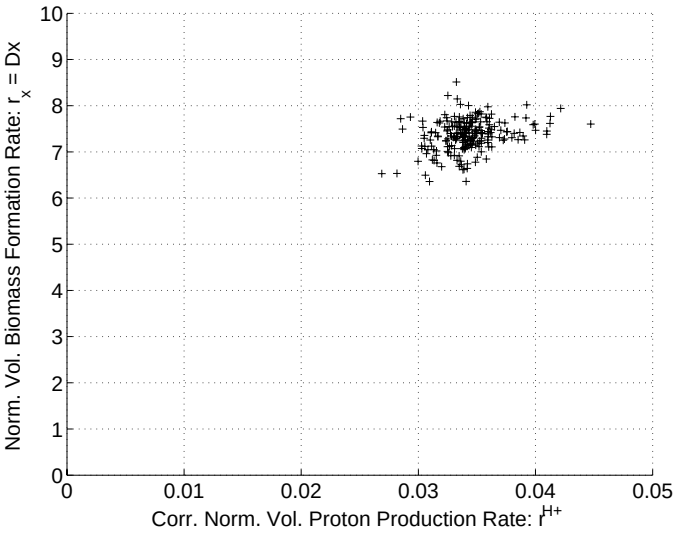


Figure 6.3. The illustration shows how the normalized biomass production rate as a function of the corrected normalized volumetric proton consumption rate r^{H+}/α_v. Data from a number of batches in a number of different bioreactors are shown with the same legend.

6.2.3 FPEM based Sensor for Product Concentration

To develop a process soft sensor for prediction of the insulin precursor concentration, physiological knowledge of the recombinant yeast strain is used. It is known that the control of the promoter for transcription of the product gene is linked to the activity of the glycolysis of the recombinant strain (Diers *et al.*, 1991). To simplify the model formulation the following assumptions are made:

- Production rate of product (r_p) proportional to production rate of biomass (r_x)[4]
- High stability of recombinant gene
- No influence from transport and folding in organelles on production rate
- Effective excretion of product

A high stability of the recombinant gene ensures that no decay in specific productivity of the insulin precursor is experienced over time. Furthermore by assuming that the transport of the insulin precursor through the organelles of the cell does not have any influence on the production rate, combined with effective folding and excretion of the insulin precursor to the abiotic phase, the rate limiting step of the cellular production process becomes transcription of the recombinant gene.

Based on the above assumptions the following model for the production rate of the product (p) is proposed:

$$r_p \propto r_x = \frac{M_{DW} \cdot r_{H^+}}{Y_{xH}} \tag{6.16}$$

Introducing a parameter (β) accounting for the issues relating to non-ideal process conditions and variations in growth stoichiometry (Y_{xH}) and cell composition (M_{DW}) a simple dynamic mass balance on the product becomes:

$$\frac{dp}{dt} = r_p - Dp = \beta \frac{F_{NH_3} - F_s q_{s,H^+}}{V} - Dp \tag{6.17}$$

leading to the product predictor:

$$p_{k+1} = p_k \cdot exp\left(\left(\frac{\beta \cdot (F_{NH_3,k} - F_{s,k} q_{s,H^+})}{V_k \cdot x_k} - D_k \right) (t_{k+1} - t_k) \right) \tag{6.18}$$

where $\beta = g(t, M_{DW}, Y_{xH}, \text{vessel properties})$ is determined for the individual bioreactor. In this work it has been assumed that β is independent of time variations and that the contributions from physiology (M_{DW}, Y_{xH}) and vessel properties can be modelled as two constant values in a multiplicative model:

$$\beta = \beta_p \cdot \beta_v \tag{6.19}$$

where β_p represents the physiological contribution and β_v represents the contribution from vessel properties. The details of the parameter estimations are given in section 6.2.2.

[4]It is assumed that other cultivation parameteres such as temperature, pH and substrate composition are constant.

6.2.3.1 Determining β_p

Assuming that $\beta_v = \alpha_v$, β_p has been determined using an identical method to the one illustrated for α_p. $\beta_p = 376$ ($\pm$ 2) units of product / N-mole.

6.2.4 Results - FPEM based Sensor

The cultivation data reported in this chapter has been normalized by dividing the process variables with a known set of normalization parameters preserving the units of the variables. The normalization has been done in order to protect sensitive information.

6.2.4.1 Estimating Biomass Concentration

Three validation examples of the normalized biomass concentration soft sensor are illustrated in figure 6.4. The calculated *RMSEP* values indicate good agreement between soft sensor estimations and offline measurements. The top graph in figure 6.4 shows an example of a normal trajectory of the biomass concentration, from which it can be seen that the operation is carried out without major incidents. It should be borne in mind that there is a delay of more than 24 hours on the offline sample results. The offline measurements of the biomass concentration can be seen to be rather constant during continuous operation, and it seems that no offline measurements were taken during the fed-batch operation except for the point at the very start of this operation.

The middle graph in figure 6.4 illustrates how a slight change in the process conditions at normalized time 3 results in a slight decrease in the biomass concentration. No information on the cause of this decrease has been found, but a decrease in the glucose concentration in the substrate might be a valid explanation. It is interesting to notice that the soft sensor correctly captures this change in the process conditions. The offline measurements of the biomass concentration does not clearly indicate this change, however the more frequently sampled offline measurements of the product concentration shown in figure 6.5 (middle graph) indicate that the decrease indicated by the soft sensor for the biomass concentration actually happens.

The bottom graph in figure 6.4 also shows an example of how a soft sensor can provide valuable online information of the process conditions. At normalized time 3, an operator made a mistake while correcting the feed rate of glucose-containing substrate. This led to a decrease in the biomass concentration in the bioreactor as indicated by the soft sensor and supported by the offline measurements. The fault went unnoticed for several operator shifts, since no direct indication of the process condition was available at the time and the results of the offline measurements were still not available. When the fault was noticed the cultivation was brought back on track by reestablishing the correct feed rate momentarily. The soft sensor provides two pieces of valuable information in this scenario. First, the online signal would directly indicate to the operators that the biomass concentration was decreasing. Second, as the fault was noticed the estimated biomass concentration would provide an indication of how quickly the correct feed rate could be reestablished, since a risk of triggering ethanol production is present when increasing the feed rate too quickly (Postma *et al.*, 1989b).

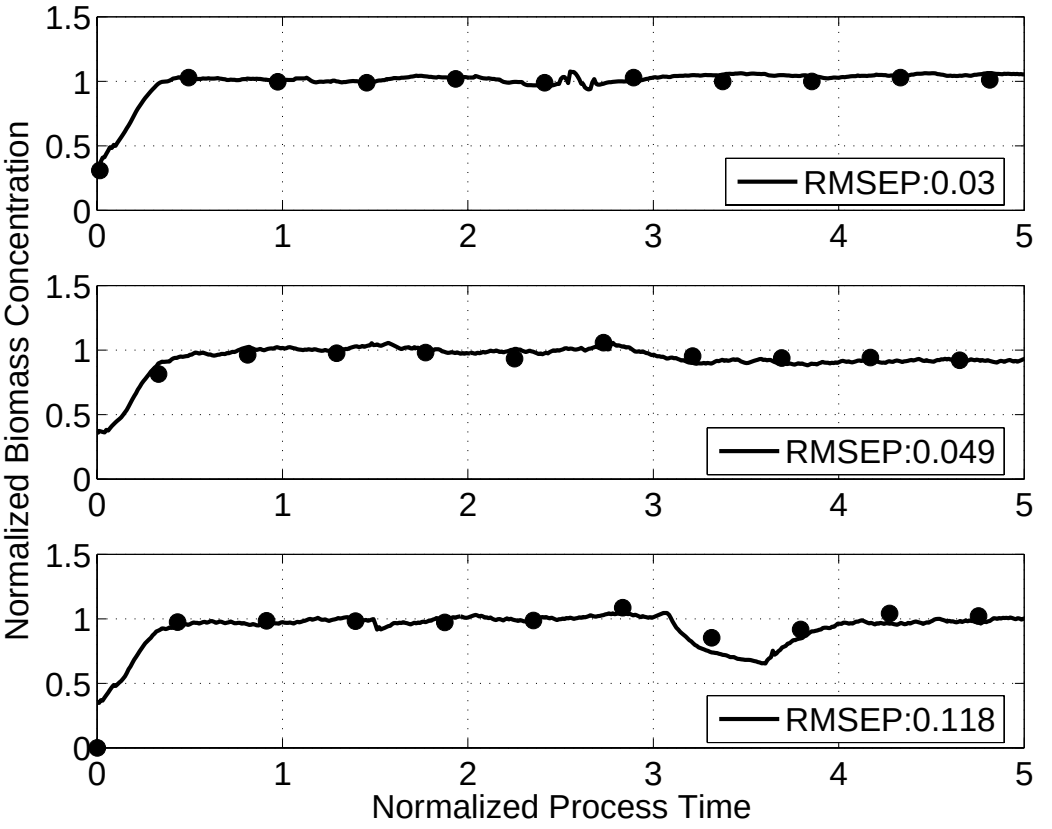


Figure 6.4. Comparison of signals from soft sensors (−) and analytical measurements (⧫) of the biomass concentration for three cultivations (top, middle and bottom). Only data from fed-batch and continuous operation is considered. Note that in an online scenario the results of analytical measurements are delayed more than 24 hours after sampling the culture broth, while the soft sensor provides an online signal. The change from fed-batch to continuous operation occurs between 0.4 and 0.5 of the normalized process time. $RMSEP$: root mean square error of prediction (see equation 6.1) between offline measurement and soft sensor signal at the corresponding time point. The relatively large $RMSEP$ in the fed-batch period is mostly caused by to an erroneous offline measurement at the beginning of the fed-batch operation.

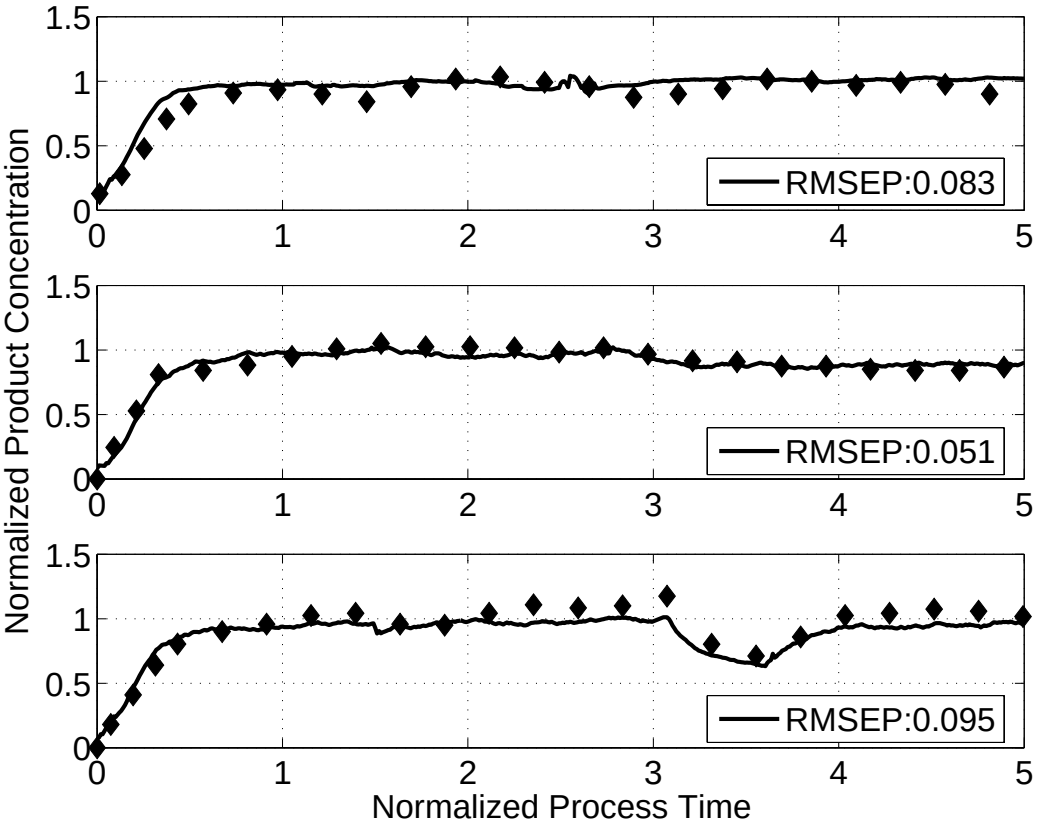


Figure 6.5. Comparison of signals from soft sensors (−) and analytical measurements (⧫) of the product concentration for the same three cultivations as in figure 6.4 (top, middle and bottom). Only data from fed-batch and continuous operation is considered. Note that in an online scenario the results of analytical measurements are delayed up to 24 hours after sampling the culture broth, while the soft sensors provide an online signal. The change from fed-batch to continuous operation occurs between 0.4 and 0.5 of the normalized process time. $RMSEP$: root mean square error of prediction (see equation 6.1) between offline measurement and soft sensor signal at the corresponding time point.

6.2.5 Estimating Product Concentration

Three application examples of the product concentration soft sensor are illustrated in figure 6.5. The soft sensors are activated after the batch phase and used for the fed-batch and continuous phases of the cultivation with constant β values. More offline measurements of the product concentration are available, since it is the product concentration trajectory rather than the biomass concentration trajectory that is used as a measure of cultivation performance with respect to quality control and process economics. The calculated *RMSEP* values indicate good agreement between soft sensor estimations and offline measurements, although higher values are seen than when compared to the soft sensor of the biomass concentration.

The top graph in figure 6.5 shows an example of a normal (and normalized) trajectory of the offline measurements of the product concentration. Apparently cyclic fluctuations with a period between 1 to 2 time units in the product concentration during continuous operation are present, however no explanation for this phenomenon is known. In the last stage of the fed-batch operation around time point 0.5 it can be seen that there is a discrepancy between the soft sensor signal and the offline measurements; this difference has not been explained.

The middle graph in figure 6.5 illustrates an apparent decrease in the product concentration starting at time point 3. A decrease was also seen in the biomass concentration (middle graph, figure 6.4). As described above an explanation for the observed decrease could be a decrease in the glucose concentration in the substrate. The decrease in the product concentration is more easily seen in the product concentration trajectories due to the more frequent sampling, and is supported by both soft sensors.

The bottom graph in figure 6.5 shows the impact on the product concentration, when a wrong set point for the feed rate of the substrate containing glucose was entered into the process control system. Again the more frequently sampling of the product concentration provides a good illustration of both the effect of the process upset and the performance of the soft sensor. The soft sensor correctly indicates the decreasing product concentration and also that the cultivation is brought back on track as the faulty set point is discovered and corrected.

6.2.6 Additional Remarks

The performance of the two FPEM soft sensors appear to be good, providing valuable online information for monitoring and control of a yeast cultivation. As illustrated in this work, the use of soft sensors also provides valuable information for batch analysis and documentation after the cultivation is terminated. The models used in the soft sensors are very simple as is the method for estimation of the model parameters. In an attempt to improve the performance of the soft sensors more complex models can be investigated as can methods for estimating the model parameters.

6.2.7 Multiway Projection to Latent Structures (MPLS)

Process monitoring and prediction of end quality using MPLS have been illustrated by a number of research groups e.g. Nomikos and MacGregor (1995), Louverse *et al.*

(1999) and Gregersen and Jørgensen (1999). The general idea behind MPLS is that an empirical model is built on measurements from reference batches operated under normal operating conditions producing a good quality product in terms of high concentration. This work has focused on the prediction possibilities of the MPLS. The available online measurements are used to estimate or predict product quality, which is desirable, since only a limited number of analytical measurements of the quality variables is available in an offline fashion. The online measurements are arranged in a three-way array $\underline{\mathbf{X}}$ $(I \times J \times K)$ where I is the number of data batches, J is the number of variables and K is the number of samples. In general the quality measurements are arranged in a two-way array $\mathbf{Y}$ $(I \times M)$ where M is the number of quality variables usually measured at the end of the data batch. By unfolding of $\underline{\mathbf{X}}$ $(I \times J \times K)$ by placing K $(I \times J)$-slides next to each other as illustrated in figure 6.6, a two-way array $\mathbf{X}$ $(I \times JK)$ is formed and an ordinary PLS can be performed on the $\mathbf{X}$ and $\mathbf{Y}$ after column-wise mean centering and scaling to unit variance of the two arrays (Nomikos and MacGregor, 1995).

By the MPLS-algorithm the arrays are decomposed into C score vectors ($\mathbf{t}$ $(I \times 1)$, $\mathbf{u}$ $(I \times 1)$), weighting vectors ($\mathbf{w}$ $(JK \times 1)$) and loading vectors ($\mathbf{p}$ $(JK \times 1)$, $\mathbf{q}$ $(M \times 1)$) along with two residual matrices ($\mathbf{E}$ $(I \times JK)$, $\mathbf{F}$ $(I \times M)$):

$$\mathbf{X} = \sum_{c=1}^{C} \mathbf{t}_c \mathbf{p}_c^T + \mathbf{E} \qquad \mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \tag{6.20}$$

$$\mathbf{Y} = \sum_{c=1}^{C} \mathbf{u}_c \mathbf{q}_c^T + \mathbf{F} \qquad \mathbf{Y} = \mathbf{U}\mathbf{Q}^T + \mathbf{F} \tag{6.21}$$

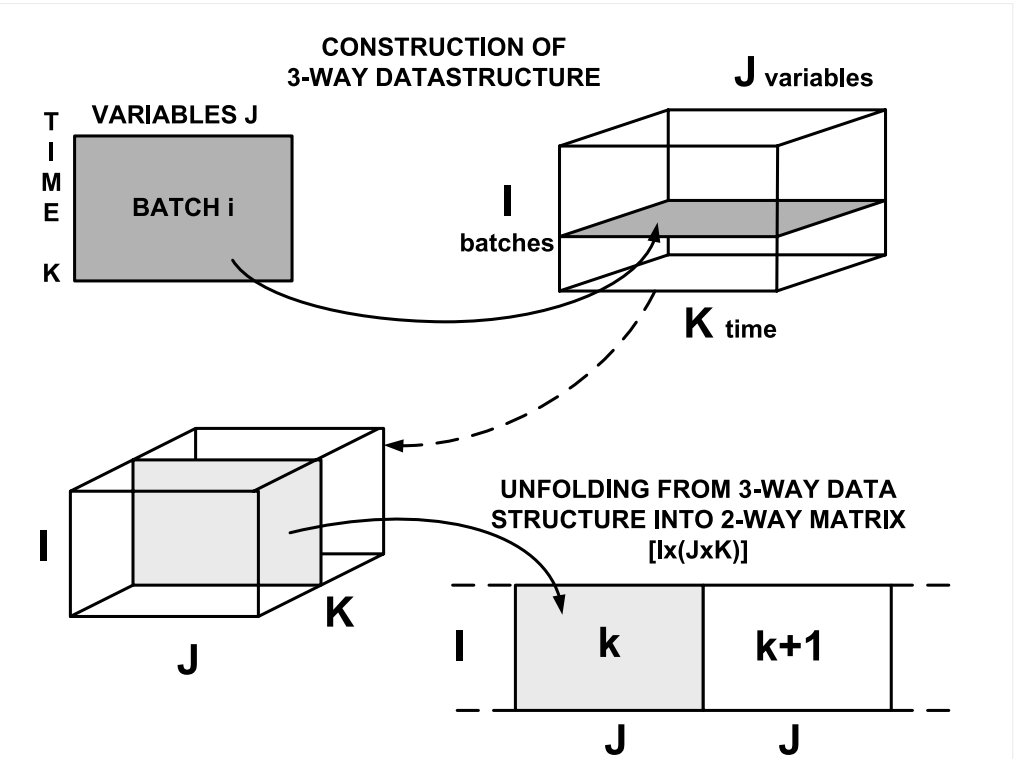


Figure 6.6. Illustration of construction of 3-way data array of size $(I \times J \times K)$ followed by an unfolding to form a 2-way data array of size $(I \times JK)$. I: number of batches, J: number of variables and K: number of sampling points.

where $\mathbf{T}$ $(I \times C)$, $\mathbf{P}$ $(JK \times C)$, $\mathbf{U}$ $(I \times C)$ and $\mathbf{Q}$ $(M \times C)$ results when gathering the C vectors of $\mathbf{t}$, $\mathbf{p}$, $\mathbf{u}$ and $\mathbf{q}$ in matrices.

Using the MPLS-algorithm a regression equation can be formulated:

$$\hat{\mathbf{Y}} = \mathbf{XB}, \text{ with } \mathbf{B} = \mathbf{W}\left(\mathbf{P}^T\mathbf{W}\right)^{-1}\mathbf{Q}^T \tag{6.22}$$

This regression model $\mathbf{B}$ $(JK \times M)$ can then be used for online prediction of the end quality of the batch provided that a suitable method for the estimation of future online measurements is available (Nomikos and MacGregor, 1995). This work has been applying the method using the J measurements obtained at the last sampling number k to fill in the empty spaces $(1 \times J(K-k))$ as illustrated in figure 6.7.

The number of PLS-components (C) necessary to obtain a desired level of regression can be evaluated using different (cross-)validation techniques. In this work the root mean square error of prediction $(RMSEP)$ is used (see equation 6.1). For increasing numbers of PLS-components C used for model identification the $RMSEP$ is evaluated on validation data, where the lowest value of $RMSEP$ indicates the number of PLS-components C to be used.

6.2.8 MPLS for On-line Prediction and Estimation

In the case where quality measurements are taken frequently during the batch run, the MPLS- framework can be used for estimation and prediction of the intra-batch quality. For all the batches considered in this work, both online measurements and offline quality measurements in each batch have been subsampled to the same frequency by interpolation. The offline data series have been linearly interpolated. For the online measurements a kernel estimator for smoothing using a tricubic kernel with a local linear fit of 3 nearest neighbors have been applied (Hastie *et al.*, 2001).

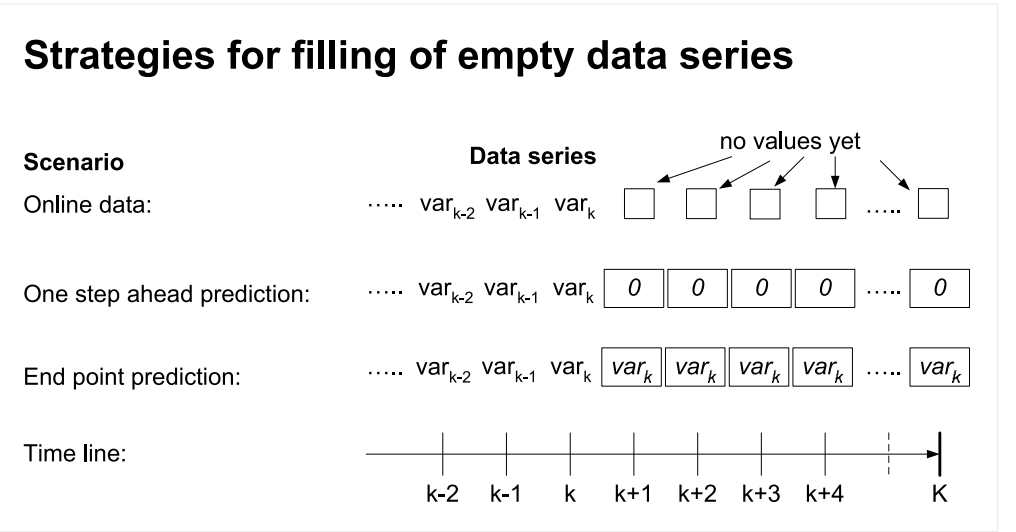


Figure 6.7. The strategy for filling in future values in data series. Top: online data series with data up until sample point k. Middle: One-step ahead prediction. Bottom: End-point prediction. From sample point $k+1$ and to the end of the data sequence the value at sampling point k is repeated.

With the interpolated data a $\mathbf{Y}$ array is obtained. At sample number k in a new batch the full batch profiles of the quality variables $\hat{Y}_k$ can be obtained by filling in the empty spaces in X_k as described above and applying the regression matrix $\mathbf{B}$:

$$\hat{Y}_k = X_k\mathbf{B} \tag{6.23}$$

6.2.9 MPLS Applied on Industrial Data

In this work the only quality variable to be regressed was the product concentration. 11 online measured variables were sampled 180 times during the fed-batch operation and the initial period of the continuous operation until the product concentration profile was stationary as shown in figure 6.5 at time point 1. Data from this part of the cultivation will be referred to as the data batch. Information related to the last data point in the data batch will be referred to as the end point *e.g.* end point concentration, although the process is operated beyond this time point. No actual measurement of the end point product concentration is available, so the end point concentrations are estimated using linearly interpolated values. The online measured variables that have been used are shown in table 6.1.

13 cultivations conducted under normal operating conditions were used for the model identification (M-data batches), while 7 validation batches (V-data batches) were used to determine the number of PLS-components to be included in the model evaluated by the $RMSEP$ as describe above. The explained variance and $RMSEP$ for the 5 first PLS-components are shown in table 6.2. It is interesting to note that the $RMSEP$ evaluation indicates that 3 PLS-components should be included in the model, explaining 71 % of the variation in $\mathbf{Y}$.

Table 6.1. On-line variables used in the MPLS regression model. Substrate dosing 1 and 2 are the carbon and growth factor sources respectively. These are fed in a constant ratio. CO_2 production and O_2 consumption are calculated as a difference in the concentration between aeration inlet and off-gas outlet of the bioreactor.

1. Ammonia flow rate	7. CO_2 production
2. pH	8. O_2 consumption
3. Aeration rate	9. Weight of broth in bioreactor
4. Stirrer speed	10. Temperature
5. Substrate dosing 1	11. Dilution rate
6. Substrate dosing 2	

Table 6.2. Explained variance of $\mathbf{X}$ and $\mathbf{Y}$. Mean $RMSEP$ from the validation.

	No	of	PLS	Comp	
Expl. var	1	2	3	4	5
$\mathbf{X}$	12 %	25 %	37 %	45 %	55 %
$\mathbf{Y}$	38 %	59 %	71 %	85 %	90 %
$RMSEP$	0.085	0.071	**0.071**	0.100	0.100

The model performance was then investigated using the 7 V-data batches along with 13 additional batches (A-data batches), the latter having normal end-point concentrations of the product, but undergoing small process upsets during operation. A comparison between the MPLS estimated and linearly interpolated product concentration at the end of the batch is shown in figure 6.8. The latter estimator is assumed to be comparable to the analytical measurements of the product concentration. From the figure it is seen how the MPLS estimates at worst are within 12 % of the linearly interpolated product concentrations for model, validation and A-data batches, which can be considered good for an industrial process in light of the uncertainties related to the determination of the actual product concentration at the end point *e.g.* sampling uncertainty and analytical measurement uncertainty.

Figure 6.9(a) shows the prediction for a validation (V-)data batch. A reasonable description of the variations in the offline analytical measurements can be seen by the one-step ahead MPLS-prediction. From time 0.65 and to the end of the batch some variations in both the one-step ahead and end point prediction (starting at coordinates (0,1)) can be noticed. For unknown reasons the feeding of substrate to the cultivation was temporarily stopped between a normalized time of 0.66 to 0.82 and again between 1.42 and 1.66. Despite these severe disturbances to the process both the one-step ahead and end point predictions provide reasonable and robust descriptions of the product concentration. The behavior of the end point predictor at the first stop in the process between a normalized time of 0.66 to 0.82 indicates that the process is far from its normal operating range and maintaining the process

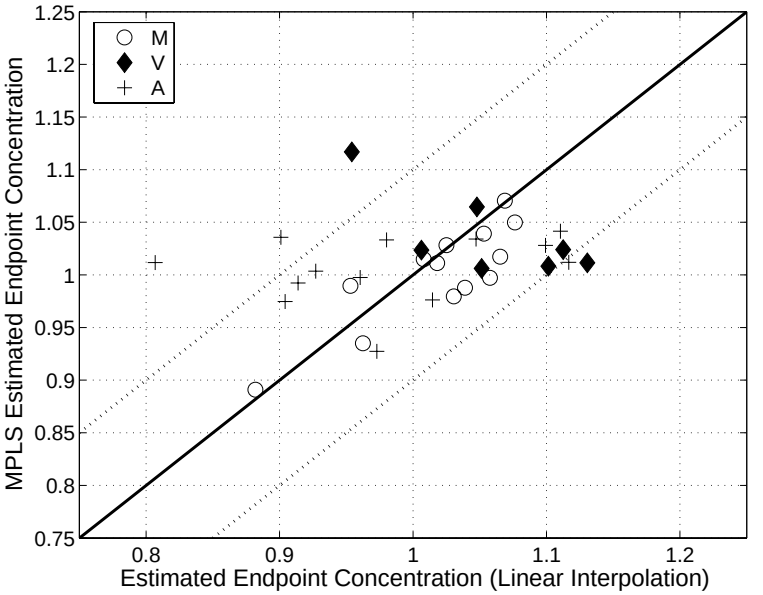


Figure 6.8. Estimated product concentrations at end of the data batches. No measurement of the actual product concentration is available, but linearly interpolated values are used instead. Linearly interpolated concentrations vs. MPLS estimated concentrations are shown for: (○) M-, (♦) V- and (+) A-data batches. The full diagonal line indicate x=y, while the two dotted lines above and below the full diagonal line indicate offsets of ±0.10 to illustrate an analytical uncertainty of 10% on the measurement of the product concentration.

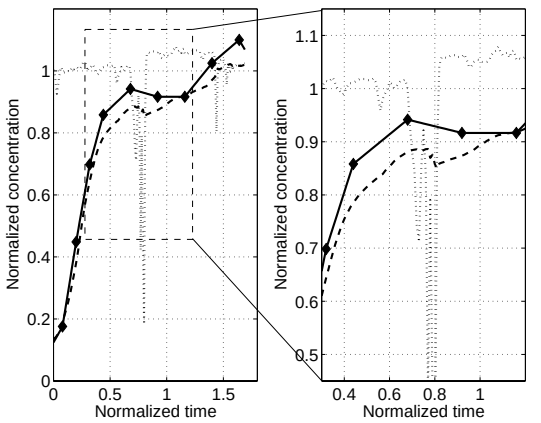
in this state will have a strong negative effect on the product concentration at the end point. As the feeding is resumed the MPLS regression model indicates that a positive effect on the end point concentration has been obtained, since the predicted end point concentration is slightly above the normal average of 1.

Figure 6.9(b) presents the results of using the MPLS regression models on one of the additional (A-)data batches, which has neither been used for modeling nor for validation of the models. This cultivation was conducted in parallel but slightly delayed compared to the validation cultivation described above and the same two process upset occurred, however appearing to occur earlier in the cultivation due to the time delay. Both process upsets were quite similar to the ones in the validation batch. The impact on the predicted product concentration was however slightly different, since no such process upsets have been present in the modeling batches. Therefore comparing and discussing the predicted values during the process upset is not relevant, since it is outside the scope of the model. It is however interesting to note the development of the one-step ahead prediction trajectory following the process upset, where an increase in the product concentration is predicted, which also is indicated by the analytical measurements shown here, as in figure 6.9(a), at a slower rate.

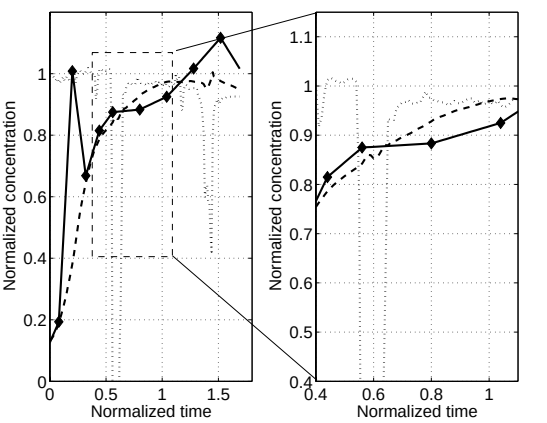
Figure 6.9(c) shows another example of a cultivation where a process upset occurs around a normalized time of 1.0 to 1.2. It appears that both the one-step ahead prediction and the end point predictions are off for the rest of the data batch, when compared to the analytical measurements. This might be true however the discrepancies are almost solely highlighted by the analytical measurement at the normalized time of 1.45, while it appears that the subsequent analytical measurement, which falls outside the range of the MPLS model, it is in fact close to the normal product concentration of 1 as indicated by the linearly interpolated line.

Figure 6.9(d) shows an example of a cultivation carried out according to the production recipe except for the time period between normalized time of 1.55 to 1.7, where the feeding of substrates is stopped. It is interesting to see how well the product concentration profile during fed-batch operation is captured by the one-step ahead predictor up until the process upset occurs. It is also interesting to note how the end point prediction slowly increases and coincides with the one-step ahead prediction just as the process upset occurs, indicating how balanced the process was operated up until this time.

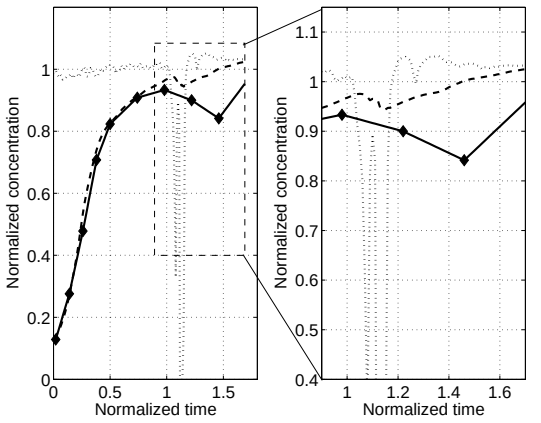
The information on the predicted end point product concentration in figure 6.9 is highlighted in figure 6.10, where the trajectories of the prediction error between the linearly interpolated end point product concentration and the one-step ahead prediction are shown. The prediction errors are illustrated for the same four data batches presented in figure 6.9. The dotted lines represents the $\pm$ 10 % value of the linear interpolated end point concentration and it can be seen that for 3 of the 4 cases the predicted end point concentration is within $\pm$ 10 % of the linearly interpolated end point concentration except during process upsets. It is interesting to note how the predicted end point concentration of data batch A3 is consistently above the final end point concentration until the process upset occurs as described above.



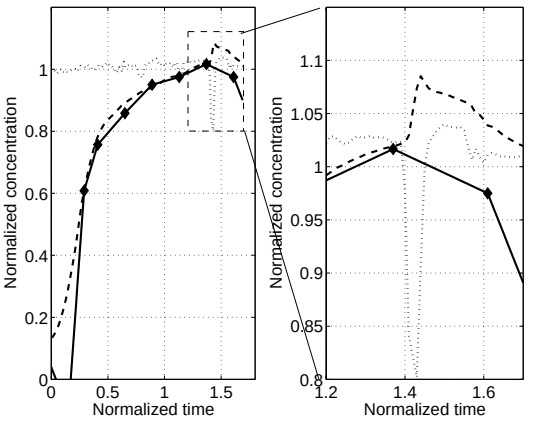
(a) V-data batch



(b) A-data batch 1



(c) A-data batch 2



(d) A-data batch 3

Figure 6.9. MPLS product concentration predictions in a validation data batch (a) and three test A-data batches (b-d). (−) linear interpolations between analytical measurements (⧫), (−−) one-step ahead prediction and (··) end of batch prediction. Zooms of interesting process behavior are provided for each data batch as indicated on the individual figures. For each modeling/sampling point, two models have been determined: a one-step ahead prediction model and an end of batch prediction model. The strategy for filling in future values in data series is described in figure 6.7.

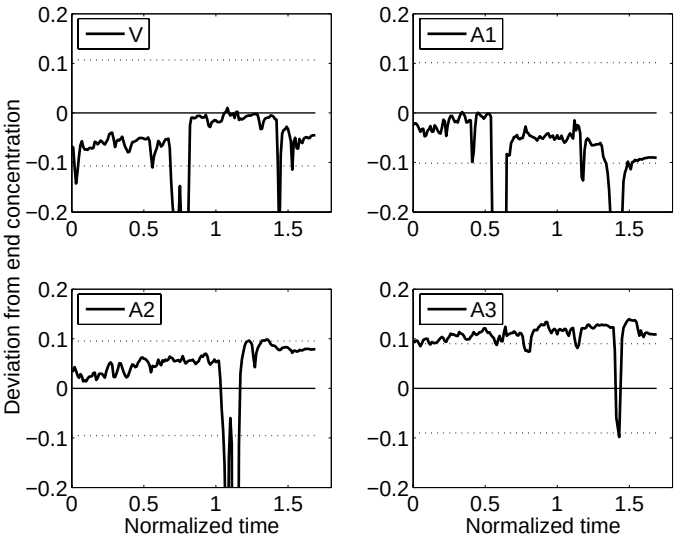


Figure 6.10. Prediction error of end point product concentration in 1 V-data batch (top left, V) and 3 A-data batches (A1-A3). (−) MPLS-end point prediction and (··) ± 10% errors of the linearly interpolated product concentration from offline analytical measurements.

6.3 Performance Comparison of FPEM- versus MPLS-predictor

In the above two methods for one-step ahead prediction of the product concentration have been developed and tested, namely a soft sensor based upon a FPEM model (FPEM-predictor) and a soft sensor based upon a MPLS model (MPLS-predictor). The performance of the two predictors will be compared in the following.

Figure 6.11 illustrates the performance of the two soft sensors compared to the analytical offline product concentration measurements. Furthermore the figures display deviations from the average modeling trajectories of the product concentration as well as the ± 10 % trajectories from this average. As a performance measure the deviations between the two soft sensors and the average modeling trajectory are shown, rather than comparing the soft sensors to the offline measurements. This distintion has been made, since a number of uncertainties are associated with the steps of sampling and handling of the offline samples prior to the analytical measurement. These sampling and handling uncertainties are very difficult to quantify and only single measurements are available at each sampling point. Therefore it has been assumed that a better basis for comparision is the average modeling trajectory, despite the fact that a number of data batches are known to behave different from the modeling batches.

From the four illustrations in figure 6.11 it appears that the MPLS-predictor provides a better description of the analytical offline measurements than the FPEM-predictor. It is however important to mention that the FPEM-predictor is very sensitive to the choice of initial conditions *i.e.* initial biomass concentration. Furthermore the FPEM-predictor is based on assumption of constant pH level. At the transistion from batch to fed-batch operation fluctuations in pH can occur *e.g.*

if the onset of substrate feeding is delayed. As a consequence the assumption of a proportionality between ammonia addition and biomass formation might not be valid during the first hours of the fed-batch operation. This is actually a general observation in all of the four illustrations in figure 6.11. This discrepancy can be compensated for by extending the modeling by making the parameter β (and α) time-dependent or online adaptation of model parameter values based on offline measurements.

In general it appears that the MPLS-predictor only shows a small variation around the average trajectories of the product concentration compared to the FPEM-predictor and the analytical measurement, which is seen from the $RMSEP$ values listed in the figure and based upon the average trajectories. It is however highly relevant to note that despite the fluctuation and deviation in the FPEM-predictor signal, the values are mostly within 20% of both the analytical offline measurements as well as the MPLS-predictor signal.

From the comparison it appears that the FPEM-predictor is inferior to the MPLS-predictor. However with the earlier discussion on initialization problems of the FPEM-predictor in mind, the FPEM- predictor can be seen to perform reasonably well 20% into the data-batch. Here it is important to keep in mind that in the comparison above cultivations in which process upsets occurred were included. Since the MPLS-predictor makes use of much more information than does the FPEM-predictor, and considers a longer time horizon than the FPEM-predictor, it is not surprising that the MPLS-predictor has a slightly better performance. However as was shown in figure 6.4 the FPEM-predictor is not limited to a certain time interval as is the MPLS-predictor in the current form and is therefore more easily applicable.

6.4 Discussion and Conclusion

In this chapter two different methods for obtaining quantitative information from a cultivation process has been presented and preliminarily compared. The methods have been applied using online process data from an industrial cultivation process to illustrate the type and quality of information obtainable with these methods.

A soft sensor was developed for monitoring of the biomass concentration based on FPEM using the feed rate of ammonia, volume of broth and the dilution rate as inputs. Application of the soft sensor using online process data gave a reasonable description of the variations seen in the analytical measurements, leading to the conclusion that the implementation of this device will enable a very simple yet highly attractive way of providing online monitoring of the biomass concentration.

A similar soft sensor was developed for monitoring of the product concentration using the same framework as the biomass concentration soft sensor. Although complex cellular processes are involved in the processes for generating the insulin precursor a very simple model was developed by only slightly modifying the FPEM used for modelling the biomass concentration. Applied on the industrial data this simple soft sensor was also able to give a reasonable description of the general product concentration trajectory thereby making online monitoring of the product concentration possible if implemented. Despite the complexity of the process and the limited knowledge of the mechanisms behind the expression of the product, it was possible

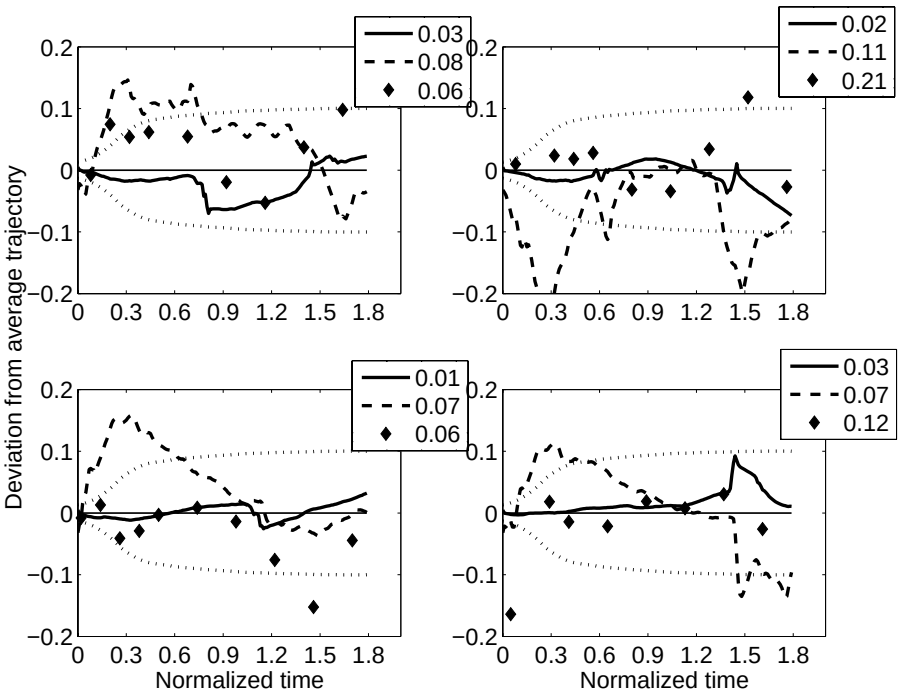


Figure 6.11. Prediction errors of one validation data batch (V,top left) and three test A-data batches (A1-3, the same data batches and positioning as illustrated in figures 6.9 and 6.10) reported as deviations from average values of linear interpolated product concentration profiles in the modeling data batches, and analytical measurements and one-step ahead product concentration predictors, respectively. MPLS-predictor ($-$, bold) and FPEM-predictor ($--$, bold). ($\blacklozenge$) analytical measurements and $(\cdot\cdot) \pm 10\%$ on the average values of linear interpolated product concentration profiles in the modeling data batches. $RMSEP$ estimates have been added.

with very simple relations to provide reasonable online estimation of the concentration trajectory of the product; estimations that can result in a whole new way of monitoring and controlling industrial cultivation processes.

An alternative approach for monitoring component concentrations in a process is through process chemometrics. A model for predicting the product concentration based on the MPLS-algorithm driven by existing on-line process measurements was developed producing a linear model describing changes around an average trajectory. The model was tested on the industrial data and indicated that both one-step ahead and end point predictions of the product concentrations came within 5-10 % of analytical offline measurements, but of course without analytical time delay.

The MPLS-predictor for the one-step ahead prediction was compared with the simple product concentration soft sensor (FPEM-predictor), where the first gave a slightly better description of the variations in the product concentration. It is however important to keep in mind that the MPLS-predictor in its current shape, has its application limited to the data period chosen in the modeling, which is the fed-batch operation and a part of the continuous operation. The FPEM-predictor can be used throughout both fed-batch and continuous operation.

In conclusion this work has provided insight into tools for monitoring a given cultivation process with respect to biomass concentration and product concentration. Using only data available in the process control system of an industrial cultivation plant, it is possible to provide online information on the evolution of the biomass concentration and product concentration profiles using very simple relations.

7

Estimating The Elemental Composition of Biomass of an Industrial Recombinant Strain of *Saccharomyces cerevisiae*

Two methodologies were used to determine the elemental composition of biomass of an industrial recombinant strain of S. cerevisiae for production of an insulin precursor. The traditional approach using elemental analysis revealed an elemental composition similar within 5% to the ones reported in the literature for other strains of S. cerevisiae, while an approach using macroscopic mass balance also resulted in a composition within 5%. It was not possible to determine which of the two methodologies provided the more robust estimate of the biomass composition, since only a limited number of samples were available. It was assumed that 3.6 w% of the biomass concentration measured as dry weight was due to water, and an ash content of 3.8 w% accounting for the contributions of metals, sulfur and phosphorous to the biomass composition. The combined water and ash contribution to the biomass is referred to as the residual. The following elemental composition of the recombinant strain of S. cerevisiae is suggested based on the methodology using elemental analysis[1]*:*

$$CH_{1.82}O_{0.576}N_{0.146} \quad \kappa_x = 4.23; \quad M_{DW} = 27.1 \text{ } g/C\text{-}mole \text{ } (f_r\text{: } 7.3 \text{ } w\%)$$

In a cultivation process the biomass serves the role as catalyst for the conversion of substrates into products, and for this reason quantitative information of the biomass is important in order to understand, improve and enhance the performance of the bioreactor. The work in this thesis is focussed on the construction of simple models in order to provide tools for analysis and monitoring of cultivation processes. Since the production of biomass accounts for a large portion of the conversion of the carbon source in the substrate, it is important to have as accurate information as possible on the biomass conversion rate and concentration, since these play a pivotal role for the evaluation of mass and component balances used for model based monitoring and control of cultivation processes. An alternative to the estimation of the elemental composition of biomass, would be to use elemental compositions cited in literature on *Saccharomyces cerevisiae*. This is however not desirable for two reasons. 1) Genetic

[1] κ_x represents the degree of reduction of biomass per C-mole biomass.

modifications have been applied to the recombinant industrial strain investigated in this work in order to improve productivity of the peptide product, an insulin precursor. It is not known to what extent the genetic modifications has had an effect on the elemental composition. 2) Even small deviations between assumed and the true elemental composition will have an effect on models assuming the elemental mass balances to close. So rather than addressing uncertainties by postulating assumptions, a quantitative investigation of the elemental composition is performed.

Complex substrates are used for the industrial cultivation of the recombinant yeast at Novo Nordisk A/S, as opposed to the defined substrates often used in scientific studies of physiology and biomass composition. The presence of di- and trisaccharides (*e.g.* maltose, maltotriose[2]) in the glucose syrup combined with yeast extract being part of the complex nutrient substrate could influence the enzyme pool/protein content of the cell and therefore the composition of the cell.

In this work the elemental composition of biomass during continuous operation will be investigated using macroscopic elemental balancing on data obtained from pilot plant cultivations. The identified composition will be compared to results from a chemical analysis of the elemental composition and finally the results will be compared to compositions reported in the open literature. The purpose of estimating the biomass composition is two-fold. Firstly, it is relevant to obtain information onto the composition of biomass for further analysis of the metabolism and physiology of the recombinant strain, and secondly it is interesting to see how accurately the biomass composition derived from overall macroscopic mass balances can describe the elemental composition, since this could open up for an extended use of mass balances in industrial cultivations. It should be noted that all of the data reported below has been normalized, except for the elemental compositions of biomass.

Following this introduction, a brief review of the literature on elemental composition of biomass is provided in section 7.1. The materials and methods used in this chapter is then described in section 7.2, and followed by a presentation of the methodologies used for setting up macroscopic elemental balances in section 7.3. The results of the analysis of the macroscopic elemental balancing and the elemental analysis are presented in section 7.4. Section 7.5 provides a discussion of the results, while a conclusion is presented in section 7.6.

7.1 Literature Review

For *S. cerevisiae* grown on glucose in carbon-limited chemostat cultivations a number of biomass compositions have been reported in the literature. Herwig *et al.* (2001) investigated the biomass composition of the research strain CEN.PK 113-7D at different dilution rates (0.04-0.37 L/L/hr), all using 20 g/L glucose in the feed. Based on results obtained using elemental analysis, they reported the average biomass composition to be $CH_{1.75}O_{0.50}N_{0.16}$ (ash 7.2 w%) with a increasing content of nitrogen at increasing dilution rates, explained by an increased accumulation of protein on behalf of storage carbohydrates that were accumulating at decreasing dilution rates. For the same strain Lange and Heijnen (2001) performed an extensive investigation of the elemental and molecular biomass composition at dilution rates ranging

[2]Byproducts from the enzymatic degradation of starch to glucose.

from 0.02-0.2 L/L/hr using glucose concentrations between 5-10 g/L. Using both elemental analysis and measuring the intracellular concentration of macromolecules (*e.g.* carbohydrates, proteins, lipids) combined with statistical data reconciliation, they also found an increasing content of nitrogen at increasing dilution rates and decreasing content of oxygen at increasing dilution rates, and explained this by an increase in the protein concentration, as did Herwig *et al.* The ash content[3] of Lange *et al.* was approximately 3.7 w% at all dilution rates, and they discussed how an additional small presence of water was unavoidable (3.6 w%), due to the method used for analysis (drying samples at 70°C for 48hr and storage in desiccator). Lange and coworker reported the average biomass composition to be $CH_{1.748}O_{0.596}N_{0.148}$ (ash 3.7 w%). For another strain Roels (1983) reports a value of $CH_{1.82}O_{0.58}N_{0.16}$ (ash 7.3 w%) at a dilution rate of 0.1 L/L/hr of the strain CBS 425.

Stückrath *et al.* (2002) reported elemental compositions of biomass for the strain CEN.PK 113-7D grown on glucose, ethanol and acetate at 0.1 L/L/hr as shown in table 7.1. Stückrath and coworkers pointed out that the glucose-grown cells had a higher concentration of carbohydrates, while cells grown on either ethanol or acetate were richer in lipids and proteins resulting in larger contents of elemental nitrogen on behalf of elemental oxygen.

7.2 Materials and Methods

The cultivation data reported in this chapter has been normalized by dividing the process variables with a known set of normalization parameters. The same set of normalizing variables have been used in all batches. The normalization has been done in order to protect sensitive information.

7.2.1 Strain and Medium

An industrial recombinant *Saccharomyces cerevisiae* strain was grown under aerobic, carbon- limited conditions in a sequence consisting of successive batch, fed-batch and continuous operation. Complex industrial substrates were used, which after sterilization were fed to the bioreactor as a dilute glucose syrup and a complex nutrient substrate containing salts, growth factors and yeast extract.

[3]The ash content being elemental sulfur, phosphor, metals *etc.*

Table 7.1. Elemental compositions of biomass reported by Stückrath *et al.* (2002) for the strain CEN.PK 113-7D, when grown on glucose, ethanol and acetate as limiting substrates

Glucose	$CH_{1.77}O_{0.597}N_{0.149}$ (ash 3.8 w%)
Ethanol	$CH_{1.77}O_{0.547}N_{0.152}$ (ash 4.0 w%)
Acetate	$CH_{1.78}O_{0.545}N_{0.164}$ (ash 3.9 w%)

7.2.2 Equipment

Industrial pilot plant bioreactors with a working volume of approximately 10 L (10 kg) were used for the experiments. The process values were sampled once every minute. The weight of the contents of the bioreactors were monitored and controlled gravimetrically, while the flow rates of substrates, aeration and ammonia were controlled by mass flow controllers. A constant set point of pH in the culture broth was maintained by a control loop adding gaseous ammonia to the aeration stream; the added ammonia also serving as the primary nitrogen source for cellular activities. Constant set points for temperature, aeration rate and stirring rate were used throughout the fed-batch and continuous operation of the experiments. The level of dissolved oxygen of the broth in the bioreactor was measured by an oxygen electrode, which was calibrated with air before each cultivation at nominal cultivation conditions.

7.2.3 Analytical Methods

In order to provide the means for analyzing possible changes in the state of the cultivation, it is necessary to collect extensive information in the form of measurements of the culture broth and the offgas from the bioreactor. Some of these measurements are standard procedures at the pilot plant, while others required the use of analytical equipment (HPLC) at Biocentrum, DTU. Below a description of the analytical methods is used in order to provide the additional information for the interpretation of the process behavior.

7.2.3.1 Biomass Concentration

Samples were taken from the bioreactor (1-3×60 mL). Determination of the biomass concentration was made by centrifugation (5min, 5000rpm) of 10 mL broth in a preweighted heat resistant glass tube, followed by two cycles of washing the pellets in 0.9% g/L NaCl, centrifugation (5min, 5000rpm) and removal of supernatant, before a final drying in an oven for 24-36 hours at 105°C, followed by storage in desiccator for 20-40 min and weighing. Duplicate or triplicate determinations were made with maximal deviation of triplicates within 1%. During sampling between 60-180 mL of culture broth is taken from the bioreactor. This corresponds to under 2 % of the total volume during continuous operation[4] and up to 2.5 % during batch operation.

7.2.3.2 Elemental Analysis

Samples for elemental analysis were prepared by centrifugation (5min, 5000rpm) of 10 mL broth in a preweighted heat resistant glass tube, followed by two cycles of washing the pellets in 0.9% NaCl, centrifugation (5min, 5000rpm), removal of supernant and addition of 10 mL water. The samples were stored at -20°C until

[4]During continuous operation the removal of culture broth to maintain a constant bioreactor weight is carried out by an ejection mechanism that rapidly ejects approximately 50 g broth as the bioreactor weight exceeds the set point value with approximately 100 g. For more details, refer to section 9.1.11.

analysis. The samples were thawed in cold water, centrifuged and pellets were freeze dried to remove water and stored in dessicator, before analysis in a CHNS-O Elemental Analyzer (EA1108-CHNS-O, Fisons Instruments, Beverly, MA, USA). To secure that the storage in water had not led to breaking of the cells, both the thawed sample and the supernant from the last centrifugation were inspected through a microscope for evidence of disrupted cells, however no indication of this was found.

7.2.3.3 Metabolite Analysis

To investigate the variations in the extracellular concentrations of a number of metabolites samples for HPLC were taken from the bioreactor and rapidly filtered within 5 seconds after sampling through a sterile filter with a pore size of 0.45 μm, and cooled to 0°C within 20 sec after sampling from the bioreactor, and then stored at -20°C until further analysis. During this sampling, metabolites are being consumed by cells present in the sample up until filtering *i.e.* for 5 seconds. The amount of glucose that can be consumed by the cells during this period can be estimated by considering the overall glucose uptake rate in the bioreactor, assuming that the residual glucose concentration in the bioreactor is negligible compared to the amounts fed. In this case an approximate value of the normalized glucose consumption rate is 1 mg/L/s, so up to 5 mg/L glucose is consumed in the time it takes from sampling to filtering of the sample. It is assumed that the uptake rates of acetate and ethanol are comparable to the glucose uptake rate. Concentrations of glucose, glycerol, ethanol and acetate were measured by HPLC using an Aminex HPX- 87H column from Bio-Rad for separation. The column was kept at 60°C and eluted with 5mM H $_2SO_4$. Glucose, glycerol and ethanol were detected by a refractometer, while acetate was detected by UV-spectrometry, both sensors being positioned at the exit of the column. No quantification of the measurement uncertainty has been performed. A few selected samples were furthermore analyzed using Boehringer Mannheim kits for acetate and glycerol.

7.2.3.4 Ammonium Concentration

The concentration of ammonium in the culture broth has not been measured, but measurements from similar experiments suggest that a constant value of 0.088 g/kg can be assumed during steady state continuous operation.

7.2.3.5 Off-gas Analysis

Mass spectroscopy on the off-gas from the bioreactor provided measurements of ethanol concentration c_e^g, change in percent O_2 (OXC, inlet: dry air) and change in percent CO_2 (CXC) between inlet and outlet streams. Measurements were conducted every 10 minutes, and the data was subsampled to a frequency of 1 min $^{-1}$ by using constant intervals between the original sampling points. The bioreactor is not equipped with a condenser and therefore water evaporates with the off-gas. Duboc and von Stockar (1998) has investigated this phenomena and found that at $T = 30$°C and 0.63 vvm the water vapor content $y_{w,out}$ is 4.2%. Using this value and assuming that the ethanol concentration in the off-gas is negligible, the following estimation of the off-gas flow rate $F_{air,out}$ is used, which is based on a nitrogen (N_2)

flux balance and suggested by Duboc and von Stockar (1998):

$$F_{air,out} = F_{air}\frac{1-(y_{CO_2,in}+y_{O_2,in})}{1-(y_{CO_2,out}+y_{O_2,out}+y_{w,out})} \tag{7.1}$$

where F_{air} is the aeration rate measured at the inlet, $y_{CO_2,in}$ and $y_{O_2,in}$ are mole fractions in the aeration stream of CO_2 and O_2, respectively, while $y_{CO_2,out}$ and $y_{O_2,out}$ are the mole fractions in the off-gas. Assuming that $y_{CO_2,in} = 0$ and $y_{O_2,in} = 0.21$, one obtains that $y_{CO_2,out} = CXC/100$ and $y_{O_2,out} = y_{O_2,in} - OXC/100 = 0.21 - OXC/100$.

The molare oxygen uptake rate (OUR) can then be calculated as:

$$OUR = \frac{OXC}{100}\cdot F_{air,out}\cdot\frac{p}{RT}\,[mole/hr] \tag{7.2}$$

and molar carbon evolution rate (CER) calculated as:

$$CER = \frac{CXC}{100}\cdot F_{air,out}\cdot\frac{p}{RT}\,[mole/hr] \tag{7.3}$$

where p is the pressure, R is the gas constant and T is the temperature. Since the aeration rate is measured as NL/min, $p = 1$ atm and $T = 273$ K.

7.2.4 Substrates

In the conducted experiments two substrate feeds have been used. The primary substrate, an industrial grade glucose syrup, contained the carbon source, glucose, while the secondary substrate contained other nutrients *e.g.* salts, vitamins and minerals. The indicators *cdos* and *ndos* will be used for the primary and secondary substrates, respectively. During continuous operation the substrates are fed in a mass flow ratio of F^m_{cdos}:F^m_{ndos} = 1:2.

The main components in the secondary substrate were phosphate and sulfate salts, an antifoam agent and growth factors *e.g.* vitamins. Furthermore an industrial grade yeast extract (hydrolyzed yeast cells) was used. The major components are: $MgSO_4$, KH_2PO_4, K_2SO_4, citric acid and yeast extract, and the components were dissolved in water and sterilized before storage in sterile holding tanks.

7.2.5 Cultivation Conditions

The process investigated in this work was a scaled-down industrial production process carried out in bioreactors with several m^3 of culture broth. The production process was carried out as a sequence of operations in the same bioreactor: batch, fed-batch and continuous process. During both fed-batch and continuous operation the two substrates were added in a fixed ratio of mass flow rates as described above. During the fed-batch operation the volume of the culture broth was increased approximately 2.5 times before continuous operation was initiated. Due to the construction of the pilot plant bioreactors the production process could not be scaled-down with respect to the feeding strategy during the fed-batch operation, since it was required that the minimum volume should be above 6 L in order to

cover the pH and DOT electrodes. To overcome this constraint only glucose syrup (*cdos*) was fed to the bioreactor during fed-batch operation, while the pre-calculated amount of nutrient substrate (*ndos*) that should be fed during the fed-batch operation, was instead added to the bioreactor prior to the inoculation of the bioreactor.

The cultivations were carried out in bioreactors with a 10 L working volume. The amount of culture broth in the bioreactor was measured gravimetrically. A constant pH value was maintained by automatic addition of gaseous ammonia to the inlet air stream. Both temperature, speed of stirrer N= 800 rpm, back pressure of $p_{t,g}$ = 0.8 atm (gauge) and aeration rate F_{air}= 18 NL/min were kept constant during cultivation. The dissolved oxygen tension was measured online and for standard cultivations the dissolved oxygen tension was above 20 % of air saturation. Table 7.2 provides an overview of experiments presented in this chapter.

7.2.5.1 Preparation of Inoculum

A vial containing colonies of the recombinant *S. cerevisiae* strain was taken from -80°C storage and transferred to a shake flask containing 200 mL growth medium with glucose as the main substrate. The shake flask was placed in an incubator at 200 rpm and 30°C for 90-120 hours, before the whole content of the shake flask was transferred to a bioreactor.

7.2.5.2 Batch Operation

Prior to the inoculation of the bioreactor glucose and nutrient substrates were charged to the bioreactor. During the first stage of the batch operation the glu-

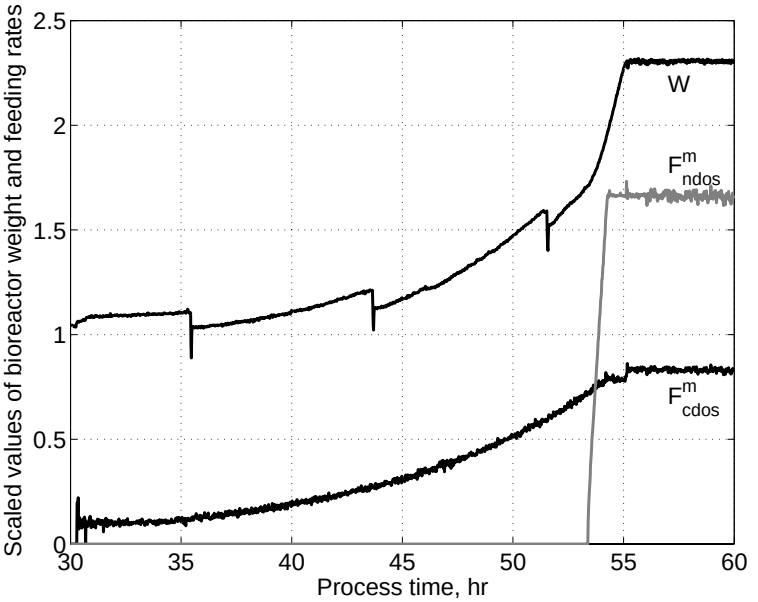


Figure 7.1. Feeding profiles of glucose syrup (black line, F^m_{cdos}), nutrient substrate (grey line, F^m_{ndos}) and the profile of the bioreactor weight (black line, W) during fed-batch and the beginning of continuous operation, where the effects of sampling can be observed as upsets in the weight signal. Continuous operation is initiated as the set point of the bioreactor weight is achieved at 55.1 hours. Data from cultivation MTS05.

Table 7.2. Overview of presented experiments. Except for cultivation MTV03, continuous operation has been carried out without upsets to the dilution rate. In the experiments with prefix *MTS*, small perturbations have been added to the standard dilution rate of 1 kg/kg/hr. Online data in a 25 hours time span is assumed to be representative for the conditions during continuous operation. Abbreviations: *approx*: approximate, *conc.*: concentration, *cont.*: continuous, *det.*: determined, *D*: dilution rate, *dist.*: disturbances, *norm.*: normalized, *op*: operation

Cultivation	MTS05	MTS06	MTS07	MTS11
Batch operation (hr)	0-30	0-30	0-30	0-31
Fed-batch operation (hr)	30-55	30-55	30-54	30-54
Continuous operation (hr)	55-120	55-121	54-120	54-122
Approx. norm. D in continuous op. (kg/kg/hr)	0.98	1.01	0.99	1.01
Number of samples	3	3	3	3
Total number of det. biomass conc.	6	6	6	6
Sampling times (hr)	74.3	74.5	74.3	77.2
	98.0	98.2	98.9	103.3
	119.0	119.2	120.1	121.7
Online data sampled (hr)	90-115	90-115	90-115	90-115

Cultivation	MTS12	MTV02	MTV03
Batch operation (hr)	0-30	0-29	0-29
Fed-batch operation (hr)	30-55	29-57	29-55
Continuous operation (hr)	55-123	57-120	55-122
Approx. norm. D in continuous op. (kg/kg/hr)	1.00	0.99	0.99
Number of samples	3	2	2
Total number of det. biomass conc.	6	6	6
Sampling times (hr)	77.2	76.5	-
	103.3	98.7	100.4
	121.7	-	121.1
Online data sampled (hr)	90-115	90-115	90-115

Table 7.3. Feeding strategies during different operations. W_{cdos} and W_{ndos} are the accumulated weights of glucose syrup and nutrient substrate respectively, fed to the bioreactor during batch and fed-batch operation. The ratio $W_{cdos} : W_{ndos} = 1{:}2$. In the batch phase the substrates are charged prior to inoculation, while during the fed-batch operation a continuous feeding profile of glucose syrup is applied as illustrated in figure 7.1.

Operation	*cdos*	*ndos*
Batch	1/10 W_{cdos}	W_{ndos}
Fed-batch	9/10 W_{cdos}	0
Continuous (ratio)	1	2

cose was consumed forming biomass and ethanol, while in the second stage ethanol was consumed to form more biomass. As the ethanol was exhausted from the culture broth, evaluated using gas analysis on the off-gas, the fed-batch operation was initiated by continuous feeding glucose syrup.

7.2.5.3 Fed-batch and Continuous Operation

As described above only glucose syrup was fed to the process during fed-batch operation. As the ratio between the accumulated amounts of added glucose syrup and nutrient substrate in both batch and fed-batch operations became 1:2, the feeding of the nutrient substrate was initiated in the ratio 2:1 of the glucose syrup feed rate. When the desired filling of the bioreactor for continuous operation was obtained, the effluent stream was opened and continuous operation initiated.

The feeding strategies of the batch, fed-batch and continuous operations are listed in table 7.3, while an example of an applied feeding profile during fed-batch operation is shown in figure 7.1. Notice that the upsets in the weight profile shown in figure 7.1 is due to sampling of the culture broth.

7.2.5.4 Cultivation MTV03

A couple of hours into continuous operation in cultivation MTV03 (at 56 hours), large amounts of acetate and ethanol were being observed and corrective actions were taken in order to get the system back into purely aerobic growth by lowering the dilution rate and then slowly increasing the normalized dilution rate back to the desired normalized steady state value of 1 kg/kg/hr at 68 hours. At 79 hours a series of step changes to the dilution rate was performed, bringing the system into oxido- reductive growth for 2.5 hours. From 85 hours and for the remainder of the cultivation a constant normalized dilution rate of 1 kg/kg/hr was applied.

7.3 Methodology

Investigating the elemental composition of biomass can be addressed in a number of ways depending on the level of detail that is needed or desired. Despite the fact that frequent measurements of the composition of the macromolecular pool undoubtly would provide relevant and important information for process improvement and optimization, this is not carried out in the process studied in this work. Instead the elemental composition of the biomass is deducted from macroscopic elemental balances and validated by elemental analysis of two samples.

7.3.1 Macroscopic Elemental Balancing

The biomass composition can be deduced through the use of macroscopic elemental balances for carbon, nitrogen, hydrogen and oxygen combined with the degree of reduction, κ. During aerobic growth the major contributors to the elemental and degree of reduction balances are: glucose, ammonia, biomass, carbon dioxide, oxygen and water as illustrated with the following black box model:

$$CH_{ax}O_{bx}N_{cx} + Y_{xc}CO_2 + Y_{xw}H_2O - Y_{xs}CH_2O - Y_{xo}O_2 - Y_{xn}NH_3 = 0 \qquad (7.4)$$

where Y_{xc}, Y_{xw}, Y_{xs}, Y_{xn} and Y_{xo} are the yield coefficients of carbon dioxide, water, glucose, ammonia and oxygen on biomass respectively, while ax, bx and cx are the mole contents of hydrogen, oxygen and nitrogen respectively per C-mole biomass. Using equation 7.4, a matrix E can be constructed based on elemental balances of C,N and the degree of reduction κ-balance:

$$E = \begin{bmatrix} C \\ N \\ \kappa \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ cx & 0 & 0 & 0 & 1 \\ \kappa_x & 0 & -4 & 4 & 0 \end{bmatrix} \tag{7.5}$$

where the degree of reduction of the biomass is calculated from the degree of reduction of the involved elements: $\kappa_x = 4{+}ax{-}2bx{-}3cx$. The columns represent steady state volumetric conversion rates (r)[5] in moles of biomass (C-mole), carbon dioxide (C-mole), oxygen (O_2-mole), glucose (C-mole) and ammonia (N-mole) and the steady state component balances are written as:

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = E \times \begin{bmatrix} r_x \\ r_c \\ r_o \\ r_s \\ r_n \end{bmatrix} \tag{7.6}$$

The identification of the parameters ax, bx and cx can not be carried out before the molecular weight per C-mole biomass (M_{DW}) is available. Therefore this parameter needs to be determined together with the three other parameters: ax, bx and cx.

The identification problem is solved as a least squares nonlinear minimization problem using the Levenberg-Marquardt algorithm of MatLab with the objective function:

$$J = \min_{\theta}(f(\theta))^2 \quad \text{with: } \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix} = \begin{bmatrix} ax \\ bx \\ cx \\ M_{DW} \end{bmatrix} \tag{7.7}$$

$$\text{where: } f(\theta) = E \cdot \begin{bmatrix} r_x^m \\ r_c \\ r_o \\ r_s \\ r_n \end{bmatrix} = \begin{bmatrix} 1/\theta_4 & 1 & 0 & 1 & 0 \\ \theta_3/\theta_4 & 0 & 0 & 0 & 1 \\ (4+\theta_1-2\theta_2-3\theta_3)/\theta_4 & 0 & -4 & 4 & 0 \end{bmatrix} \cdot \begin{bmatrix} r_x^m \\ r_c \\ r_o \\ r_s \\ r_n \end{bmatrix}$$

r_x^m is the volumetric conversion rate of biomass based on mass rather than C-moles. From the estimated biomass composition the residual content[6] can be calculated as:

$$res = 1 - \frac{M_{DW,-res}}{M_{DW}} = 1 - \frac{12 + ax + 16bx + 14cx}{M_{DW}} \tag{7.8}$$

where $M_{DW,-res}$ is the molecular weight per C-mole biomass calculated on the basis of the elemental composition of carbon, hydrogen, oxygen and nitrogen alone.

[5]The rates are positive when a component is being produced and negative when consumed.

[6]The reason for using the term residual content rather than ash content is explained in section 7.3.1.1.

The conversion rate of glucose (r_s), biomass (r_x^m) and ammonia (r_n) are calculated as:

$$r_s = D(c_{g,f}^m - c_{g,b}^m)/M_{glu} \tag{7.9}$$

$$r_x^m = Dc_x^m \tag{7.10}$$

$$r_n = F_{NH_3}/W - (Dc_{n,b}^m)/M_{NH_3} \tag{7.11}$$

where $c_{g,f}^m$ is the mass concentration of glucose in the total feed[7], $c_{g,b}^m$ is the mass concentration of glucose in the bioreactor, c_x^m is the mass concentration of biomass in the bioreactor, c_n^m is the mass concentration of ammonia in the bioreactor, while F_{NH_3} is the molar flow rate of ammonia. W is the weight of the bioreactor contents and D is the true dilution rate in kg/kg/hr; M_{glu} and M_{NH_3} are the molecular weights of glucose (per C-mole) and ammonia, respectively.

7.3.1.1 Ash Content

The biomass is composed of other elements than carbon, hydrogen, oxygen and nitrogen, *e.g.* phosphor, sulfur, potassium, magnesium and other trace metals (Roels, 1983), however these elements are only present in very small amounts compared to carbon, hydrogen, oxygen and nitrogen. To account for these elements in mass balances, the parameter *ash content* is introduced referring to the residual from an elemental analysis. The ash content varies between 3-13 w% of the biomass investigated for the laboratory strain *S. cerevisiae* CBS 426 under aerobic glucose-limited cultivation for dilution rates between 0.073 and 0.259 according to Roels (1983), while Herwig *et al.* (2001) report the ash content to vary between 6-8 w% in aerobic glucose-limited cultivations of *S. cerevisiae* CEN.PK 113-7D for dilution rates ranging between 0.04 to 0.40 L/L/hr. Lange and Heijnen (2001) pointed out that water contributed with approximately 3.6 w% of the biomass for samples dried at high temperatures (70°C for 48hr) followed by storage in a dessicator and subsequent elemental analysis. From the data provided by Lange and coworker, an approximate ash content of 3.8 w% can be calculated (based on the mass fraction listed in table VII in Lange and Heijnen (2001)).

Although freeze drying has been used in the preparation of samples for elemental analysis rather than drying at high temperatures as used by Lange and coworkers, both sample preparation methods use storage in dessicator prior to analysis (see section 7.2). Therefore the data obtained from elemental analysis is assumed to contain a contribution from water content of 3.6 w%, which needs to be corrected for before calculating the biomass composition.

In the macroscopic mass balance the water content does not directly influence any of the balances used for calculating the composition, since hydrogen and oxygen contents are determined by the κ-balance in equation 7.5. However the water content will influence the measurement of the biomass concentration, since the method applied uses high temperature drying and storage in dessicator (see section 7.2). The term residual content (f_r) is therefore used to account for the joint contributions from both ash and water contents:

$$(1 - f_r) = (1 - f_a) \cdot (1 - f_w) \tag{7.12}$$

[7] In the combined *cdos* and *ndos* in a 1:2 ratio.

where f_a and f_w are the fractions of ash and water respectively.

In the macroscopic mass balance the residual content is estimated using equation 7.8 and combined with equation 7.12 the ash content in percent is calculated as:

$$f_a = 1 - \frac{1 - f_r}{1 - f_w} \tag{7.13}$$

For the results obtained by elemental analysis both the ash content and the water content are assumed to be constant at 3.8 w% and 3.6 w%, respectively. This yields a residual content of:

$$(1 - f_r) = (1 - f_{a,EA}) \cdot (1 - f_w) = (1 - 0.038) \cdot (1 - 0.036) \Rightarrow res = 7.3w\% \tag{7.14}$$

where f_w is the fraction of water, while $f_{a,EA}$ signifies the ash fraction used in calculations based on the elemental analysis.

7.3.2 Elemental Analysis

Two samples for elemental analysis were taken in the continuous part of cultivation MTS12. These samples were analyzed for elemental carbon, hydrogen and nitrogen using a CHNS-O-analyzer. Assuming an ash content of 3.8 w% and a water content of 3.6% the elemental oxygen composition can be calculated based on the residuals when the other components, ashes and water is accounted for.

7.4 Results

7.4.1 Cultivation Results

Average values of measurements obtained during a number of continuous cultivations are listed in table 7.4, all of which have been normalized using a fixed set of normalization variables. As mentioned above the residual contents is assumed to be 7.3 w% when considering the results obtained by elemental analysis, while a water content of 3.6 w% is assumed in case of the macroscopic mass balances, since the applied method provides an estimate of the residual content. It is assumed that the most important contributions to the elemental balances are accounted for and it can be seen that the contributions of ethanol, acetate, succinate and pyruvate are negligible compared to glucose. Furthermore the concentration of glucose in the feed has not been determined explicitly for all of the cultivations. The glucose concentration in the feed is similar in all of the cultivations, evaluated by inspection of the ratio between CER and glucose syrup flow rate. The average of this ratio only differs with less than 1 % in the cultivations listed in table 7.4. The listed normalized value of 37 g/kg is determined from the substrates used in cultivation MTV03. The use of yeast extract in the nutrient substrate has an impact on the nitrogen balance of the process, since it is known that the yeast extract provides amino acids for the growth processes of the yeast. It is assumed that the contribution to the overall nitrogen balance is negligible.

Table 7.4. Averages of measured concentrations and flow rates during pseudo-steady state conditions from 90 to 115 hours in continuous operation. Where available, standard deviations are listed in parenthesis below the mean values. Data has been normalized, using the same value except for the dilution rate.

	Unit	MTS05	MTS06	MTS07	MTS11	MTS12	MTV02	MTV03
Component								
Biomass	g/kg	20	20	19	19	19	20	20
		(0.38)	(0.46)	(0.46)	(0.51)	(0.27)	(0.53)	(0.07)
Glucose (feed)	g/kg	37	37	37	37	37	37	37
		-	-	-	-	-	-	-
Ammonia (feed)	mole/hr	0.11	0.11	0.11	0.11	0.11	0.11	0.11
		(0.0010)	(0.0008)	(0.0010)	(0.0007)	(0.0009)	(0.0005)	(0.0008)
Glucose (broth)	g/kg	0.037	0.042	0.024	0.036	0.038	0.039	0.042
		(0.0030)	(0.0021)	(0.0010)	(0.0037)	(0.0012)	(0.0008)	(0.0022)
Ethanol	g/kg	0	0	0	0	0	0	0
		0	0	0	0	0	0	0
Acetate	g/kg	0.045	0.049	0.009	0.047	0.049	0.026	0.015
		(0.0072)	(0.0064)	(0.0008)	(0.0079)	(0.0072)	(0.0200)	(0.0174)
Succinate	g/kg	0.052	0.048	0.150	0.119	0.136	0.050	0.048
		(0.0037)	(0.0035)	(0.0323)	(0.0409)	(0.0019)	(0.0006)	(0.0009)
Pyruvate	g/kg	0.0029	0.0025	0.0034	0.0032	0.0034	0.0044	0.0036
		(0.00030)	(0.00015)	(0.00018)	(0.00043)	(0.00020)	(0.00015)	(0.00010)
CER	mole/hr	0.48	0.50	0.49	0.50	0.49	0.49	0.48
		(0.0027)	(0.0023)	(0.0028)	(0.0022)	(0.0029)	(0.0007)	(0.0020)
OUR	mole/hr	0.44	0.47	0.45	0.45	0.45	0.44	0.44
		(0.0025)	(0.0022)	(0.0026)	(0.0021)	(0.0027)	(0.0007)	(0.0018)
Ammonia (broth)	g/kg	0.088	0.088	0.088	0.088	0.088	0.088	0.088
		-	-	-	-	-	-	-
Dilution rate	kg/kg/hr	0.980	1.010	0.990	1.010	1.000	0.990	0.990
		(0.032)	(0.036)	(0.049)	(0.042)	(0.054)	(0.012)	(0.009)

7.4.2 Elemental Analysis

Results from analyzing two samples from the continuous operation of cultivation MTS12 are listed in table 7.5.

7.4.3 Elemental Composition

Based on the results shown in table 7.4 it can be concluded that the most important components to the overall mass balance during continuous operation appear in equation 7.4. Using equation 7.7 and 7.8 the parameters are estimated using all 7 cultivation data sets available:

$$\theta = \begin{bmatrix} ax \\ bx \\ cx \\ M_{DW} \end{bmatrix} = \begin{bmatrix} 1.82 \\ 0.583 \\ 0.142 \\ 26.9 \end{bmatrix} \tag{7.15}$$

Table 7.5. Elemental composition measured by elemental analysis (w%) from two samples from cultivation MTS12. The data has not been corrected for ash and water contents. $a-c$ refers to repeated measurements. *Rest* is the amount not accounted for by carbon, hydrogen or nitrogen.

Sample	Carbon	Hydrogen	Nitrogen	Rest
MTS12p17a	44.03	7.11	7.52	41.34
MTS12p17b	44.59	7.27	7.60	40.54
MTS12p17c	44.31	7.13	7.54	41.02
MTS12p18a	43.88	7.00	7.49	41.63
MTS12p18b	44.20	7.03	7.55	41.22

which leads to the following biomass composition:

$$CH_{1.82}O_{0.583}N_{0.142} \quad \kappa_x = 4.23; \quad M_{DW} = 26.9 \text{ g/C-mole } (f_r\text{: } 6.7 \text{ w\%}) \tag{7.16}$$

and the ash content can be calculated by use of equation 7.13; the ash content is 3.1 w%.

An analysis of how well the three balances, C, N and κ from equation 7.5 close, when using the identified parameters of θ is shown in table 7.6. Here the individual deviations in the three balances are calculated for each experiment and listed as a percentage of the total value. From the table it can be seen that the standard deviations are small for all three balances with most variation present in the N-balance.

Table 7.6. Relative deviations in C-, N- and κ-balance for each experiment, along with average values and standard deviations. C-balance: %dev=$(\hat{r}_x+r_s+r_c)/|r_s|$; N-balance: %dev=$(\hat{cx}\hat{r}_x+r_n)/|r_n|$; κ-balance: $(\hat{\kappa}_x-4r_o+4r_s)/\hat{\kappa}_x$. A negative value indicates that the use of the estimated parameters from equation 7.15 leads to a balance closing below 100%. Notice that the values were not obtained under stationary continuous steady state conditions, but rather as average values during continuous operation with perturbations to the feeding rate.

Experiment	C-dev. (%)	N-dev (%)	κ-dev (%)
MTS05	1.73	4.43	0.42
MTS06	3.11	-0.81	1.17
MTS07	-0.08	-0.30	-0.11
MTS11	-0.74	-1.92	-0.21
MTS12	-1.54	-3.09	-0.36
MTV02	-0.25	1.38	-0.18
MTV03	-2.37	0.37	-0.63
Average	-0.02	0.01	0.01
St.dev.	1.88	2.44	0.60

The data listed in table 7.5 obtained by elemental analysis on samples from cultivation MTS12 results in an elemental composition of the biomass as listed in table 7.7. Unfortunately it was not possible to analyze more samples and other batches with this methodology.

Table 7.7. Calculated elemental composition of biomass ($CH_{ax}O_{bx}N_{cx}$) from data obtained using elemental analysis. For the calculations an ash content of 3.8% has been assumed as well as a water content of 3.6 w%. Data listed in table 7.5, along with degree of reduction (κ_x) and the molecular weight (M_{DW} in g/C-mole) at a residual content of 7.3 w% (see equation 7.14). The average is calculated as $0.5 \cdot (mean(p17) + mean(p18))$, where $mean(p17)$ signifies the mean value from sample p17. Standard deviations (St.dev) are calculated based on the average value.

Sample	ax	bx	cx	κ_x	M_{DW}
MTS12p17a	1.83	0.580	0.146	4.23	27.1
MTS12p17b	1.85	0.560	0.146	4.29	26.8
MTS12p17c	1.82	0.571	0.146	4.24	27.0
MTS12p18a	1.80	0.587	0.146	4.19	27.2
MTS12p18b	1.80	0.576	0.146	4.21	27.0
Average	1.82	0.576	0.146	4.23	27.1
St. dev.	0.0100	0.0052	0.00012	0.0193	0.084
St. dev. (%)	0.548	0.91	0.081	0.458	0.309

The data in table 7.7 suggest that the elemental composition of biomass should be:

$$CH_{1.82}O_{0.576}N_{0.146} \quad \kappa_x = 4.23; \quad M_{DW} = 27.1 \text{ g/C-mole } (f_r\text{: } 7.3 \text{ w\%}) \qquad (7.17)$$

7.5 Discussion

7.5.1 Comparing Biomass Compositions

The two methods of determining the biomass composition provides almost identical results. The small discrepancies in the amounts of elemental oxygen and hydrogen are negligible. It is however interesting to notice that all 5 of the measurements in table 7.7 show an elemental composition of nitrogen of 0.146 with a standard deviation of less than 0.1%. All of the samples analyzed using elemental analysis originates from cultivation MTS12. A number of explanations for the discrepancies in the estimated nitrogen composition will be discussed here.

Examining the deviations of the carbon, nitrogen and the degree of reduction balances from cultivation MTS12 in table 7.6, it is interesting to notice how a relatively large amount of nitrogen is unaccounted for (-3.1 %) in the nitrogen balance. The ammonia concentration in the culture broth has been assumed to have a constant value of 0.088 g/kg in all experiments, which corresponds approximately to 4% of the total ammonia balance, so violation of this assumption could explain the observed discrepancy. In the other experiments, other large discrepancies are seen in the nitrogen content and it is therefore concluded that the assumption of a constant ammonia concentration in the culture broth is likely to be untrue since the ammonia concentration in the broth has a large influence on the determination of the nitrogen content of the biomass by macroscopic mass balances.

An alternative explanation of the lack of fit of the nitrogen balance in cultivation MTS12 can be obtained by reestimating the nitrogen content of biomass (cx) solely on the data from cultivation MTS12, while keeping all other parameters constant.

This leads to a value of $cx = 0.142/0.97 = 0.146$ an interesting observation, since this corresponds to the average value of the nitrogen content in biomass estimated using elemental analysis data, where only samples from cultivation MTS12 has been analyzed. Although the biomass compositions in equations 7.16 and 7.17 appear to have different nitrogen compositions, this might actually be as a result of the samples analyzed rather than differences due to the use of different algorithms. With the data at hand it is not possible to investigate these issues further and more investigations are needed for clarification.

Table 7.8. Identified elemental compositions of biomass using macroscopic mass balancing (MMB) and elemental analysis (EA)

Method	Elemental Composition	κ_x	M_{DW}	f_r
MMB	$CH_{1.82}O_{0.583}N_{0.142}$	4.23	26.9 g/C-mole	6.7 w%
EA	$CH_{1.82}O_{0.576}N_{0.146}$	4.23	27.1 g/C-mole	7.3 w%

From the comparison of the results obtained in equations 7.16 and 7.17 and tables 7.6 and 7.7 (summarized in table 7.8), it appears that the estimation of biomass composition obtained using the results from the elemental analysis is more accurate than the one estimated using macroscopic mass balances. It is however still interesting to note how close the two estimates are, suggesting that in other applications, a good estimate of the biomass composition can be obtained using macroscopic mass balances.

7.5.2 Effects From Ash Content

Although only small discrepancies between the two identified biomass compositions have been found, it is still relevant to test the sensitivity of the results with respect to the assumed ash content, hence the sensitivity of assuming the residual content to be 7.3 w% in the results obtained from elemental analysis is discussed here.

The results for varying amounts of ash content are used on the data in table 7.5 for ash contents varying between 2 to 6 w%. The results are listed in table 7.9. Since the ratio between carbon, hydrogen and nitrogen is determined by the measurements in table 7.5, it is only the amounts of elemental oxygen that varies, having a direct effect on the degree of reduction balance and therefore κ_x.

From table 7.9 it can be seen that assuming that the estimated value of the ash content obtained using macroscopic mass balancing of 3.1 w% does not lead to a better agreement between the estimated biomass compositions of the two approaches. It is therefore concluded that the an ash content of 3.8 w% is reasonable.

7.5.3 Selected Values from the Literature

The estimated compositions of biomass of this work are in accordance with the results presented in the open literature. Lange and Heijnen (2001), providing the most elaborate study of the biomass composition of the strain CEN.PK 113-7D, arrive at an average biomass composition of $CH_{1.748}O_{0.596}N_{0.148}$ (f_r: 7.3 w%), while Herwig *et al.* (2001) presents an average composition of: $CH_{1.75}O_{0.50}N_{0.16}$ (f_r: 7.2 w%) for the same strain. The difference in oxygen and nitrogen contents by the

Table 7.9. Elemental biomass composition ($CH_{ax}O_{bx}N_{cx}$) determined by elemental analysis as a function of the ash content including the degree of reduction (κ_x) and the molecular weight (M_{DW} in g/C-mole). Data based on elemental analysis results listed in table 7.5

Ash content w%	ax	bx	cx	κ_x	M_{DW}
2.0	1.82	0.606	0.146	4.19	27.6
3.0	1.82	0.589	0.146	4.21	27.3
3.8	1.82	0.576	0.146	4.23	27.1
4.0	1.82	0.573	0.146	4.23	27.0
5.0	1.82	0.557	0.146	4.25	26.7
6.0	1.82	0.540	0.146	4.28	26.4

two groups using the same strain is probably due to different cultivation conditions, especially the range of dilution rates.

Comparing these results with the results obtained in this work indicates that the hydrogen and nitrogen contents are a little different from the ones reported by Lange and coworkers and Herwig and coworker. A possible explanation for the discrepancy in nitrogen content could be the dilution rates investigated, as both research groups concluded that nitrogen content was dependent on dilution rates. A second explanation for discrepancies could be that the cultivations in this study were not in steady state. The cultivations with prefix *MTS* were perturbed in the glucose syrup feed rate is discussed in section 8.1.8, while neither cultivation MTV02 nor MTV03 had spent the recommended 5-7 residence times in continuous steady state conditions (Lange and Heijnen, 2001). A further explanation could be that the observed differences are due to the use of different strains. The one used in this work is a recombinant strain and this could have an effect on the macromolecular composition of the strain compared to the laboratory research strain CEN.PK 113-7D.

Further experiments are needed to elucidate the reason for the small observed discrepancy in the nitrogen content. It is also interesting to elucidate the changes to the biomass composition brought about by changing process conditions as the operations progress from batch through fed-batch to continuous operation. Based on the biomass compositions reported by Stückrath *et al.* (2002) and listed in table 7.1 in the introduction to this chapter, it appears that during growth on ethanol and acetate, the nitrogen content of the biomass is larger than during growth on glucose, whereas the oxygen content is lower, explained by a large content of carbohydrates in cells grown on glucose at the expense of lipids and proteins. In the light of these observations it would be interesting to investigate the changes to the biomass concentration through the different operating conditions to evaluate how and how fast the cells adapt to the changing conditions.

7.6 Conclusion

Using two different approaches, elemental analysis and macroscopic mass balancing, the investigations in this chapter have attempted to determine the elemental composition of a recombinant strain of *S. cerevisiae* used for production of an insulin

precursor during aerobic glucose-limited growth. Inspired by the work of Lange and Heijnen (2001) it was assumed that 3.6 w% of the biomass concentration measured as dry weight was due to water, and an ash content of 3.8 w% accounting for the contributions of metals, sulfur and phosphor to the biomass composition. The contribution of water and ashes is referred to as the residual (see equation 7.14). With these assumptions the following estimates of the biomass composition was obtained[8]:

$$\text{MMB: } CH_{1.82}O_{0.583}N_{0.142} \quad \kappa_x = 4.23; \quad M_{DW} = 26.9 \text{ g/C-mole } (f_r\text{: } 6.7 \text{ w\%}) \tag{7.18}$$

$$\text{EA: } CH_{1.82}O_{0.576}N_{0.146} \quad \kappa_x = 4.23; \quad M_{DW} = 27.1 \text{ g/C-mole } (f_r\text{: } 7.3 \text{ w\%}) \tag{7.19}$$

Based on the results obtained and the discussions presented above, the following elemental composition of the recombinant strain of *S. cerevisiae* is suggested based on the methodology using elemental analysis:

$$CH_{1.82}O_{0.576}N_{0.146} \quad \kappa_x = 4.23; \quad M_{DW} = 27.1 \text{ g/C-mole } (f_r\text{: } 7.3 \text{ w\%}) \tag{7.20}$$

The estimate of the elemental biomass composition can now be used in dynamic modeling for a cultivation process using the investigated strain. This is the topic of the following two chapters where the information provided in this chapter will serve an important role in the construction of models and identification of model parameters.

[8] MMB: macroscopic mass balancing. EA: elemental analysis.

8

Indicator for Onset of Oxidoreductive Growth

A modelbased indicator has been constructed to monitor the onset of oxidoreductive growth of an industrial recombinant strain of Saccharomyces cerevisiae for production of an insulin precursor. Using a macroscopic proton flux balance over the bioreactor and assuming a constant pH of the culture broth, an estimate of the ammonia flow rate demand during aerobic growth was compared to the measured ammonia flow rate, providing a residual that could indicate the onset of oxidoreductive growth. Inputs to the model were effects from changes in the volume of the culture broth, effluent flow rate and glucose syrup feed rate. Parameters were identified from a number of operating conditions in three batches. During normal operating conditions of oxidative growth the estimated ammonia demand was within 10% of the measured addition rate of ammonia, and the onset of process upsets and oxidoreductive growth could easily be identified as the difference between estimated and measured values of the ammonia addition rate exceeded a detection limit of 1.5 mmole NH_3/mole biomass/hr.

In chapter 6 a soft sensor for online estimation of biomass concentration was presented for a production of insulin precursor, however using a slightly different recombinant strain than in this chapter. The soft sensor was based on the assumption that the ammonia flow rate could be used for estimation of the biomass production rate and concentration. The same approach was attempted using a different strain of *Saccharomyces cerevisiae*, however surprising results were obtained. In a number of cultivations conducted, a peculiar behavior of the ammonia flow rate just prior to the onset of oxidoreductive growth[1], $F^m_{NH_3}$ has been observed in the late part of fed-batch phases, indicating that a number of effects influence the proton flux balance of the system. In some cases an elevated flow rate of ammonia has been followed by ethanol production 1-2 hours later, an example of this behavior is shown in figure 8.1. In other cases HPLC analysis of the culture broth during periods of peculiar behavior of the ammonia flow rate have indicated the presence of acetate. Both ethanol and acetate are metabolites formed during oxidoreductive growth where the fermentative pathway is active (see section 3.4.5). The presence of these metabolites is unwanted during the fed-batch and continuous operating regimes in the production of the insulin precursor, since metabolic energy is redirected from the two main purposes: 1) obtaining and maintaining a high concentration of biomass to support

[1]Superscript m indicates mass flow rate.

2) a high productivity of the insulin precursor.

It is highly relevant to develop observers and indicators that can alert operators and/or the process control system, when the metabolic activities in the process are about to change from oxidative to oxidoreductive growth. The purpose of this chapter is to develop a framework, that can provide an indicator for the onset of oxidoreductive growth. A model will be constructed that will be able to estimate the necessary ammonia demand for neutralization of different effects on the proton flux balance during purely oxidative growth. For this purpose the impact of different effects on the proton flux balance will be quantified. The residual between estimated and measured ammonia flow rate can then be used for further investigations with the aim of elucidating what metabolic response may be responsible for the behavior seen in figure 8.1. As the model is based on the assumption of oxidative growth without formation of ethanol or acetate, the residual can be used to evaluate whether or not this assumption is valid or not *i.e.* as an indicator of the state of the metabolism.

This chapter consists of a section describing the data used for model identification followed by a section on the applied strategy adapted for modeling. After this, a section presents the model identification along with the parameter estimation, before the identified model is validated using cultivation data.

8.1 Materials and Methods

8.1.1 Normalized Data

The cultivation data reported in this chapter has been normalized in order to blur sensitive information. The requirement to do so makes it rather difficult to compare the findings reported in this work with results from the open literature. Furthermore it can be difficult to check and verify calculations and results since normalization has to be used in a

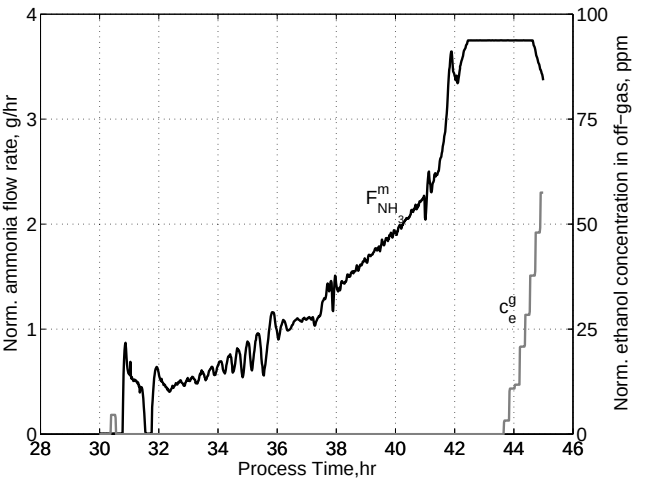


Figure 8.1. Profiles of normalized ammonia flow rate ($F^m_{NH_3}$, black line, left hand axis) and ethanol concentration in the offgas (c^g_{eth}, grey line, right hand axis) in cultivation MTF04. Up until 42.5 hours the culture broth pH does not deviate more than 0.02 pH-units from its set point value. At 42.5 hours the maximum ammonia flow rate is reached and the pH starts to decrease below the set point value (not shown here).

way that also blurs ratios *e.g.* dilution rates and specific productivities. The author apologies for the inconvenience that the normalization can cause and for the distortion it might have on understanding the issues addressed in this thesis.

8.1.2 Strain and Medium

Details on this topic are given in section 7.2.1. The biomass composition was investigated in chapter 7 and estimated to: $CH_{1.82}O_{0.576}N_{0.146}$ (f_r: 7.3 w%).

8.1.3 Equipment

Details on this topic are given in section 7.2.2.

A number of process parameters and variables are measured and data series of these measurements are available with a sampling frequency, ω_s, of 1 min^{-1}. These variables are listed in table 8.1.

Table 8.1. Variables available as online measurements at a sampling frequency $\omega_s = 1$ min^{-1}. CER and OUR are presented in section 7.2.3.5. Note that CXC, OXC and c_e^g are sampled every minute, however the mass spectrometer only provides new values every 10 minutes.

Variable	Unit	Description
t	[hr]	Elapsed time from start of cultivation
F_{air}	[NL/min]	Aeration rate of bioreactor using sterilized air
$F_{NH_3}^m$	[g/hr]	Ammonia dosed to the bioreactor by injection into air flow
F_{cdos}^m	[g/hr]	Feed flow rate of primary substrate: glucose syrup
F_{ndos}^m	[g/hr]	Feed flow rate of secondary substrate: nutrients
pH	[-]	Acidity of culture broth measured online
DOT	[%]	Dissolved oxygen tension; percentage of saturation
$p_{t,g}$	[bar]	Back pressure at the top of the bioreactor (gauge)
N	[rpm]	Speed of stirrer
CXC	[%]	Difference between offgas and air inlet concentration of CO_2
OXC	[%]	Difference between air inlet and offgas concentration of O_2
c_e^g	[ppm]	Concentration of ethanol in off-gas
RQ	[-]	Respiratory quotient: RQ = CER/OUR
θ	[$^\circ C$]	Temperature of culture broth inside tank
W	[kg]	Estimated weight of culture broth

8.1.4 Analytical Methods

Details on this topic are given in section 7.2.3.

8.1.5 Substrates

Details are given in section 7.2.4.

8.1.6 Cultivation Conditions

Details are given in section 7.2.5. Table 8.2 provides an overview of experiments presented in this chapter.

8.1.7 Preparation of Inoculum

Details on this topic is found in section 7.2.5.1.

8.1.8 Perturbations

In the cultivations with the prefix *MTS* small amplitude perturbations were introduced to the nominal substrate feed rates during both fed-batch and continuous operation to provide more dynamic data for modeling. The feeding profile during the fed-batch phase was composed of ramps approximating an exponentially increasing feeding profile, while in continuous operation a constant feed rate was used. A grid of nodes was spanned along the combined feeding profiles, with a node for every 2 hours. Between every node linear ramps of feed rates were used to create the nominal feeding profile. Perturbations were introduced to the nominal feeding profile by multiplying the feeding rates at the nodes by a factor f and then reconstructing the feeding profile by using piecewise linear ramps between the new values at the nodes[2].

$$f = (1 + p_f) \quad p_f \in \mathcal{N}(0, \sigma^2) \text{ where } \sigma = 5\% \text{ of nominal dilution rate} \tag{8.1}$$

[2]This form of simple piecewise linear perturbations was chosen, since the process control system could only handle 16 points in the construction of the dosing profile.

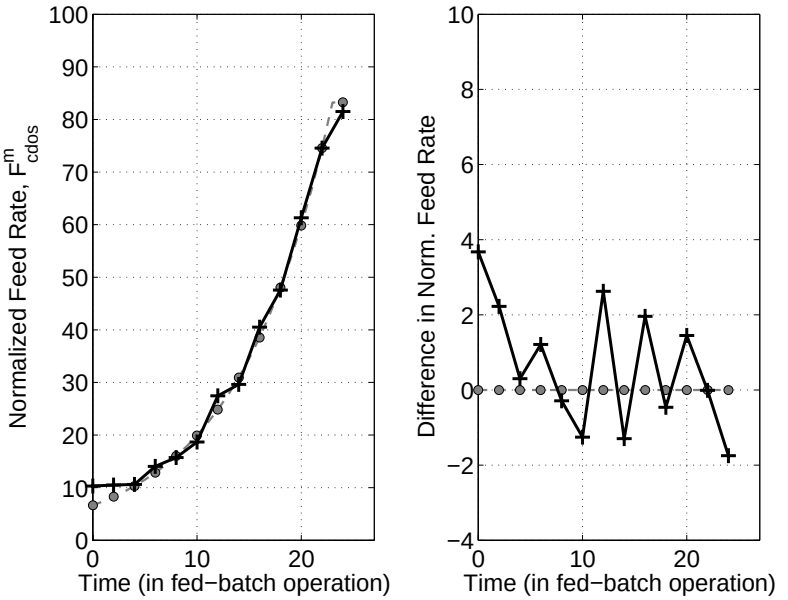


Figure 8.2. Left: Exponential feeding profile (grey dashed line) with grid points (grey filled circles) and perturbed feeding profile (black line, +). Right: Difference between exponential feeding profile (grey dashed line, grey filled circles) and perturbed feeding profile (black line, +).

Table 8.2. Overview of presented experiments. A more elaborate description of the entries marked by a $\star$ can be found in the text. Standard cultivations are experiments were no disturbances (planned or unplanned) where observed during the fed-batch and continuous operation. The control of substrate feed rate is carried out as a preprogrammed feeding profile by a series of set points with linear ramps in between (*Recipe*). Abbreviations: *op*: operation, *D*: dilution rate, *approx*: approximate, *norm*: normalized.

	MTS01	**MTS05**	**MTS07**	**MTS12**
Standard cultivation	No	Yes	Yes	Yes
Perturbations in fed-batch and continuous operation*	Yes	Yes	Yes	Yes
Planned disturbance*	Ramp in D	No	No	No
Unplanned disturbance (type)	Ethanol	No	No	No
Batch operation (hr)	0-26	0-30	0-30	0-30
Fed-batch operation (hr)	26-51	30-55	30-54	30-55
Continuous operation (hr)	51-144	55-120	54-120	55-123
Approx. norm. D in continuous operation (kg/kg/hr)	-	0.98	0.99	1.00
Control of substrate feed rate	Recipe	Recipe	Recipe	Recipe
Results displayed in	Fig. 8.13	Fig. 8.12	Fig. 8.10	Fig. 8.11
Data used for (Modeling/Validation)	Validation	Validation	Modeling	Validation

	MTE01	**MTV01**	**MTV02**
Standard cultivation	No	No	Yes
Perturbations in fed-batch and continuous operation*	No	No	No
Planned disturbance*	Yes	No	No
Unplanned disturbance (type)	No	Yes	No
Batch operation (hr)	0-30	0-28	0-28
Fed-batch operation (hr)	30-51 64-84	28-52	29-57
Continuous operation (hr)	84-120	52-123	57-120
Approx. norm. D in continuous operation (kg/kg/hr)	1.00	0.98	1.00
Control of substrate feed rate	Recipe	Recipe	Recipe
Results displayed in	Fig. 8.7	Fig. 8.15	Fig. 8.8
Data used for (Modeling/Validation)	Modeling	Validation	Modeling

An example of perturbations added to a nominal feeding profile is illustrated in figure 8.2 were also a lower constraint on the minimum feed rate at a normalized value of 10 g/hr has been added[3].

8.1.9 Modeling Batches; Feeding Strategies

For the modeling of the flow rate of ammonia, data from cultivations MTE01, MTV02 and MTS07 were used. The course of these cultivations are described in the following.

MTV02: This cultivation was carried out using the normal feeding strategy. Prior to the batch phase nutrient substrate containing growth factors, vitamins and yeast extract, had been added to the bioreactor in sufficient amounts to cover both the batch and fed-batch phases. During the fed-batch phase from 29 to 54.5 hours only glucose syrup was fed to the bioreactor. Feeding of nutrient substrate was initiated as the weight of the bioreactor contents exceeded 8.8 kg, bringing the process into continuous operation at 56.3 hours. The fed-batch phase of this cultivation was rather conservative, where the specific growth rate dropped below the nominal feeding profile value halfway for 27 hours of fed-batch operation, and averaged 90 % of the nominal feeding profile value for the second half of the fed-batch operation. The subsequent continuous operation was conducted without any incidents for more than 60 hours. Figure 8.3(a+b) show the pH and substrate feeding profiles of cultivation MTV02.

MTS07: During continuous operation of cultivation MTS07, perturbations in the feeding rates were planned and carried out in order to investigate the response of the system to such disturbances. The fed-batch phase was conducted without any incidents at a normalized specfic growth rate of 0.9-1.0. During continuous operation, perturbations in the substrate feeding were introduced, which resulted in piecewise linear changes in the nominal feeding profile value between 90 - 110 %. Figure 8.3(c+d) show the pH and substrate feeding profiles of cultivation MTS07.

MTE01: In this cultivation two types of fed-batch operation were performed. The first of these followed the normal feeding strategy, but as the desired filling of the bioreactor was achieved, the process was stopped. In this experiment a second strategy for fed-batch operation was to be investigated, which was the reason for terminating feeding before entering into continuous operation. To prepare for the second fed-batch phase, 5/6 of the bioreactor contents were removed and 3 kg of nutrient substrate was fed to the bioreactor[4]. The bioreactor was then left aerated and stirred for 11 hours, during this batch period the pH rose to 8.3. Low amounts of amino acids are present in the broth due to the addition of yeast extract and these might be consumed by the biomass. This phenomenon has not been further

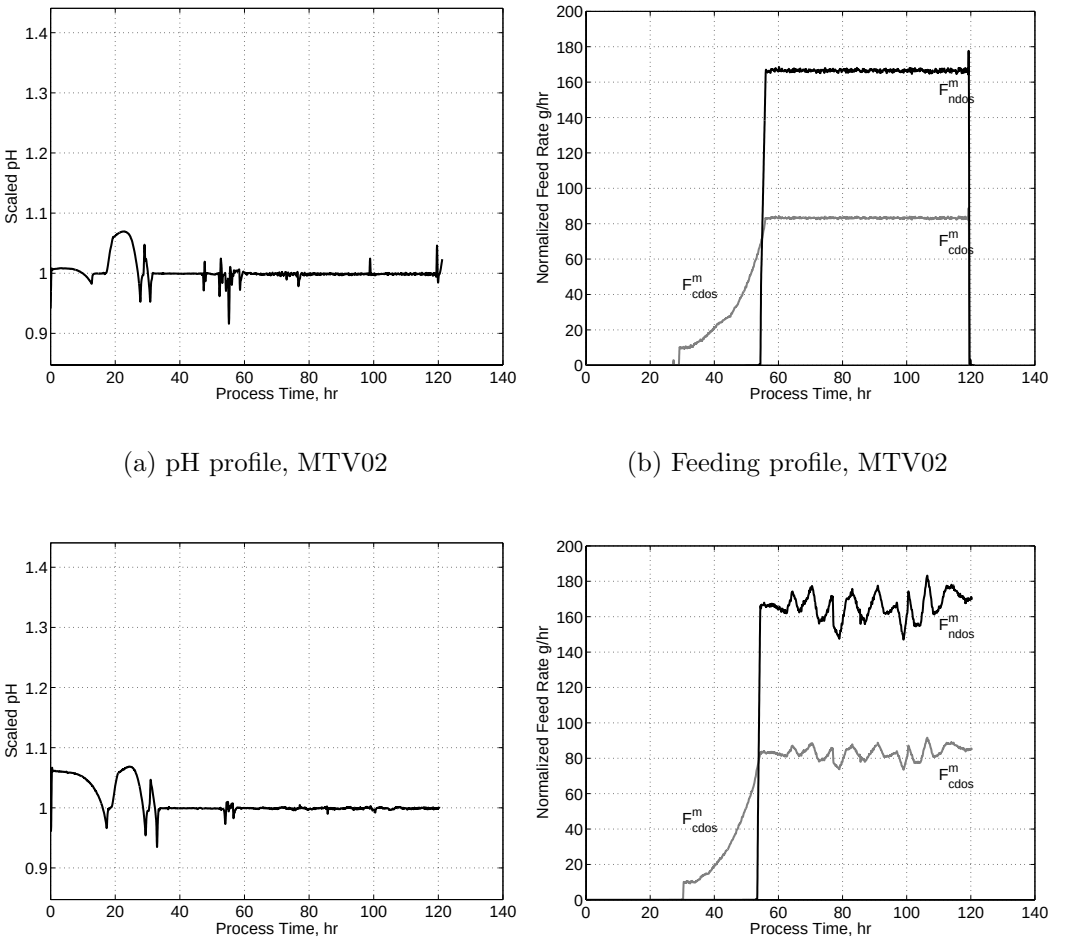
[3]The lower calibration limit of the mass flow controller.

[4]The reason for the design of this experiment was to investigate the fed-batch operation with a lower initial biomass concentration.

studied[5]. From 63.5 hours the second fed-batch operation was carried out using both glucose syrup and nutrient substrate, in the 1:2 relation normally used during continuous operation. The fed-batch phase lasted until 84.5 hours, after which continuous operation was initiated. The feed rates reached their maxima at 86.5 hours after which they were lowered to normal continuous operation levels. A larger filling level (weight) had been used as set point in this second fed-batch phase. This was readjusted to the normal level before the experiment was ended by 30 hours of continuous operation.

Figure 8.4(a+b) show the profiles of pH and ammonia flow rate. Excursions in pH are seen in the batch phase (0-30 hours) as well as between the two fed-batch phases (50 to 64 hours). The drops in pH around 80 hours are effects from synchronization of growth resulting in oscillations. Figure 8.4(c) show the flow rates of the two substrates, where the difference between the two fed-batch operations can be seen; no F^m_{ndos} used for the first fed-batch operation. The apparent specific

[5]At this low level of tank filling the pH sensor is barely under the surface of the broth and the stirring is far from effective due to the positioning of the impellers.



(a) pH profile, MTV02

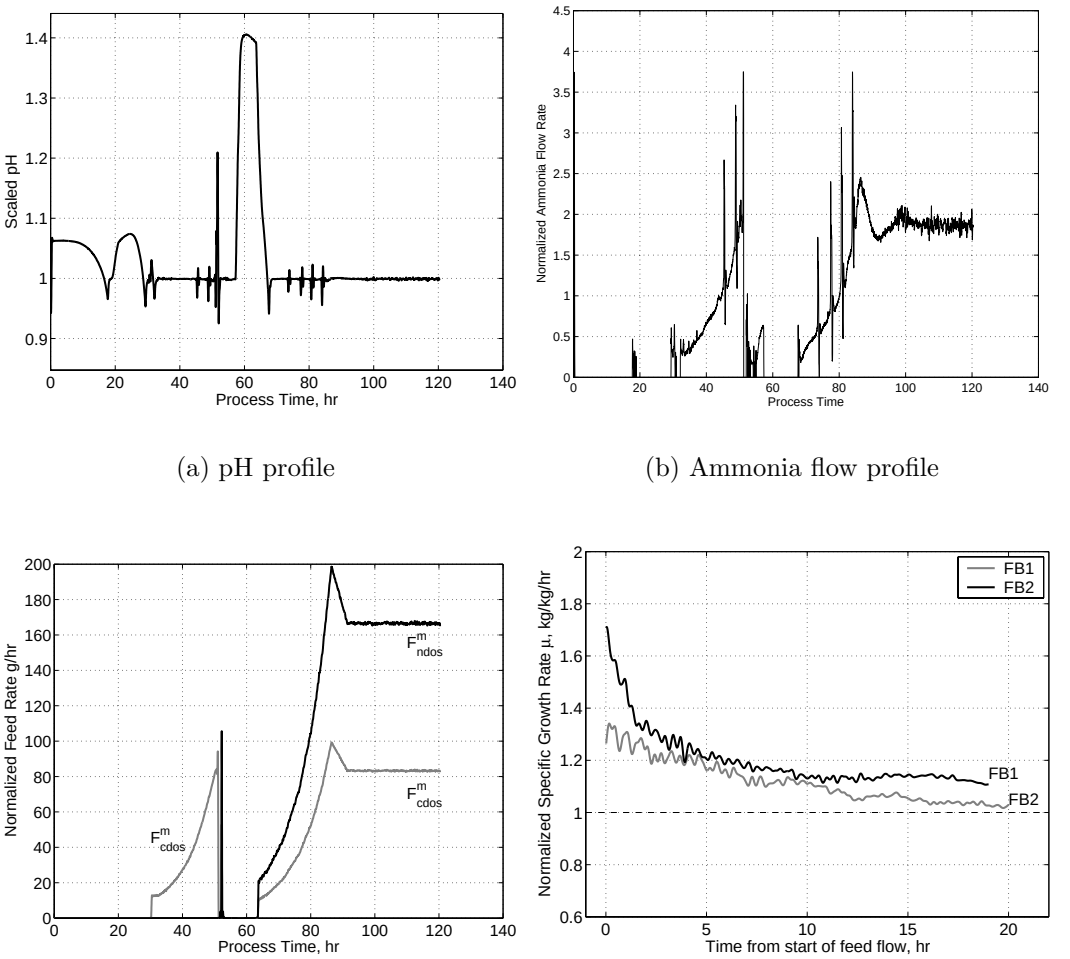
(b) Feeding profile, MTV02

(c) pH profile, MTS07

(d) Feeding profile, MTS07

Figure 8.3. pH and normalized feeding profiles of glucose syrup (F^m_{cdos}) and nutrient substrate (F^m_{ndos}) from cultivation MTV02 (top row) and MTS07 (bottom row).

growth rates (μ) for the two fed-batch operations have been calculated based on the assumptions of balanced growth, constant yield of biomass on glucose (Y_{sx}) and a constant normalized glucose concentration of 110 g/kg in the glucose syrup. Figure 8.4(d) show the normalized growth rates being above 100 % of the nominal feeding profile value throughout the two fed-batch phases and ending between 100 and 110 %.



(a) pH profile

(b) Ammonia flow profile

(c) Feeding profile

(d) Growth rate profile

Figure 8.4. Profiles of (a) pH, (b) normalized ammonia flow rate, (c) normalized substrate feeding rates and (d) calculated normalized growth rate during the two fed batch operations in cultivation MTE01.

8.1.10 Filtering

Filtering of signals can be used in order to remove unwanted process dynamics or sensor noise from interesting and relevant process information. In this work filtering has been used in order to remove high-frequency information from the process data signals. This has been done since the process dynamics that are in focus have a characteristic time constant of approximately 10 min and above. Even faster process dynamics are also highly interesting, however these have not been considered here.

A fourth order Butterworth filter (see equation 8.2) with a cut-off frequency of 0.15 $\cdot\omega_s$ (ω_s is the sampling frequency) is used to filter the originally sampled data signals.

To avoid an introduction of a time delay in the data series the signals are filtered both in forward and backward (reverse) directions; the latter is conducted by reversing the data sequence and applying the same Butterworth filter again.

The Butterworth filter is formulated as:

$$G_f(s) = \frac{\omega_B^2}{s^2 + 2\zeta_1\omega_B s + \omega_B^2} \cdot \frac{\omega_B^2}{s^2 + 2\zeta_2\omega_B s + \omega_B^2} \tag{8.2}$$

where the filter parameters were chosen to be: $\zeta_1 = 0.38$, $\zeta_2 = 0.92$ and $\omega_B = 0.3$ ω_s (Åström and Wittenmark, 1995).

8.1.11 Estimating Effluent Flow Rate and Rate of Weight Change

The details on this subject can be found in section 9.1.11. The Kalman Filter algorithm is used for the estimation of the effluent flow rate F_{out} and rate of weight change $\frac{\Delta W}{\Delta t}$.

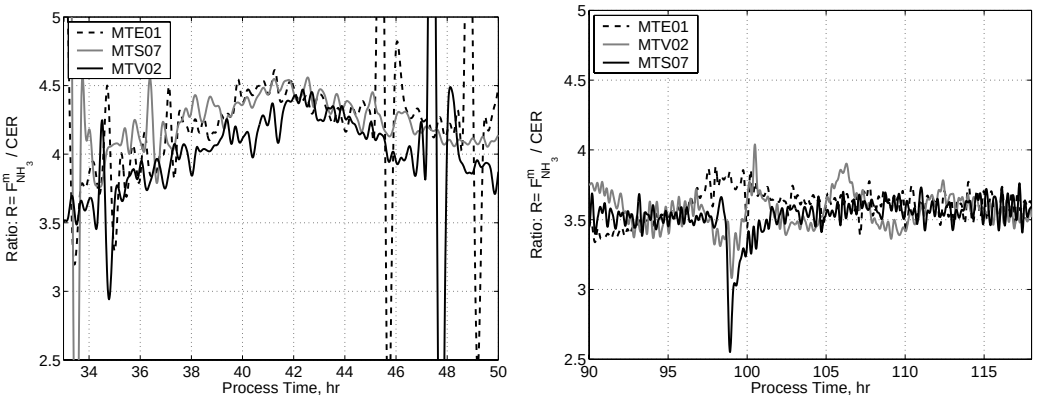
8.2 Modeling Strategy and Construction

As was mentioned in the introduction to this chapter the purpose of this investigation is to construct a model, that will be able to estimate the necessary ammonia demand for neutralization of different effects on the proton flux balance during purely oxidative growth. This is needed in order to use the residual between the estimated and measured ammonia flow rate to evaluate whether the growth is oxidative or oxido-reductive. With such tools at hand it will be possible to construct other models *e.g.* for the estimation of the biomass concentration.

This section will describe how the strategy behind the model is developed, based on observations from process data. Three cultivations are used for model construction and parameter estimation: MTV02, MTS07 and MTE01. These cultivations are rather different, since different feeding strategies were used. Therefore they contained a lot of information on how variations in different variables influence the proton flux balance. Thereafter a section on model construction and parameter estimation follows, where also the influence from possible contributions to the proton flux balance are discussed and evaluated.

The model for the ammonia flow rate, $F^m_{NH_3}$, is developed from a proton balance, where a controller is assumed to vary $F^m_{NH_3}$ to maintain a constant pH. Data from the fed-batch phases of the three modeling cultivations are used together with data from continuous operation to model the effects of the two substrates, the effluent flow and the buffer capacity of the cultivation broth on pH and subsequently on $F^m_{NH_3}$ used to maintain a constant pH. By inspection of the data it was found that in the initial part of the fed-batch phase the ratio between the ammonia flow rate and carbon emission rate (CER) is not the same as in the stationary part of the continuous operation. This is demonstrated in figure 8.5. Both $F^m_{NH_3}$ and CER

reflect cellular activity and during balanced growth the ratio of these two signals is expected to be constant as is seen during continuous operation. This is not the case in the fed-batch phase, where a maximum around 40 hours can be seen. The ratio is in general larger than during the continuous phase. A number of effects influence the ratio between $F^m_{NH_3}$ and CER. The presence of buffer pairs in the substrates fed to the system will lead to an additional flow of ammonia to maintain the desired pH. Dilution and effluent flow will decrease the need for ammonia by removal of protons. These effects will be discussed and included in the modeling.



(a) Fed-batch operation (b) Continuous operation

Figure 8.5. Ratio between $F^m_{NH_3}$ and CER. Left hand figure (a) shows a window of the fed-batch phase, while the right hand figure (b) shows a window of the continuous operation.

Table 8.3 summarizes the data used for model identification. It can be seen that not all the data available in the cultivations has been used. This is because many of the flow rates are highly correlated leading to an ill-conditioned identification problem.

Table 8.3. Time frame in hours of data used for modeling

Cultivation	Fed-batch	Continuous
MTE01	70-80	95-100
MTV02	32-47	85-95
MTS07	35-47	51-58

8.2.1 Estimating Buffer Effect of the Substrates

A proton balance for a bioreactor is shown in equation 8.3 and illustrated in figure 8.6:

$$\frac{dV[H^+]}{dt} = V\frac{d[H^+]}{dt} + [H^+]\frac{dV}{dt} = F_{ndos}\cdot[H^+]_n + F_{cdos}\cdot[H^+]_c + F_{H^+} - F_{NH_3} - F_{out}\cdot[H^+]_e \tag{8.3}$$

where V is the working volume of the bioreactor, F_{ndos} is the volumetric flow rate of nutrient substrate, F_{cdos} is the volumetric flow rate of glucose syrup, F_{NH_3} is the

molar flow rate of ammonia, F_{out} is the volumetric effluent flow from the bioreactor and $[H^+]$ is the concentration of protons with subscripts b, c, e and n referring to broth, glucose syrup, effluent and nutrient substrate respectively. Rather than using the concentration of hydrogen ion $[H^+]$, the buffer equivalent, β, will be used where $\beta_n = 1/\rho[H^+]_n$. F_{H^+} is the net molar rate of proton production or consumption in the broth due to cellular activities.

Assuming a constant pH of the broth, and converting to mass flows for substrate and effluent streams, equation 8.3 simplifies to a proton flux balance:

$$\beta_b \frac{dW}{dt} = \beta_n \cdot F^m_{ndos} + \beta_c \cdot F^m_{cdos} + F_{H^+} - F_{NH_3} - \beta_e \cdot F^m_{out} \tag{8.4}$$

During oxidative growth a constant ratio between the proton production rate (F_{H^+}) and the biomass production rate has been reported by Lei (2001) and Castrillo *et al.* (1995). Assuming a constant yield of biomass on glucose, F_{H^+} can be calculated from:

$$F_{H^+} = \alpha V r_x = \alpha \frac{Y_{sx}}{M_{glu}} c^m_{glu,cdos} F^m_{cdos} \tag{8.5}$$

where r_x is the volumetric biomass production rate ([mole/L/hr]), Y_{sx} is the molar yield of biomass on glucose, M_{glu} is the molecular weight of glucose ([g/C-mole]), $c^m_{glu,cdos}$ is the concentration of glucose in the glucose syrup ([g/kg]) and α is the molar ratio of protons per C-mole biomass.

Using the assumptions on F_{H^+} (equation 8.5) and $\frac{dW}{dt}$ calculated using a Kalman Filter as described in section 9.1.11, identification of the parameters in the proton flux balance of equation 8.4 can be performed by reformulating the flux balance to estimate the molar flow rate of ammonia:

$$F_{NH_3} = \beta_n \cdot F^m_{ndos} + (\beta_c + \alpha c^m_{glu,cdos} \frac{Y_{sx}}{M_{glu}}) F^m_{cdos} - \beta_e \cdot F^m_{out} - \beta_b \cdot \frac{\Delta W}{\Delta t} \tag{8.6}$$

which is then formulated as a least squares problem:

$$Y = X \cdot \theta^T \tag{8.7}$$

with

$$\begin{aligned} Y = F_{NH_3} \quad , \quad X = [F^m_{ndos} \, , \quad F^m_{cdos} \, , \quad F^m_{out} \, , \quad \frac{\Delta W}{\Delta t}] \\ \theta = [\beta_n \, , \quad \beta_c + \alpha c^m_{glu,cdos} \frac{Y_{sx}}{M_{glu}} \, , \quad \beta_e \, , \quad \beta_b] \end{aligned} \tag{8.8}$$

The parameters are estimated using least squares estimation. Using normalized values the solution yields the results listed in table 8.4.

From the values of the estimated parameters and their standard deviations it appears that only $[\beta_c + \alpha \cdot c^m_{glu,cdos} \cdot \frac{Y_{sx}}{M_{glu}}]$ is significant. It should be kept in mind however that variables such as F^m_{out}, F^m_{ndos} and $\frac{\Delta W}{\Delta t}$ are needed to explain the behavior during transition between different operational modes *e.g.* transition from fed-batch to continuous operation. Taking these matters into consideration it still appears that the contribution from F^m_{ndos} can be ignored since the standard deviation on the

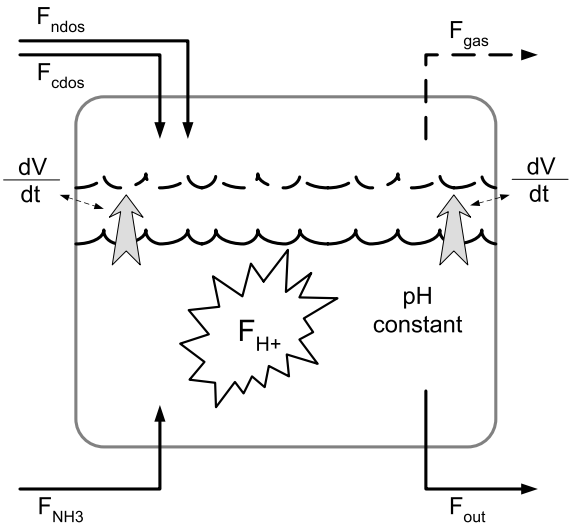


Figure 8.6. Schematic illustration of the contributions to the proton balance of a bioreactor shown in equation 8.3. The pH is assumed to be constant and protons are produced due to cellular activities (F_{H^+}). A contribution to the proton balance from the effluent gas is indicated, however this contribution is assumed to be negligle *i.e.* all ammonia fed to the bioreactor is absorbed by the broth.

Table 8.4. Estimated parameters of the model in equation 8.7 and 8.2.1. The column *St.dev* lists the standard deviation of the estimated parameter in percentage of the parameter value.

Parameter	Estimate	St.dev	Unit
β_n	$7.3 \cdot 10^{-4}$	$\pm$ 230%	mmole NH_3/g Ndos
$[\beta_c + \alpha \cdot c^m_{glu,cdos} \cdot \frac{Y_{sx}}{M_{glu}}]$	$3.3 \cdot 10^{-1}$	$\pm$ 0.76%	mmole NH_3/g Cdos
β_e	$1.7 \cdot 10^{-3}$	$\pm$ 50%	mmole NH_3/g eff
β_b	$1.5 \cdot 10^{-3}$	$\pm$ 60%	mmole NH_3/g broth

parameter is more than 200% and its contribution to the molar flow rate of ammonia evaluated with average values during continuous operation is less than 1%.

Leaving the contribution of F^m_{ndos} out, the reestimation of the parameters yields the results listed in table 8.5.

Table 8.5. Estimated parameters of the model in equation 8.7 and 8.2.1 leaving out the contribution of F^m_{ndos}. The column *St.dev* lists the standard deviation of the estimated parameter in percentage of the parameter value.

Parameter	Estimate	St.dev	Unit
$[\beta_c + \alpha \cdot c^m_{glu,cdos} \cdot \frac{Y_{sx}}{M_{glu}}]$	$3.3 \cdot 10^{-1}$	$\pm$ 0.76%	mmole NH_3/g Cdos
β_e	$1.6 \cdot 10^{-3}$	$\pm$ 45%	mmole NH_3/g eff
β_b	$1.5 \cdot 10^{-3}$	$\pm$ 54%	mmole NH_3/g broth

The pH of the glucose syrup used for the experiments is pH $\approx$ 4, then:

$$\rho\beta_c \approx [H^+]_c < 0.0001M \approx 0.0001\text{mmole } H^+/\text{g Cdos} \tag{8.9}$$

hence the contribution of β_c can be ignored. Therefore it can be assumed that $\alpha \cdot c_{glu,cdos} \cdot Y_{sx}/M_{glu} = 0.33$ mmole NH_3/g Cdos (normalized value), and from this

α can be determined to be:

$$\alpha = \frac{0.33 \text{ mmole } NH_3/\text{g Cdos} \cdot M_{NH_3}}{c^m_{glu,cdos} \frac{Y_{sx}}{M_{glu}}} = 0.024 \text{ g } NH_3/\text{gDW} = 0.15 \frac{\text{mole } NH_3}{\text{C-mole biomass}} \tag{8.10}$$

In chapter 7 the biomass composition was found to be: $CH_{1.82}O_{0.576}N_{0.146}$ (f_r 7.3 w%) and it can be seen that $\alpha = 0.15$ corresponds well to the nitrogen content of the biomass $cx = 0.146$.

The final model for estimation of the molar flow rate of ammonia becomes:

$$F_{NH_3} = \alpha \frac{Y_{sx}}{M_{glu}} c^m_{glu,cdos} F^m_{cdos} - \beta_e \cdot F^m_{out} - \beta_b \cdot \frac{\Delta W}{\Delta t} \tag{8.11}$$

which can be converted to mass flow rate of ammonia using:

$$F^m_{NH_3} = M_{NH_3} \cdot F_{NH_3} = 17 \text{ g } NH_3/\text{mole} \cdot F_{NH_3} \tag{8.12}$$

The assumptions behind the model in equation 8.11 are listed in table 8.6.

Table 8.6. Assumptions behind the model in equation 8.11. The assumptions are listed according to priority with the most important assumption at the top.

1	Asynchronous balanced oxidative growth with constant yield (Y_{sx})
2	Negligible offset or bias in measurements of ammonia flow rate
3	No significant variation in (constant) buffering capacity of the broth
4	No significant effect from feeding of nutrient substrate
5	pH controlled at set point

8.3 Modeling and Validation

Equation 8.11 has been validated using data from different cultivations and the results are shown in the following. Filtering has been used on the data as described in section 8.1.10 and to quantify the precision of the estimations, the root mean square error (RMSE) is presented for selected time periods[6].

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{F}^m_{NH_3,i} - F^m_{NH_3,i})^2} \tag{8.13}$$

N being the number of measurements. RMSE has the same unit as $F^m_{NH_3}$ [g NH_3/hr].

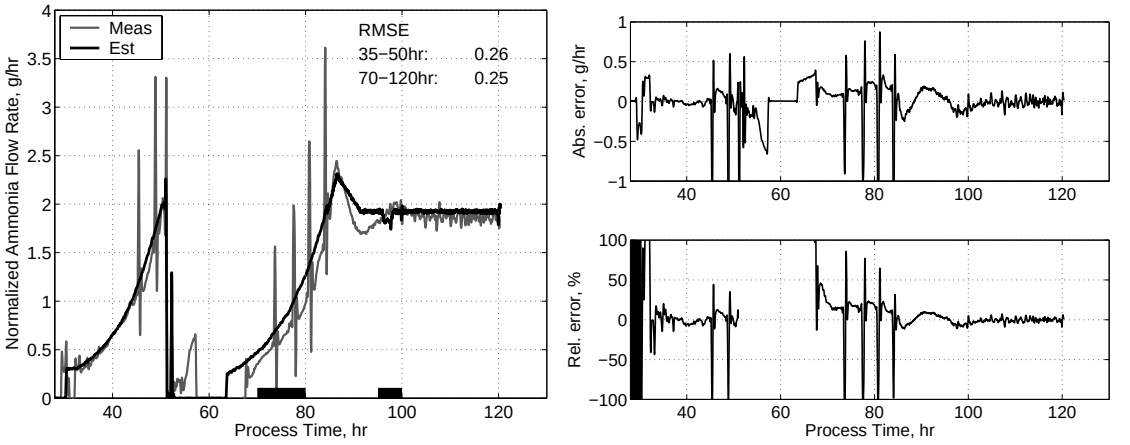
8.3.1 Modeling Batches

Figure 8.7(a) shows how well the ammonia flow rate can be estimated during cultivation MTE01. Although the RMSE is around 10-15% of $F^m_{NH_3}$ during continuous operation, this also includes oscillations and a different fed-batch environment in the second fed-batch phase. Here it is interesting to note that the estimation of the

[6]RMSE is both used on modeling batches and validation batches.

ammonia flow rate is above the measured flow rate for most of the second fed-batch phase from 64 to 84 hours. The first of the fed-batch phases has not been used for parameter estimation and a very good description of the ammonia flow rate is seen. The overall performance of the estimation is rather good, capturing the most important features of the two fed-batch phases and the continuous operation. At the start of the second fed-batch operation, indicated by the signal of the estimated ammonia flow rate rising at 64 hours, it can be observed that no flow of ammonia is measured for the first 4 hours (64 to 68 hours). This is because the pH has risen significantly during the process stop as shown in figure 8.4(a), and therefore no ammonia is needed for pH control. The pH is slowly brought back to the set point value by the acidification of the culture broth due to protons being produced as discussed in section 8.2.1. The bioreactor was not equipped with an inlet for acid, which could have been used to control the pH during the process stop.

Figure 8.7(b) show the absolute (top) and relative (bottom) deviation between estimation and measurement. Closer investigation of these figures reveal that the relative deviation is less than 10 % for most of the time when glucose syrup is being fed.



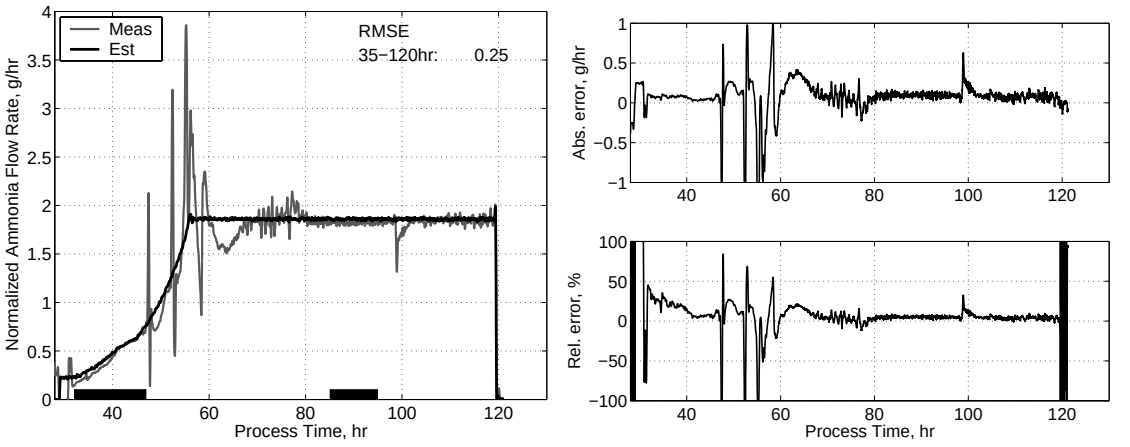
(a) Normalized ammonia flow rate F_{NH_3}. Measured (light) and estimated (dark) values.

(b) Differences between measured and estimated values. Absolute (top) and relative (bottom) values.

Figure 8.7. Results from estimating $F^m_{NH_3}$ in cultivation MTE01. a) Measured (light) and estimated (dark) trajectories of $F^m_{NH_3}$. b) Absolute and relative (1-$\hat{F}^m_{NH_3}/F^m_{NH_3}$) differences between measured and estimated values. Note that only data from the periods 70-78 hours and 95-100 hours of this batch has been used for parameter estimation. This has been indicated by the black bars just above the x-axis. pH and substrate dosing profiles are shown in figure 8.4.

Figure 8.8 show the same plots for cultivation MTV02. Again the fed-batch phase show oscillations in the $F^m_{NH_3}$ signal as seen in figure 8.8(a). Just before the change to continuous operation around 56 hours, a large peak in $F^m_{NH_3}$ can be seen, which is followed by 2-3 additional peaks. The first peak appears as feeding of nutrient substrate (F^m_{ndos}) is initiated, increasing from normalized values of 0 to 170 g/hr over 1.5 hours. From investigating the acidity and buffer capacity of the nutrient

substrate based on its composition, it was not expected that this substrate flow would be responsible for the rapid increase in $F^m_{NH_3}$. This can also be deducted from the signals in the continuous operation, where both high growth activity and a high flow rate of the nutrient substrate is present at the same time. The $F^m_{NH_3}$ during this phase is constant around a normalized value of 1.9 g NH_3/hr as seen in figure 8.8(a).



(a) Normalized ammonia flow rate $F^m_{NH_3}$. Measured (light) and estimated (dark) values.

(b) Differences between measured and estimated values. Absolute (top) and relative (bottom) values.

Figure 8.8. Results from estimating $F^m_{NH_3}$ in cultivation MTV02. a) Measured (light) and estimated (dark) trajectories of $F^m_{NH_3}$. b) Absolute ($\hat{F}^m_{NH_3}$-$F^m_{NH_3}$) and relative ($\hat{F}^m_{NH_3}/F^m_{NH_3}$-1) differences between measured and estimated values. Note that only data from the periods 32-47 hours and 85-95 hours of this batch has been used for parameter estimation. This has been indicated by the black bars just above the x-axis. pH and substrate dosing profiles are shown in figure 8.3.

The characteristic peaks at the change to continuous operation do not suggest it to be part of culture oscillations. This is seen from figure 8.9, where the fluctuations in CER and the ammonia flow rate are shown. Between 55 and 60 hours a number of the fluctuations in the two signals are different. The peaks in this time period are neither due to oscillations nor due to changing feed rates, since the estimation of the ammonia flow in figure 8.8(a) does not show such variations. This indicates that some sort of unmodeled metabolic response triggers the excessive flow of ammonia; the model assumption of balanced growth is violated. Four hours into the continuous phase, as the measured and estimated values of the ammonia flow seems to merge, an eight hour dip (between 60 - 68 hours) is observed in $F^m_{NH_3}$.

At 98.7 a sudden dip in the ammonia addition rate appears as a consequence of adjusting for a difference between in-line measurement of pH and a off-line measurement off pH performed by sampling of culture broth. After 100 hours the F_{NH_3} and $F^m_{NH_3}$ are similar for the remainder of the cultivation. Figure 8.8(b) show how the deviation between the two signals is mostly within a normalized value of 0.1 g NH_3/ hr, corresponding to less than 5% of the measured value $F^m_{NH_3}$.

The last of the batches used for modeling was cultivation MTS07, where primarily

the fed-batch phase was used for parameter identification. Figure 8.10(a) show how well the measured trajectory in the fed-batch phase is estimated. Again a peak of 200% of the estimated flow rate of ammonia appears as the feeding of nutrient substrate is initiated increasing from 0 to 170 g/hr over less than 1 hour; followed by a dip to approximately 50% of the estimated flow rate. Three hours after the initial peak, $F^m_{NH_3}$ seems to settle, however then a seven hour dip (between 57 - 64 hours) is observed, as was also the case of cultivation MTV02. In the continuous phase the perturbation of the feeding rates can be seen to influence the ammonia flow rate in a way which is not explained by the model observed as variations in the absolute error illustrated in figure 8.10(b). It has not been possible to explain these variations as effects from changes in the feeding rates, effluent flow or dilution, and the discrepancies are therefore ascribed to changing metabolic activities, indicating that the system is highly sensitive to rapid changes in the operation conditions. It is noticeable how the RMSE is as low as 0.14 g NH_3/hr even with discrepancies around the onset of continuous operation and the perturbations during continuous operation. The two plots in figure 8.10(b) further highlights how well the model estimates the ammonia flow rate to within 0.1 g NH_3/hr.

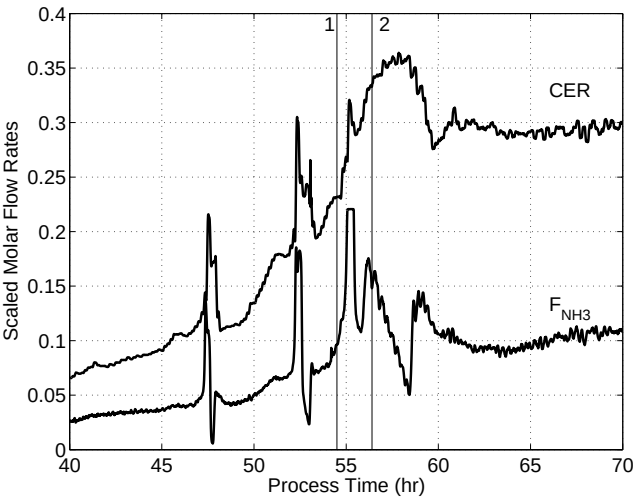
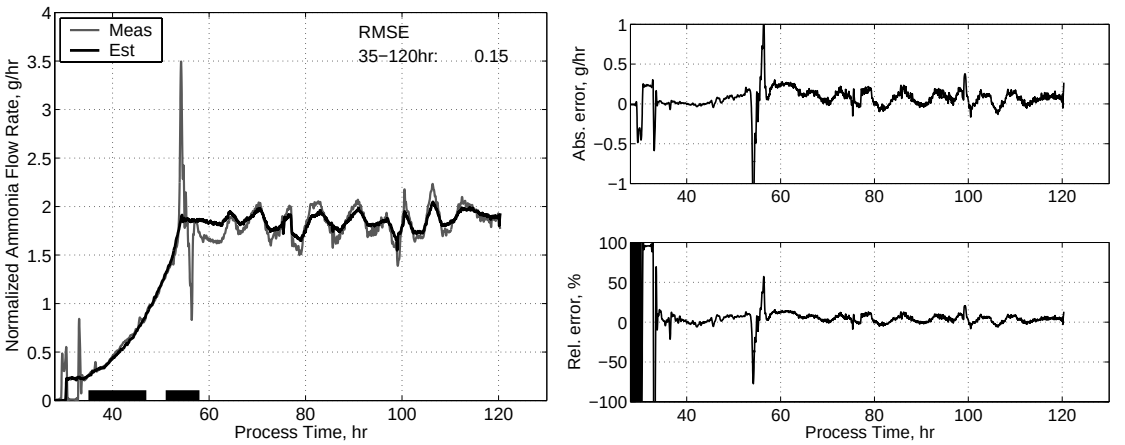


Figure 8.9. Scaled trajectories of normalized carbon emission rate (CER) and ammonia flow rate ($F^m_{NH_3}$) in the time periods 40-70 hours showing oscillations at 47 and 52 hours. Fluctuations are also seen between 55 and 60 hours, but are not believed to be oscillations due to synchronized growth. The vertical lines labelled 1 and 2 indicate the onset of nutrient substrate feeding and transition from fed-batch to continuous operation, respectively Data from cultivation MTV02.



(a) Normalized ammonia flow rate $F^m_{NH_3}$. Measured (light) and estimated (dark) values.

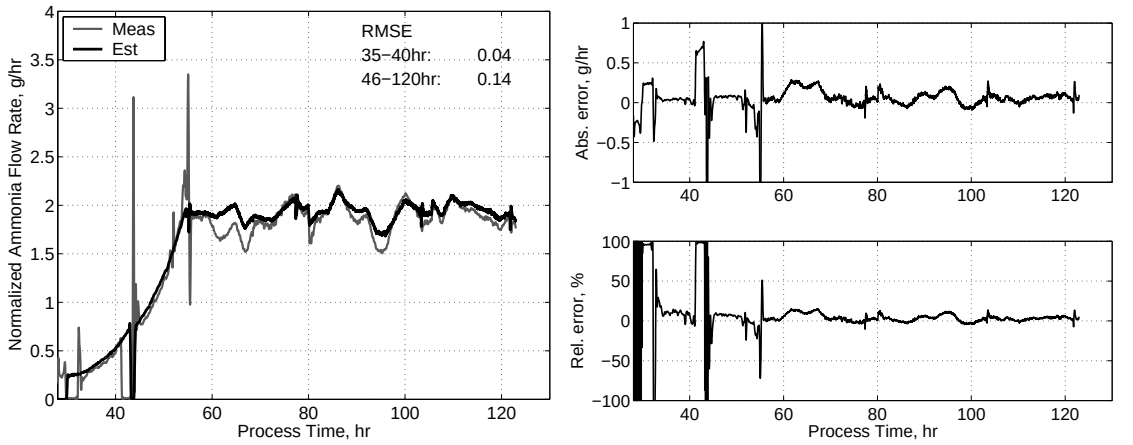
(b) Differences between measured and estimated values. Absolute (top) and relative (bottom) values.

Figure 8.10. Results from estimating $F^m_{NH_3}$ in cultivation MTS07. Note that only data from the periods 35-47 hours and 51-58 hours of this batch has been used for parameter estimation. This has been indicated by the black bars just above the x-axis. See caption for figure 8.8 for calculation of relative values. pH and substrate dosing profiles are shown in figure 8.3.

8.3.2 Validation Batches

To illustrate the performance of the model for estimation of the ammonia flow rate, it has been applied to batches, where data have not been used in the model parameter estimation. Both examples of normal and problematic batches will be demonstrated.

Figure 8.11 shows the result of estimating the ammonia flow rate in cultivation MTS12, which is characterized as normal except for an error in the ammonia supply system leading to a stop in ammonia dosing between 41 and 43.5 hours. The process was back on track at 45 hours. From 50 to 55 hours, where continuous operation is initiated, a small discrepancy between measured and estimated flow rates can be observed. At 53.5 hours the feeding of nutrient substrate is initiated and increases from 0 to 170 g/hr over less than 1 hour. As the system settles into continuous operation the previously observed dip in measured ammonia flow rate in cultivations MTS07 (57-64 hours) and MTV02 (60-68 hours), can again be seen between 60 - 68 hours, however disturbed by a change in dilution rate occurring at the same time. A RMSE of 0.13 g NH_3/hr is similar to the observed for cultivation MTS07.

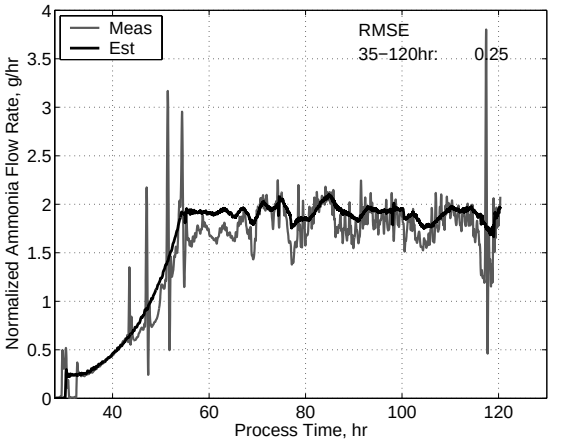


(a) Normalized ammonia flow rate $F^m_{NH_3}$. Measured (light) and estimated (dark) values.

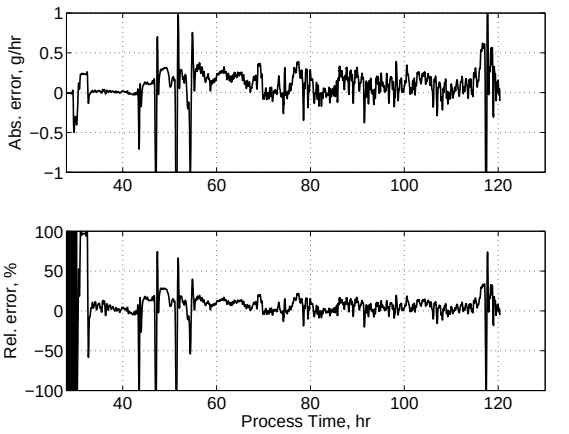
(b) Differences between measured and estimated values. Absolute (top) and relative (bottom) values.

Figure 8.11. Results from estimating $F^m_{NH_3}$ in cultivation MTS12. Validation batch. See caption for figure 8.8 for calculation of relative values.

Figure 8.12 illustrates the results from cultivation MTS05. In the fed-batch phase three oscillations can be seen at 43.5, 47.2 and 51.5 hours. The nutrient substrate feeding is initiated at 53.5 hours and increases from 0 to 170 g/hr over less than 1 hour; a small increase in $F^m_{NH_3}$ is seen. As continuous operation is initiated, the measured ammonia flow rate is lower than the estimated flow rate for ten hours, however more variations in the measured signal is seen in MTS05 than in the other cultivations. For the remaining duration of continuous operation, perturbations in the feeding rates are introduced, which lead to variations and discrepancies as reported previously, however it can also be observed that the noise (or oscillations) in the $F^m_{NH_3}$ signal seems to be larger than observed in MTS07 and MTS12. With these observations in mind the RMSE of 0.22 g NH_3/hr indicates a reasonable estimation of the ammonia flow rates.

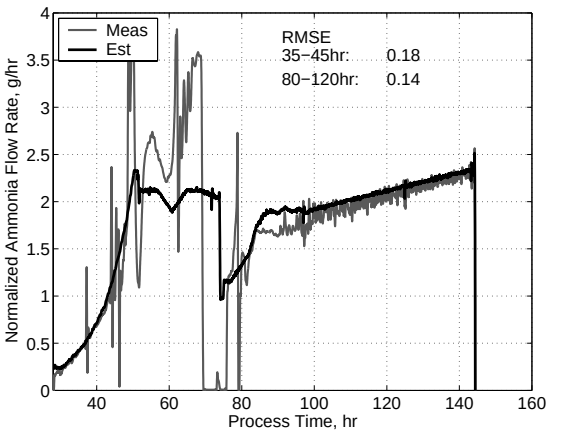


(a) Normalized ammonia flow rate $F^m_{NH_3}$. Measured (light) and estimated (dark) values.

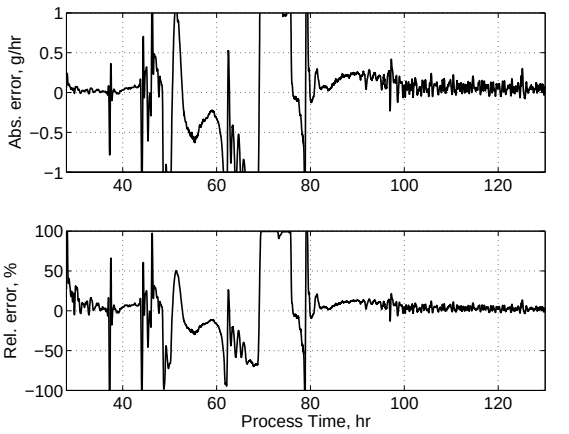


(b) Differences between measured and estimated values. Absolute (top) and relative (bottom) values.

Figure 8.12. Results from estimating $F^m_{NH_3}$ in cultivation MTS05. Validation batch. See caption for figure 8.8 for calculation of relative values.



(a) Normalized ammonia flow rate $F^m_{NH_3}$. Measured (light) and estimated (dark) values.

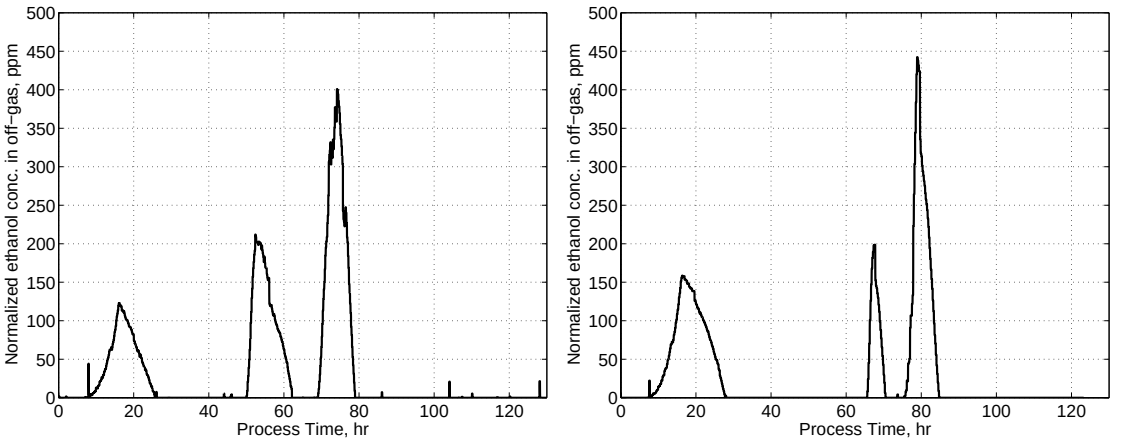


(b) Differences between measured and estimated values. Absolute (top) and relative (bottom) values.

Figure 8.13. Results from estimating $F^m_{NH_3}$ in cultivation MTS01. Validation batch. See caption for figure 8.8 for calculation of relative values.

Cultivation MTS01, shown in figure 8.13, was one of the more problematic cultivations. Not only were oscillations observed at 36 and 44 hours, but subsequently the $F^m_{NH_3}$ signal showed a lot of variations. At 48 hours the measured ammonia flow rate increased rapidly to the maximum value of 3.75 g NH_3/hr for 0.5 hours. The signal can be seen to decrease for a short while followed by yet another peak. The increase is initiated 20 min before the nutrient substrate feeding is started, hence the peak can be seen as the joint effect of the onset of nutrient feeding and metabolic activities. For the next thirty hours (50-80 hours) a peculiar pattern can be seen in the measured ammonia flow rate, indicating that the cultivation is far from balanced

growth. The possible reasons for these discrepancies is discussed in chapter 9, and is related to conversion of acetate and ethanol, products of oxido-reductive growth. Figure 8.14(a) show the trajectory of ethanol concentrations measured in the off-gas for MTS01 and MTV01. Balanced growth conditions are reestablished before 90 hours, after which the feed rate is increased until the end of the cultivation. The two signals are similar in this period, which can also be seen from figure 8.13(b).

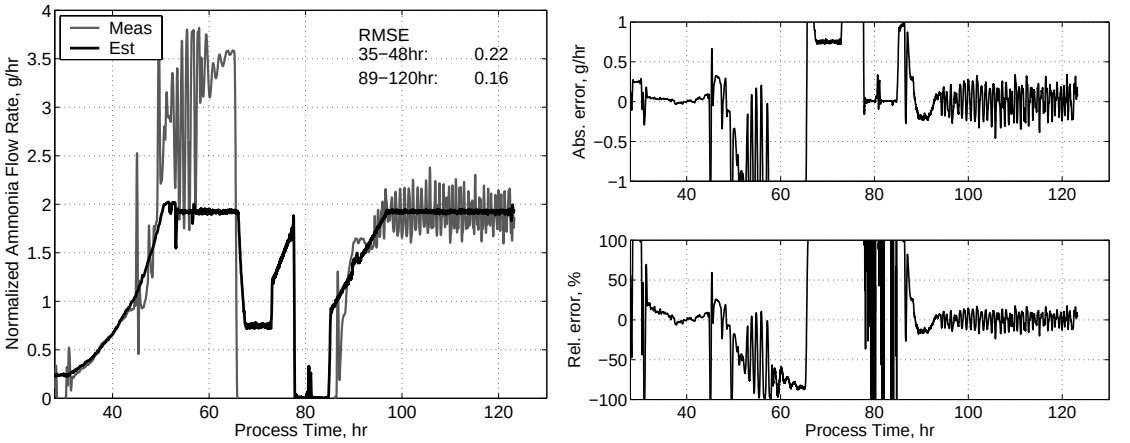


(a) MTS01 (b) MTV01

Figure 8.14. Normalized trajectories of ethanol concentrations (ppm) in the off-gas. The process is operated as a batch process for the first 30 hours. Ethanol is formed in the beginning of this phase up until 20 hours and then consumed before fed-batch operation is initiated around 30 hours.

Cultivation MTV01 was also one of the problematic cultivations as can be seen from figure 8.15(a). Synchronization bursts occur at 45 hours and at 49 hours. During the second burst $F^m_{NH_3}$ increases towards a maximum value of 3.75 g NH_3/hr. From 49 hours and until 65 hours, the average measured ammonia flow rate is significantly above the estimated flow rate. At 51 hours nutrient substrate feeding is initiated, which by close inspection of the signal does not seem to have a large influence on the measured ammonia flow rate, which is already very high. Just before ethanol production sets in the ammonia flow rate drops to zero. At 65 hours the substrate feeding rate is automatically reduced due to the ethanol production shown in figure 8.14(b) and at a normalized dilution rate of 0.38 kg/kg/hr, the ethanol is removed by dilution and consumption. An initial attempt to reestablish the process failed due to a too aggressive feeding strategy resulting in renewed ethanol production at 75 hours as seen in figure 8.14(a). A second try with a more conservative feeding ramp was successful and after 100 hours the process was back on track.

Large variations are observed in the $F^m_{NH_3}$ signal during the last 20 hours of the cultivation. The corresponding variations is $\pm$ pH 0.05 around the pH set point. The variations seem to be growth related, since CER, OUR and DOT signals show the same variations, indicating synchronization of cell cycles. In figure 8.16(a) the scaled trajectories of $F^m_{NH_3}$, CER and DOT are shown in the time period 80-120 hours, and it can be seen that the oscillations appear at a frequency of just above



(a) Normalized ammonia flow rate $F^m_{NH_3}$. Measured (light) and estimated (dark) values.

(b) Differences between measured and estimated values. Absolute (top) and relative (bottom) values.

Figure 8.15. Results from estimating $F^m_{NH_3}$ in cultivation MTV01. Between 65 and 85 hours the culture broth pH deviates significantly from the set point value (not shown here) and therefore large discrepancies between estimated and measured ammonia flow rates can be seen in this time interval. Validation batch. See caption for figure 8.8 for calculation of absolute and relative values.

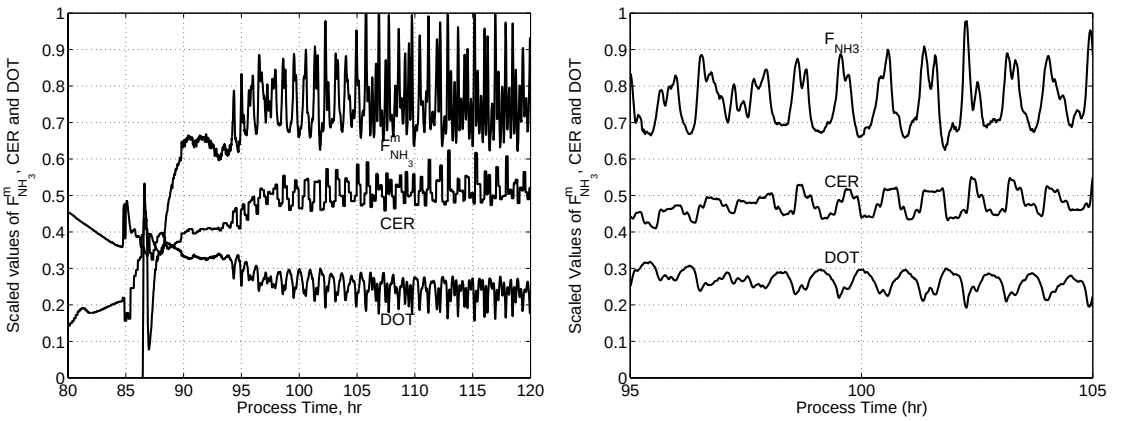
1 per hour, which can be seen from the zoom of the time period 95-105 hours provided in figure 8.16(b). The frequency suggests that a number of subpopulations have synchronized growth independently from each other, since the time needed to complete a full growth cycle is reported to be above one hour (Münch, 1992).

8.4 Indicator of Oxidative Growth and Onset of Oxidoreductive Growth

The proposed model for estimating the ammonia flow rate can be used to form an indicator for oxidative growth. Modelbased monitoring on the error ($e_n = \hat{F}^m_{NH_3} - F^m_{NH_3}$) between measured and estimated ammonia flow rates can be used to maintain balanced oxidative growth conditions, where corrective actions are taken as e_n exceeds certain warning limits. The model for estimating the ammonia flow rate assumes asynchronous growth, and as a result it would not be valid for monitoring during synchronization bursts, which were observed during some of the experiments discussed in this chapter.

When the warning limits are exceeded the sign of e_n can be interpreted to provide a possible explanation for the reason for the violation of the warning limits. As oxidoreductive growth has in this work been seen to be accompanied by formation of acetate (more precisely acetic acid), this leads to an acidification of the culture broth, requiring an increased ammonia flow rate to maintain a constant pH of the culture broth. In other words, the onset of oxidoreductive growth will result in a violation of a negative warning/detection limit on e_n.

In order to construct constant detection limits that can be used both during fed-



(a) Time period 80-120 hours (b) Time period 95-105 hours

Figure 8.16. Profiles of scaled ammonia flow rate ($F^m_{NH_3}$), carbon emission rate (CER) and dissolved oxygen tension (DOT) of cultivation MTV01 during 80-120 hours (a) and during 95-105 (b). Oscillating behavior can be observed. Scaling of normalized values have been done to facilitate the illustration of the oscillations in one figure. Note that RQ is only fluctuating between 1.09 and 1.12 during 95-120 hours.

batch and continuous operation, a relative value of the error is used rather than the absolute value presented above. This relative error value was also presented in the caption of figure 8.8.

$$e_n^r = \hat{F}_{NH_3}^m / F_{NH_3}^m - 1 \tag{8.14}$$

Considering only periods of operation where oxidative growth appears without synchronization bursts the relative errors (*e.g.* the plots in the bottom of figures 8.7(b), 8.8(b), and 8.10(b)) indicate that as long as the relative error is within ± 10% oxidative growth can be assumed. Figure 8.17 illustrates how the indicator of oxidoreductive growth performs with detection limit at - 10% of the relative errors for validation batches MTS12, MTS05, MTS01 and MTV01. The detection limit at +10% can be interpreted as the conversion of metabolites formed during oxidoreductive growth *i.e.* consumption of acetate.

In figure 8.17(a) it can be seen how cultivation MTS12 only show a few examples of process conditions where the onset of oxidoreductive growth could be present, and after 60 hours of operation no sign of oxidoreductive growth is seen. During fed-batch operation between 30 and 55 hours the indicator detects onset of oxidoreductive growth a number of times. During the inital 5 hours of fed-batch operation (30-35 hours) the process is not yet in balance in terms of pH and ammonia flow rate as can be seen from figure 8.11(a) (pH profile not shown). As mentioned above, a stop in ammonia dosing is responsible for the erroneous detections of oxidoreductive growth between 41 and 45 hours, since assumptions behind the model are violated (assumption 5 in table 8.6). At 53.5 hours the feeding of nutrient substrate is initiated, which triggers the indicator of oxidoreductive growth. Between 60 and 70 hours, during continuous operation of the process, the detection limit at +10% of the relative error is violated in two periods, however from the upper plot in figure 8.17(a) it can be seen that the upper detection limit is just barely violated. An explanation for these observations has not been determined.

Figure 8.17(b) show the performance of the indicator when used on data from cultivation MTS05. As mentioned above this cultivation was characterized by 3 oscillations in the fed-batch phase and during continuous operation noise (or oscillations) appeared in the measurement of the ammonia flow rate. These discrepancies from normal operation can be seen to influence the performance of the indicator of oxidative growth as a large number of detections are seen both during fed-batch and continuous operation. Rather than extending the detection limits, such observations should trigger an investigation of what the explanation behind the peculiar behavior could be.

Figure 8.17(c) show data from cultivation MTS01 where oxidoreductive growth is known to occur between 50 to 70 hours as discussed above and shown in figure 8.14(a). It can be seen that the indicator detects oxidoreductive growth between 53 and 55 hours of operation, however it also shows that the oxidoreductive growth continuous up until 69 hours of operation, despite the indication in figure 8.14(a) that ethanol is being consumed up until 62 hours. It is demonstrated and discussed in chapter 9, that between 53 and 69 hours of operation large amounts of acetate is being formed, whereas ethanol is only formed between 53 and 55 hours and again at 69 to 75 hours. The formation of acetate explains why oxidoreductive growth is correctly indicated to occur between 53 to 69 hours, while from 69 hours and

until 75 hours the indicator no longer responds correctly, propably explained by too extensive violation of the model assumption of balanced oxidative growth in table 8.6 (assumption 1). From 81 hours to the end of the cultivation, only a few detections of deviation from oxidative growth is seen. It is interesting to notice that the dilution rate is slowly increased with 20% from 100 to 130 hours, however no onset of oxidoreductive growth is seen in this period.

Figure 8.17(d) provides another example of how the indicator reacts to the onset of oxidoreductive growth. The consequence of a synchronization burst can be seen at 45 hours, where onset of oxidoreductive growth is indicated. More intersting is the extended period of time where oxidoreductive growth is indicated to take place between 48 and 65 hours, while figure 8.14(b) show that ethanol is not being formed up until 65 hours of operation. As ethanol is being formed the performance of the indicator can be seen to deteriorate explained by a too extensive violation of the model assumptions in table 8.6. As the system gets back on track around 92 hours of operation, oscillations in the measurements of the ammonia flow rate can be seen to trigger the indicator of oxidoreductive growth repeatedly.

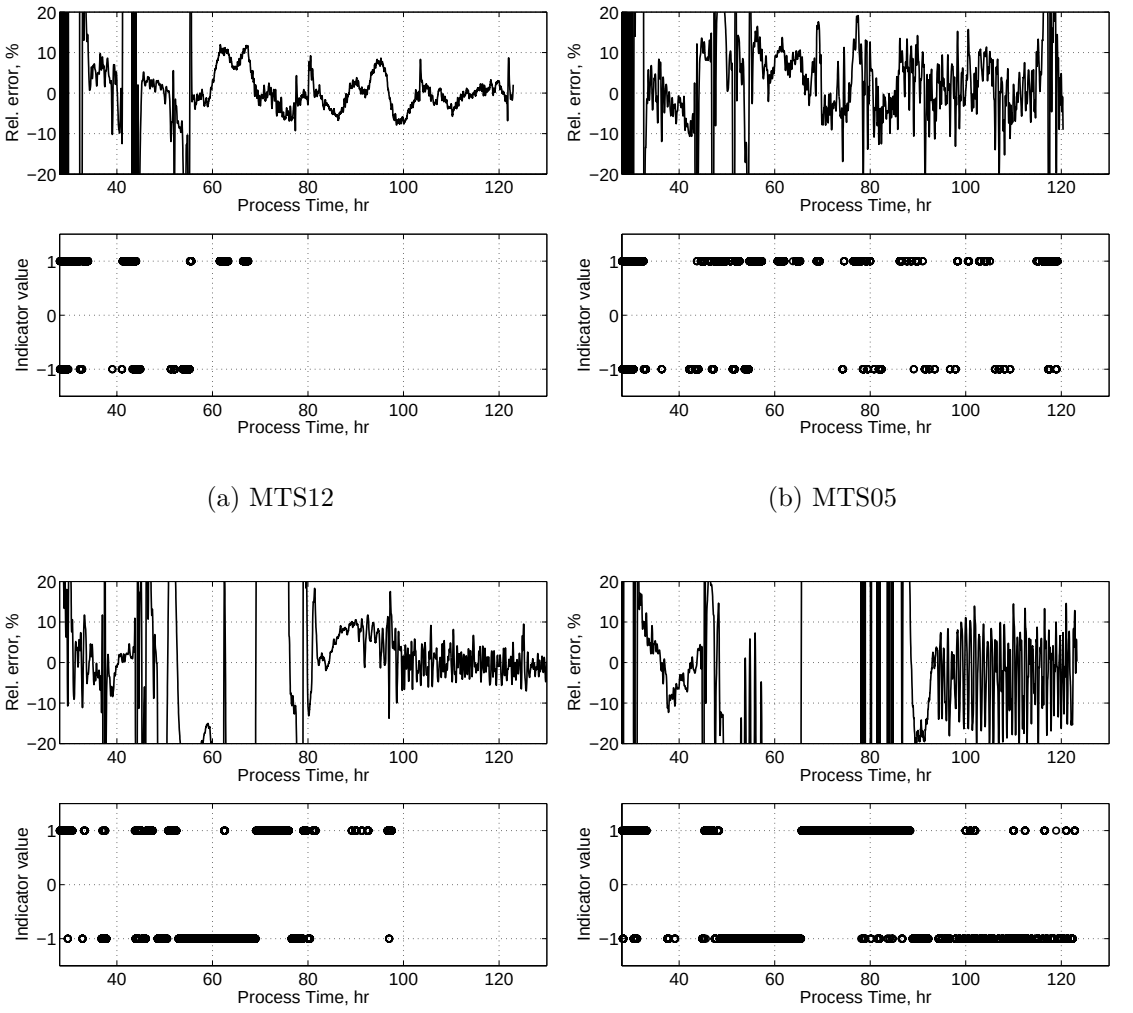
8.5 Discussion

8.5.1 Effects of Buffer Capacities

Investigation of the buffer capacity of glucose syrup and nutrient substrate has been carried out. Table 8.7 lists the ammonia demand for the two substrate to elevate their pH with 1.7 pH-units in the upper row. In section 8.2.1 it was assumed that the demand of ammonia to neutralize the effect of protons produced due to the growth mechanism was much larger than the effect from a low pH of the glucose substrate. The results listed in table 8.7 support this assumption with the need for ammonia to increase pH of the glucose syrup accounting for 4% of the total demand of ammonia to neutralize both effects. In case of the buffer capacity of the nutrient substrate, it can be seen in table 8.2.1 that the value obtained when adjusting pH prior to inoculation is four times lower than the one identified during a running cultivation, however the order of magnitude is approximately the same. The ammonia demand needed for neutralization of the buffer capacity of the nutrient substrate during a continuous operation is 0.01 g NH_3/hr, compared to 1.9 g NH_3/hr required to neutralize the effect of protons produced during growth.

Table 8.7. Buffer capacity of substrates. Normalized ammonia demand for increasing pH with 1.7 pH-units. The value for the glucose syrup in a running cultivation is listed in parenthesis, since it also includes the neutralization of protons produced due to growth.

	Nutrient	Glucose
Unit	[g NH_3/g Ndos]	[g NH_3/g Cdos]
Prior to inoculation	$6.0 \cdot 10^{-5}$	$2.4 \cdot 10^{-4}$
Running cultivation	$1.2 \cdot 10^{-5}$	$(5.6 \cdot 10^{-3})$



(a) MTS12

(b) MTS05

(c) MTS01

(d) MTV01

Figure 8.17. Indicators of onset of oxidoreductive growth. Relative errors (upper plot) shown together with indicator plots (lower plot) for cultivations MTS05, MTS12, MTS01 and MTV01. The circles indicates that the action limits at ± 10% of the relative error has been violated.

8.5.2 Process Upsets

The two examples of problematic cultivations, MTS01 (figure 8.13) and MTV01 (figure 8.15), support the hypothesis that metabolic activities different from balanced growth contribute to the proton balance and therefore influence the ammonia demand. Both cultivations illustrate that the onset of nutrient substrate feeding does not contribute significantly to the proton balance, and is not directly responsible for the high ammonia flow rates associated with the process upsets. The information gained from the discrepancy between observed and estimated flow rates of ammonia can be used for further studies of the metabolic mechanisms responsible for the process upsets.

Under the assumptions listed in table 8.6, an inverse form of the model in equation 8.11 can be used for estimation of the biomass concentration along the lines described

by Lei (2001). Combined with equation 8.5 the following estimation of the biomass formation rate can be obtained:

$$r_x = \frac{1}{\alpha V}\left(F_{NH_3} + \beta_e \cdot F^m_{out} + \beta_b \cdot \frac{\Delta W}{\Delta t}\right) \tag{8.15}$$

This approach is to be investigated further.

8.6 Monitoring of Oxidative Growth

In experiments with another strain of *S. cerevisiae* the assumption concerning balanced oxidative growth has been evaluated by considering the production of ethanol (Lei, 2001). It is not possible to apply the same approach with the strain that has been used in this work, since ethanol production does not appear to set in, if at all, until after a large deviation between estimated and measured ammonia flow rates have been observed. This can be seen by comparing figures 8.13(a) and 8.14(a) from cultivation MTS01, where large negative deviations between estimated and measured ammonia flow rates appear prior to ethanol formation. Similar observations can be seen in figures 8.14(b) and 8.15(a) for cultivation MTV01 except for the disturbance around 65 hours[7].

In this work the relative error (e^r_n) between measured and estimated ammonia flow rates has been used to monitor that the desired oxidative growth was maintained in the process, using relative warning/detection limits at $\pm$ 10% of the relative error. It would be even better to define these limits in terms of the specific ammonia flow rate *i.e.* the ammonia flow rate per amount biomass, since this would provide a stronger link to the growth metabolism and thus the interpretation of the violation of detection limits as will be demonstrated in chapter 9. The relative warning/detection limits at $\pm$ 10% of the relative error corresponds approximately to a detection limit of $\pm$ 1.5 mmole NH_3/mole biomass/hr. It would be desirable to further tighten the detection limits, however this requires that the model is extended to include such factors as *e.g.* tank-to-tank differences in measurements and the dynamics of the pH control loop, which is responsible for the addition of ammonia. Rather than extending the model, the detection limit could be time-dependent or methods from statistical process control *e.g.* the cumulated sum (CUSUM) approach could be used to monitor e^r_n.

8.7 Conclusion

The identified model, equation 8.11, is able to provide a reasonable estimate of the ammonia feed flow rate during balanced oxidative growth. This was demonstrated by using the model on validation batches MTS05 and MTS12. A number of discrepancies were observed between the estimated and measured signal of the ammonia flow rates, especially during the switch over from fed-batch to continuous operation. It is believed that these discrepancies can be ascribed to undetermined metabolic

[7] At this point the culture broth pH increases above its set point value only to return at 86 hours.

activities *i.e.* changes in metabolic flux distribution, which influence the amount of protons transported from the cell to the surrounding broth and subsequently the addition of ammonia to neutralize this effect. A better description can probably be obtained by addressing some of the assumptions and provide a better description of changes to the elemental composition of the biomass or even to the macromolecular (protein, carbohydrate, lipid *etc.*) distribution of the yeast, effects that can have an influence on the flux distribution especially in the fed-batch phase.

The use of the proposed model for monitoring of oxidative growth condition was also presented and discussed. During process conditions where the model assumptions were valid, the modelbased monitoring of the oxidative growth condition performed well. Synchronized bursts due to cell cycle synchronization resulted in violations of the warning limits for oxidative growth set at ± 10% of the relative error (e_n^r) between measured and estimated ammonia flow rate. The relative warning/detection limits at ± 10% of the relative error corresponds approximately to a detection limit of ± 1.5 mmole NH_3/mole biomass/hr. Violation of the lower warning limit for oxidative growth condition could be interpreted as an indicator of the onset of oxidoreductive growth. The indicator was shown to accurately capture the onset of oxidoreductive growth when applied on data from cultivation MTS01 and MTV01.

9

Soft Sensors for Estimating Biomass and Acetate Concentrations, Intrinsic Metabolic Fluxes and Product Concentration

For the analysis and online monitoring of aerobic cultivation of Saccharomyces cerevisiae three simple soft sensors have been constructed. The first soft sensor is used for the online estimation of biomass and acetate conversion rates and concentrations. This sensor illustrates how acetate production is seen to occur prior to ethanol formation at the onset of oxido-reductive growth. The second soft sensor provides estimates of the intrinsic metabolic flux distribution using a model of a simple metabolic network. By analysis and online monitoring of the metabolic flux distributions, it is possible to identify changes in the metabolism e.g. oxidative to oxido- reductive growth, as well as to investigate the magnitude of the individual fluxes in the model for signs of bottlenecks or other limiting conditions such as onset of repression. The third soft sensor provides estimates of the production rate and concentration of the product, an insulin precursor. It is shown that the specific productivity is not constant throughout the process. More than 80 hours of oxidative growth is needed for the specific productivity to reach a high constant level. Furthermore the productivity is affected adversely by oxido-reductive growth, increasing the time needed to reach the high constant level of specific productivity.

With the information provided by the two previous chapters it is now possible to further investigate the information available in the online process data. Due to the multivariate nature of a cultivation process and therefore also the process data provided, it is desirable to construct a framework for the interpretation of this multivariate data. By constructing simple mathematical models based on physiological insight, process knowledge and data analysis, it is possible to extract useful information, which is not directly available in the online data.

The use of mathematical modeling to describe the distribution of metabolic fluxes can facilitate the elucidation of a number of interesting attributes related to the growth and product formation of a recombinant strain of *S. cerevisiae* used in an established industrial application for production of an insulin precursor.

The process investigated in this study is an aerobic, carbon-limited cultivation ini-

tially operated as a batch process followed by a fed-batch phase leading to continuous operation of the bioreactor characterized by a high cell density (> 40 gDW/L). The process has shown peculiar behavior when tested under conditions more extreme than the normal operating conditions. The observed phenomena could not be accounted for using the accumulated experience and knowledge of the strain. It was expected that the response to substrate overfeeding and similar disturbances triggering overflow metabolisms would be ethanol formation. However it turned out that extended flow of gaseous ammonia was needed in order to maintain a constant pH of the culture broth prior to production of ethanol. Offline measurements of the broth composition using HPLC also indicated the presence of acetic acid and in some cases glycerol.

Intrigued by these observations, a more elaborate investigation of the data at hand was carried out to unveil the mechanisms behind the observed phenomena. The purpose of this chapter is to provide a simple metabolic flux model to provide the means to 1) detect and 2) diagnose changes in the dynamic flux distribution during different process conditions. First a description of the production rate of biomass and acetate is developed and presented. Secondly, using estimated and measured conversion rates a simple flux model is applied for estimation of the specific flux distribution during normal operation as well as before and during process disturbances, highlighting critical conditions leading to significant byproduct formation. Thirdly, the estimation of the biomass concentration renders it possible to propose a simple model for description of the specific production rate of the insulin precursor (the product), thereby facilitating estimation of the product concentration for online monitoring.

The data used in this chapter has been normalized which influences the numerical values of the identified model parameters. Although the models, which are presented here, are evaluated on offline data, it is the aim of this work to provide a framework and tools to facilitate online monitoring and control of cultivation processes and also to facilitate offline data analysis and comparison of process behavior to identify possible improvements in the process operation.

The chapter is structured as follows. After the introduction above, including the background for and purpose of the chapter, section 9.1 provides a description of the materials and methods used in this chapter. An important parts of this section is the description of the methods used to provide estimated concentration trajectories of important species *e.g.* estimation of effluent flow rate and weight changes. Section 9.2 describes the construction of a soft sensor for online estimation of the conversion rates and concentrations of biomass and acetate. Section 9.3 describes how a soft sensor for online monitoring of the intrinsic metabolic flux distribution can be constructed based on a simple metabolic model of *S. cerevisiae*. Section 9.4 describes how a simple soft sensor for monitoring of the insulin precursor production rate can be formulated using the information from the two other soft sensors. Following the presentation of the three software sensors, their applications are demonstrated in section 9.5 highlighting the online information that they each provide. The results and application of the three soft sensor are then discussed in section 9.6 by illustrating how the new information provided can be used for a deeper and more detailed analysis of the large amount of information available online process data. Section 9.7 summarizes and concludes on the results and observations

made in this chapter and also indicates possible future directions to extend the models behind the soft sensors and thereby also extending the knowledge of the system.

9.1 Materials and Methods

9.1.1 Normalized Data

The cultivation data reported in this chapter has been normalized in order to blur sensitive information. The requirement to do so makes it rather difficult to compare the findings reported in this work with results from the open literature. Furthermore it can be difficult to check and verify calculations and results since normalization has to be used in a way that also blurs ratios *e.g.* dilution rates and specific productivities.

The author apologies for the inconvenience that the normalization can cause and for the distortion it might have on understanding the issues addressed in this thesis.

9.1.2 Strain and Medium

Details on this topic are given in section 7.2.1. The biomass composition was investigated in chapter 7 and estimated to: $CH_{1.82}O_{0.576}N_{0.146}$ (f_r: 7.3 w%).

9.1.3 Equipment

Details on this topic are given in section 7.2.2. A number of process parameters and variables are measured and data series of these measurements are available with a sampling frequency, ω_s, of 1 min^{-1}. These variables were presented in table 8.1.

9.1.4 Analytical Methods

Some of these measurements are standard procedures at the pilot plant, while others required the use of analytical equipment (HPLC) at Biocentrum, DTU. Below is a description of the analytical methods used in order to provide the additional information for the interpretation of the process behavior.

9.1.4.1 Biomass and Product Concentration

Details on the determination of the biomass concentration is found in section 7.2.3.1. A small amount of supernatant from the first centrifugation from the analysis of the biomass concentration was used for in-house determination of insulin precursor concentration by HPLC.

9.1.4.2 Metabolite Analysis

Details on this topic is found in section 7.2.3.3.

9.1.4.3 Off-gas Analysis

Mass spectroscopy on the off-gas from the bioreactor provided measurements of ethanol concentration c_e^g, change in percent O_2 (OXC) and change in percent CO_2 (CXC) between inlet and outlet streams. Details on this topic is found in section 7.2.3.5.

The measured ethanol concentration in the off-gas (c_e^g) is filtered using a fourth order Butterworth filter with a cut-off of $0.0075{\cdot}\omega_s$. The resulting data series are then used to estimate the ethanol conversion rate (r_e) in the culture broth assuming an equilibrium between the ethanol concentration in the off-gas (c_e^g) and in the broth (c_e). By using offline measurements of the ethanol concentration in the broth (c_e) obtained form HPLC analysis and correlating this to the ethanol concentration in the off-gas (c_e^g) the following relation was found[1]:

$$c_e[g/L] = c_e^g[ppm]/120[ppm/(g/L)] \tag{9.1}$$

Using this the ethanol conversion rate can be calculated:

$$\frac{dVc_e}{dt} = Vr_e - F_{out}c_{e,out} \Leftrightarrow \tag{9.2}$$

$$r_e = \frac{dc_e}{dt} + \left(D + 1/V\frac{dV}{dt}\right)c_e \tag{9.3}$$

$$= 1/120\left(\frac{dc_e^g}{dt} + \left(D + 1/V\frac{dV}{dt}\right)c_e^g\right) \tag{9.4}$$

where $D = \frac{F_{out}}{V}$ and it has been assumed that there is no ethanol in the feed streams, that the removal of ethanol by the off-gas can be ignored and that $c_{e,out} = c_e$. For calculation of r_e please refer to the section 9.1.10.

9.1.5 Substrates

Details are given in section 7.2.4.

9.1.6 Cultivation Conditions

Details are given in section 7.2.5. The dissolved oxygen tension was measured online and for standard cultivations the dissolved oxygen tension was above 20% of air saturation. In one of the cultivations (MTF02) the dissolved oxygen tension came as low as 8% of air saturation for an extended period of time (70 hours). Table 9.1 provides an overview of experiments presented in this chapter.

9.1.7 Preparation of Inoculum

Details on this topic are given in section 7.2.5.1.

[1] It has not been possible to verify this correlation by comparison to equilibrium data from water/ethanol systems.

Table 9.1. Overview of presented experiments. Further description of the entries marked by a ⋆ can be found in the text. Standard cultivations are experiments where no disturbances (planned or unplanned) were observed during the fed-batch and continuous operation. The control of substrate feed rate is either carried out as a preprogrammed feeding profile by a series of set points with linear ramps in-between or by closed loop control using the feed rate to obtain a constant signal in an online measurement of reducible gases in the offgas from the bioreactor (see section 9.1.8.4 for details). Abbreviations: *op*: operation, *dist*: disturbances, *D*: dilution rate, *Approx*: approximate, *norm*: normalized. *BAC*: soft sensor for biomass and acetate conversion (page 142). *IMF*: soft sensor for intrinsic metabolic fluxes (page 147). *IPP*: soft sensor for insulin precursor production rate and concentration (page 152).

	MTS01	**MTS02**	**MTS05**	**MTS06**	**MTS07**
Standard cultivation	No	No	Yes	Yes	Yes
Perturbations in fed-batch and continuous op.⋆	Yes	Yes	Yes	Yes	Yes
Planned dist.⋆	Ramp in D	No	No	No	No
Unplanned dist. (type)	Ethanol	Ethanol	No	No	No
Closed loop experiments⋆	No	No	No	No	No
Batch operation (hr)	0-26	0-26	0-30	0-30	0-30
Fed-batch operation (hr)	26-51	26-51	30-55	30-55	30-54
Continuous operation (hr)	51-144	51-144	55-120	55-121	54-120
Approx. norm. D in continuous op. (kg/kg/hr)	-	1.06	0.98	1.01	0.99
Control of substrate feed rate	Recipe	Recipe	Recipe	Recipe	Recipe
BAC-**results in figure**	9.11		9.12		
IMF-**results in figure**	9.17+9.18		9.15+9.16		
IPP-**results in figure**	9.25+9.26	9.25	9.24		9.24

	MTS11	**MTS12**	**MTE02**	**MTV03**	**MTF02**
Standard cultivation	Yes	Yes	No	No	No
Perturbations in fed-batch and continuous op.⋆	Yes	Yes	No	No	No
Planned dist.⋆	No	No	Ramp in D	Steps in D	Closed loop on ethanol
Unplanned dist. (type)	No	No	No	Ethanol	Ethanol
Closed loop experiments⋆	No	No	No	No	Yes
Batch operation (hr)	0-30	0-30	0-30	0-29	0-29
Fed-batch operation (hr)	30-54	30-55	30-51	29-55	29-51
Continuous operation (hr)	54-122	55-123	51-123	55-122	51-288
Approx. norm. D in continuous op. (kg/kg/hr)	1.01	1.00	-	0.99	-
Control of substrate feed rate	Recipe	Recipe	Recipe	Recipe	Ethanol
BAC-**results in figure**			9.13		
IMF-**results in figure**			9.20+9.21		9.23
IPP-**results in figure**	9.24	9.24	9.25+9.26	9.25	

9.1.8 Feeding Strategies

This section will describe the different feeding strategies applied in the cultivations used for model construction and validation. The main reason for using different feeding strategies was to provide more information on the process dynamics. An overview of the feeding strategy used in this chapter is given in table 7.3 on page 86 in chapter 8. Below more detailed descriptions of feeding strategies are provided, starting with the fed-batch and continuous operation. Then the perturbations used in a number of cultivations are presented, followed by a description of some planned changes to the dilution rate during continuous operation. Finally a feeding strategy based on closed loop control is described, using the signal of an additional measurement of reducible gases, primarily ethanol, in the offgas from the bioreactor to control the substrate feed rates.

9.1.8.1 Fed-batch and Continuous Operation

Details are given in section 7.2.5.3. In most of the cultivations preprogrammed profiles of the substrate feed rate were used. In these cultivations short-term presence of small concentrations of ethanol in the culture broth were allowed since these provided valuable information in order to understand the process dynamics close to the onset of ethanol formation. The long-term Crabtree effect should be suppressed in these investigations, since the use of high concentrations of glucose in the substrate could lead to very high ethanol concentrations, reaching levels that are toxic for the microorganism and lead to wash-out of the cells during continuous operation. To suppress the long-term Crabtree effect, a control strategy was used, which lowered the substrate feed rates in steps of 10 % of the actual feed rates, if the ethanol concentration in the offgas from the bioreactor exceeded a certain value. Following a successful decrease of the substrate feed rates, the ethanol concentration will decrease and purely oxidative growth resumed at lower substrate feed rates.

9.1.8.2 Perturbations

Details are given in section 8.1.8. An example of perturbations added to a nominal feeding profile is illustrated in figure 9.1 were also a lower constraint on the minimum feed rate at a normalized value of 10 g/hr has been added. In addition figure 9.2(a) illustrates an example of perturbations to the feeding rates in continuous operation.

9.1.8.3 Changes to the Dilution Rate

In three of the planned cultivations, changes to the normalized dilution rate have been carried out. Figure 9.2 illustrates the normalized dilution rates along with the ethanol concentrations in the off-gas for the three cultivations and an example of a cultivation, MTS05, where only perturbations to the dilution rate has been carried out (figure 9.2(a)).

Figure 9.2(b) illustrates the trajectory of the applied dilution rate during continuous operation of cultivation MTS01. A ramp increasing the normalized dilution rate from 1.00 kg/kg/hr to 1.21 kg/kg/hr between 100 and 144 hours was carried out with the aim of obtaining a crude determination of the critical dilution rate D_{crit} at

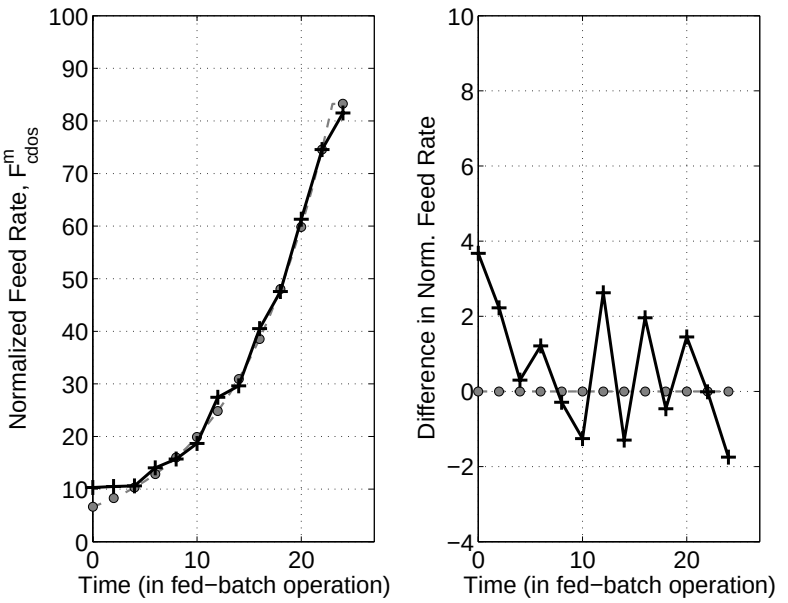
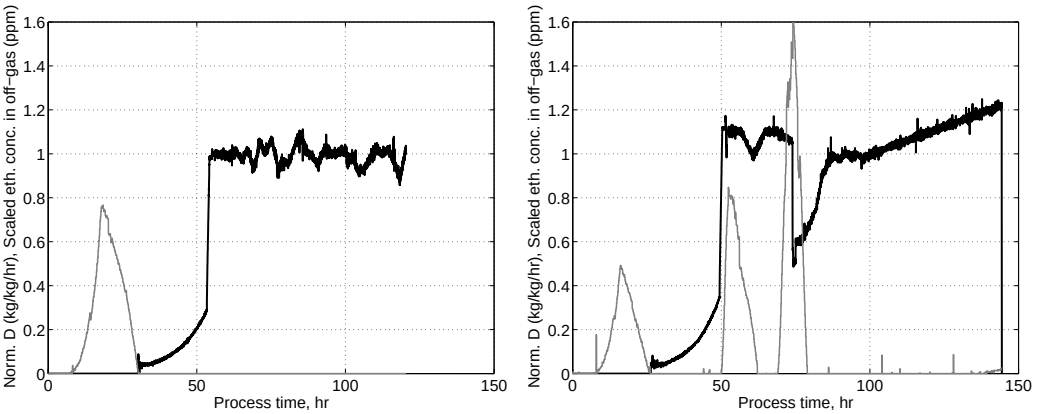


Figure 9.1. Left: Exponential feeding profile (grey dashed line) with grid points (grey filled circles) and perturbed feeding profile (black line, +). Right: Difference between exponential feeding profile (grey dashed line, grey filled circles) and perturbed feeding profile (black line, +).

which oxido-reductive growth is activated by production of ethanol (Postma *et al.*, 1989b). Prior to this planned disturbance, large unplanned disturbances had occurred just around the change from fed-batch to continuous operation and again 20 hours into continuous operation observed as two large peaks of ethanol.

Figure 9.2(c) illustrates the trajectory of the applied dilution rate in cultivation MTE02, where a steeper feeding profile during fed-batch operation was attempted. This substrate feeding profile corresponded to a linear decrease in the normalized growth rate (assuming balanced oxidative growth) from 2.00 kg/kg/hr to 1.00 kg/kg/hr shortening the fed-batch phase with approximately 4 hours (16%) from 25 to 21 hours. No ethanol was formed in the fed-batch phase indicating that the use of a more aggressive feeding profile was indeed possible. After initiation of continuous operation the cultivation was operated at an unperturbed, normalized dilution rate of 1.00 kg/kg/hr for 30 hours before a linearly, increasing ramp in the dilution rate was initiated. From 80 to 116 hours the normalized dilution rate was increased from 1.00 kg/kg/hr to 1.45 kg/kg/hr at which point ethanol was observed in the off-gas indicating activation of the fermentative pathway. At 117 hours the normalized dilution rate was lowered to 0.96 kg/kg/hr and the ethanol disappeared within 2.5 hours.

Figure 9.2(d) illustrates the trajectory of the applied dilution rate in cultivation MTV03, where three step changes in the dilution rate were carried out. During continuous operation at 79.5 hour the normalized dilution rate was instantly increased from 1.00 kg/kg/hr to 1.50 kg/kg/hr, which resulted in the onset of ethanol production indicating activation of the fermentative metabolism. At 82.5 hours the normalized dilution rate was instantly decreased from 1.50 kg/kg/hr to 0.60 kg/kg/hr, leading to removal of ethanol through consumption and dilution within



(a) MTS05, standard cultivation

(b) MTS01

(c) MTE02

(d) MTV03

Figure 9.2. Normalized dilution rates (black lines) and scaled ethanol concentrations in the off-gas (grey lines) of the four cultivations MTS05, MTS01, MTE02 and MTV03. See text for detailed information on the individual plots.

1.5 hours. At 84.5 hours the normalized dilution rate of 1.00 kg/kg/hr is reestablished. Prior to the step changes, at the onset of continuous operation, ethanol was observed in the off-gas and the dilution rate was temporarily lowered to remove the ethanol.

9.1.8.4 Closed Loop Control of Reducible Gases in Offgas

In cultivation MTF02 a sensor able to measure reducing gases, primarily ethanol, was placed in the off-gas stream from the bioreactor. Between the location of the sensor and the exit port in the bioreactor, a filter is positioned to capture droplets of culture broth in the off-gas. From the filter the main part of the off-gas is led to an exhaust pipe, while a small gas stream is led by a heat-traced pipe to be analyzed by mass spectroscopy. The reducible gas sensor, a Figaro TGS 822 sensor, was positioned approximately 4 meters after the filter along the heat traced pipe. With an aeration rate of 18 L/min and an estimated inner diameter of the pipe of 0.25 cm, this corresponds to a time delay of less than 1 second. The Figaro sensor provided a 4-20 mA signal to a PID controller, that attempted to maintain a constant level of reducible cases in the off-gas, assumed to be primarily ethanol. The detection range of the Figaro sensor was 0-0.3 g ethanol/NL.

9.1.9 Filtering of Data

Details are given in section 8.1.10. A fourth order Butterworth filter is used to filter the originally sampled data signals.

9.1.10 Mass Flux Balances

The aim of this work has been to provide a model framework for online monitoring of a number of important and critical species in the cultivation broth, such that the time evolution of both conversion rates (r_i) and concentrations (c_i) of a number of species (i) in the cultivation broth can be used to analyze and evaluate the state of the cultivation. Conversion rates and concentrations are related through component mass flux balances *e.g.* on biomass, while elemental mass flux balances, *e.g.* on carbon, can be used to estimate unmeasured components.

A mass flux balance equation for each of the species in the bioreactor can be set up:

$$\text{Accumulation} = \text{Reaction} + \text{In} - \text{Out} \Leftrightarrow \tag{9.5}$$

$$\frac{dVc_i}{dt} = r_i V + F_{in}c_{i,in} - F_{out}c_{i,out} \tag{9.6}$$

V is the volume of the cultivation broth, c_i is the concentration of species i, F_{in} and F_{out} are the volumetric flow in and out of the bioreactor respectively, while r_i is the volumetric conversion rate. It should be noted that equation 9.5 can be based on a number of units *e.g.* amounts (moles) or mass (kg) per volume (L) or mass (kg). Here amounts per volume (moles/L) is used as an example.

Since operation regimes with variable volume are considered *e.g.* fed-batch operation, the balance equation is rewritten as:

$$V\frac{dc_i}{dt} = r_i V + F_{in}c_{i,in} - F_{out}c_{i,out} - c_i\frac{dV}{dt} \tag{9.7}$$

Considering very small time differences and assuming constant intervals of variables between the sampling points, a one-step-ahead prediction of the concentration trajectory can be obtained from:

$$c_{i,k+1} = c_{i,k} + \Big(r_{i,k} + D_{in,k}c_{i,in,k} - D_{out,k}c_{i,k} - \frac{c_{i,k}}{V_k}\left(\frac{\Delta V}{\Delta t}\right)_k\Big) \cdot (t_{k+1} - t_k) \tag{9.8}$$

F_{out}, and consequently D_{out}, is not directly measured, while the determination of $\frac{\Delta V}{\Delta t}$ is subjected to a large amount of noise if the raw weight measurements are to calculate the change in volume as:

$$\left(\frac{\Delta V}{\Delta t}\right)_k \approx \frac{1}{\rho}\left(\frac{\Delta W}{\Delta t}\right)_k = \frac{1}{\rho}\frac{W_{k+1} - W_k}{t_{k+1} - t_k} \tag{9.9}$$

Better estimations of $\frac{\Delta V}{\Delta t}$ and the effluent flow rate F_{out} can be obtained using a Kalman Filter; this is discussed in section 9.1.11.

After having set up a framework to determine conversion rates and concentrations, it is relevant to evaluate the conversion rates per unit biomass. This information can be obtained by calculating the specific conversion rate q_i of species i as:

$$q_i = \frac{r_i}{c_x} \quad [mole/mole/hr] \tag{9.10}$$

where r is the volumetric conversion rate and c_x is the biomass concentration.

9.1.11 Estimating Effluent Flow Rate and Rate of Weight Change

In the section above on the use of mass flux balances for estimating conversion rates and concentration of relevant species, it was mentioned that $\frac{\Delta V}{\Delta t}$ and the effluent flow rate F_{out} need to be estimated. The reason for this is the design and configuration of the bioreactors that are used in this work. For this purpose a model for predicting of the weight of the bioreactor contents is constructed, and formulated in a way that suits implementation of the model in a Kalman Filter (Madsen and Holst, 2000). The flexibility of the Kalman Filter algorithm will be demonstrated during the change from fed-batch to continuous operation, where an elegant update of the filter states results in almost no time delay in the estimation of $\frac{\Delta V}{\Delta t}$ and F_{out}.

During fed-batch operation $F_{out} = 0$ except during the occasional sampling of culture broth. Using the raw signal of the weight measurement or even the Butterworth filtered signal to calculate the rate of weight change and subsequently the rate of change in volume of culture broth using equation 9.9 results in very noisy estimates. This is illustrated in figure 9.3. Observing the trajectories of both the raw and Butterworth filtered weight signals in figure 9.3(a) illustrates that the weight signal

does not show large fluctuations except when samples are being taken. In figure 9.3(b) it can be seen that both the raw and Butterworth filtered signals[2] are very noisy. Using such signals in the calculations of conversion rates and concentrations as discussed in section 9.1.10 will propagate the noise into these estimates.

During continuous operation the removal of culture broth to maintain a constant bioreactor weight W is carried out by an ejection mechanism that rapidly ejects approximately 50 g broth as the bioreactor weight exceeds the set point value with approximately 100 g. This is illustrated in figure 9.4 where the black line shows how 16-18 ejections occur during the time span of 1 hour. The fluctuations in the weight signal only corresponds to 0.5 % of the weight during continuous operation and is believed not to have a significant effect on the state of the cultivation. In most cases it would be sufficient to estimate the effluent flow rate as:

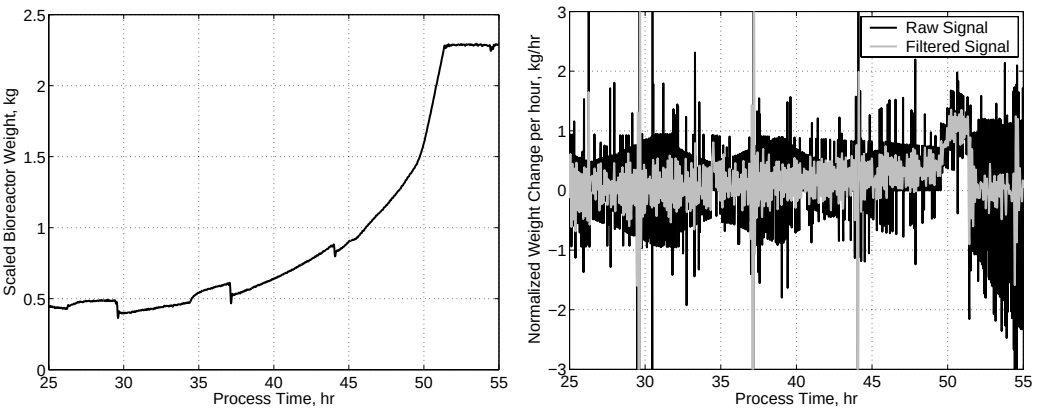
$$F_{out,k} = F_{in,k} = F_{cdos,k} + F_{ndos,k} \tag{9.11}$$

ignoring the contribution from the weight change *i.e.* $\frac{\Delta W}{\Delta t} = 0$. However in some of the cultivations changes to the set point of the weight occur during continuous operation, which means that F_{out} and $\frac{\Delta W}{\Delta t}$ are changed significantly at the same time and can not be estimated correctly, since the two assumptions $F_{out} = 0$ and $\frac{\Delta W}{\Delta t} = 0$ are not valid. Hence the assumptions of equations 9.9 and 9.11 can not be fulfilled.

In order to provide an improved estimation of both $\frac{\Delta W}{\Delta t}$ and F^m_{out}, a model is constructed:

$$W_{k+1} = W_k + \Delta W_k + F^m_{cdos,k} + F^m_{ndos,k} - F^m_{out,k} \tag{9.12}$$

[2]The filtering is performed on the raw signal and then followed by a calculation of the weight change as a difference between the values at two neighboring sampling points.



(a) Scaled weight profile

(b) Normalized profiles of rate of weight change

Figure 9.3. Left hand plot: The scaled weight profile. Both the raw signal of the weight measurement (black line) and the Butterworth filtered signal are shown (grey line), but they can not be distinguished as they appear to be identical at the resolution used in the figure. Right hand plot: The normalized profiles of the rate of weight change. Both the raw signal of the weight measurement (black line) and the Butterworth filtered signal are shown (grey line). Data in both plots are from fed-batch and onset of continuous operation in cultivation MTS01.

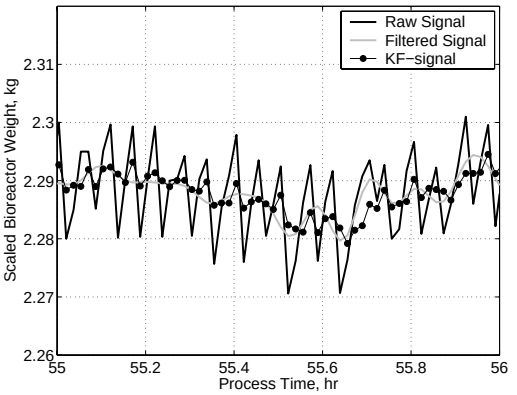


Figure 9.4. The scaled weight profile. The raw signal of the weight measurement (black line) is shown together with the Butterworth filtered signal (grey line) and the estimated weight signal from the proposed Kalman Filter (thin black line, •). Data are from fed-batch and onset of continuous operation in cultivation MTS01.

where:

$$\Delta W_k = \left(\frac{\Delta W}{\Delta t}\right)_k \Delta t_k \tag{9.13}$$

For fed-batch operation equation 9.12 is rearranged to a stochastic model with three states:

$$F^m_{out,k+1} = 0 + e^1_k \tag{9.14}$$

$$\Delta W_{k+1} = -F^m_{out,k} + F^m_{cdos,k} + F^m_{ndos,k} + e^2_k \tag{9.15}$$

$$W_{k+1} = \Delta W_k + W_k + e^3_k \tag{9.16}$$

where $\boldsymbol{e}$ represents white noise sequences. The reason for including the variable F^m_{out} in the model, although it was argued above that $F^m_{out} = 0$ during fed-batch operation, is that this variable can be used to represent different known disturbances *e.g.* sampling of culture broth. This is desirable for the calculation of ΔW in equation 9.15, yielding a much lower variance on this variable.

In matrix form this yields:

$$\begin{bmatrix} F^m_{out,k+1|k} \\ \Delta W_{k+1|k} \\ W_{k+1|k} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} F^m_{out,k|k} \\ \Delta W_{k|k} \\ W_{k|k} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} F^m_{cdos,k} \\ F^m_{ndos,k} \end{bmatrix} + \begin{bmatrix} e^1_k \\ e^2_k \\ e^3_k \end{bmatrix} \tag{9.17}$$

$$\hat{W}_{k+1} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} F^m_{out,k+1|k} \\ \Delta W_{k+1|k} \\ W_{k+1|k} \end{bmatrix} + \begin{bmatrix} \epsilon_k \end{bmatrix} \tag{9.18}$$

$\boldsymbol{e}$ and $\boldsymbol{\epsilon}$ are white noise sequences, the state noise and measurement noise respectively, with covariance matrices $E(\boldsymbol{e}^T\boldsymbol{e}) = \boldsymbol{Q}$ and $E(\boldsymbol{\epsilon}^T\boldsymbol{\epsilon}) = \boldsymbol{S}$. Equation 9.18 is added to the model for the use in the Kalman Filter algorithm presented below. The reason for using the Kalman Filter is that estimates of F^m_{out}, ΔW and W can be obtained by using a number of different measurements, each containing information relevant for the estimation. In this formulation F^m_{cdos} and F^m_{ndos} are used as

inputs, while measurements of W is used to check the consistency of the estimates, calculating the Kalman Filter gain $\boldsymbol{K}$ and correcting inconsistencies.

A general form of the model can be obtained by using:

$$\hat{W} = y \quad ; \quad \begin{bmatrix} F^m_{out} \\ \Delta W \\ W \end{bmatrix} = \begin{bmatrix} x^1 \\ x^2 \\ x^3 \end{bmatrix} \quad ; \quad \begin{bmatrix} F^m_{cdos} \\ F^m_{ndos} \end{bmatrix} = \begin{bmatrix} u^1 \\ u^2 \end{bmatrix} \tag{9.19}$$

where $\boldsymbol{x}$ represents model states and $\boldsymbol{u}$ are the inputs. A more general description of equations 9.14-9.16 can be formulated as:

$$\begin{bmatrix} x^1_{k+1|k} \\ x^2_{k+1|k} \\ x^3_{k+1|k} \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} x^1_{k|k} \\ x^2_{k|k} \\ x^3_{k|k} \end{bmatrix} + \boldsymbol{B} \begin{bmatrix} u^1_k \\ u^2_k \end{bmatrix} + \begin{bmatrix} e^1_k \\ e^2_k \\ e^3_k \end{bmatrix} \tag{9.20}$$

$$\hat{y}_{k+1} = \boldsymbol{C} \begin{bmatrix} x^1_{k+1|k} \\ x^2_{k+1|k} \\ x^3_{k+1|k} \end{bmatrix} + \begin{bmatrix} \epsilon_k \end{bmatrix} \tag{9.21}$$

The model parameters in matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are dependent on the state of operation such that during batch and fed-batch operation:

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \quad ; \quad \boldsymbol{B} = \begin{bmatrix} 0 & 0 \\ 1 & 1 \\ 0 & 0 \end{bmatrix} \quad ; \quad \boldsymbol{C} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \tag{9.22}$$

A slightly different model formulation is used for continuous operation:

$$F^m_{out,k+1} = -\Delta W_k + F^m_{cdos,k} + F^m_{ndos,k} + e^1_k \tag{9.23}$$

$$\Delta W_{k+1} = -F^m_{out,k} + F^m_{cdos,k} + F^m_{ndos,k} + e^2_k \tag{9.24}$$

$$W_{k+1} = \Delta W_k + W_k + e^3_k \tag{9.25}$$

As illustrated by the raw data signal (black line) in figure 9.4, both $F^m_{out,k}$ and ΔW_k changes significantly depending on whether the broth ejection mechanism is active or not. Therefore both these variables need to be estimated during continuous operation in order to provide a good account of the process dynamics. At a first glance the formulation of equations 9.23 and 9.24 can seem a little strange, since F^m_{cdos} and F^m_{ndos} appear in both equations and neither $F^m_{out,k}$ nor ΔW_k have a direct influence on the prediction of states $F^m_{out,k+1}$ or $\Delta W_k + 1$ respectively. The reason for this formulation is to support the ability of the Kalman Filter algorithm to update its internal states depending on the prediction error between the measurement W_{k+1} and the prediction $\hat{W}_{k+1}$.

Equations 9.23 - 9.25 leads to the model parameter matrices, when using the general model 9.20 - 9.21 :

$$\boldsymbol{A} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \quad ; \quad \boldsymbol{B} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix} \quad ; \quad \boldsymbol{C} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \tag{9.26}$$

9.1.11.1 Kalman Filter Algorithm

The two models are now ready to be implemented in the Kalman Filter algorithm. In the discrete Kalman Filter algorithm (Madsen and Holst, 2000) the state estimate $\boldsymbol{x}$ and state variance estimate $\boldsymbol{P}$ is updated by:

$$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{K}_k(\boldsymbol{y}_k - \hat{\boldsymbol{y}}_{k|k-1}) \tag{9.27}$$

$$\hat{\boldsymbol{P}}_{k|k} = \hat{\boldsymbol{P}}_{k|k-1} - \boldsymbol{K}_k\hat{\boldsymbol{R}}_{k|k-1}\boldsymbol{K}_k^T \tag{9.28}$$

where $\boldsymbol{K}_k$ is the so-called Kalman Filter gain at grid point k calculated by:

$$\boldsymbol{K}_k = \hat{\boldsymbol{P}}_{k|k-1}\boldsymbol{C}_k^T\hat{\boldsymbol{R}}_{k|k-1}^{-1} \tag{9.29}$$

The predictions of states $\boldsymbol{x}$, outputs $\boldsymbol{y}$, state variances $\boldsymbol{P}$ and output variances $\boldsymbol{R}$ then become respectively:

$$\hat{\boldsymbol{x}}_{k+1|k} = \boldsymbol{A}\hat{\boldsymbol{x}}_{k|k} + \boldsymbol{B}\boldsymbol{u}_k \tag{9.30}$$

$$\hat{\boldsymbol{y}}_{k+1|k} = \boldsymbol{C}\hat{\boldsymbol{x}}_{k+1|k} \tag{9.31}$$

$$\hat{\boldsymbol{P}}_{k+1|k} = \boldsymbol{A}\hat{\boldsymbol{P}}_{k|k}\boldsymbol{A}_k^T + \boldsymbol{Q} \tag{9.32}$$

$$\hat{\boldsymbol{R}}_{k+1|k} = \boldsymbol{C}\hat{\boldsymbol{P}}_{k+1|k}\boldsymbol{C}^T + \boldsymbol{S} \tag{9.33}$$

9.1.11.2 Covariance Matrices and Initial State Estimations

Although representing the covariance matrices of the state noise and measurement noise respectively, $\boldsymbol{Q}$ and $\boldsymbol{S}$ are often used to tune the performance of the Kalman Filter. This is done because there is often no information available to evaluate the structure and correlations in the covariance matrices. Furthermore by manually tuning the covariance matrices by choosing the structure and numeric values in the covariance matrices, different properties of the underlying model can be exploited *e.g.* prediction or estimation of filter states. In this case the Kalman Filter serves different purposes during fed- batch and continuous operation. In fed-batch operation the aim is to provide a good estimation of ΔW and subsequently of $\frac{\Delta V}{\Delta t}$, while in continuous operation good estimates of both ΔW and F_{out} are desired.

The following matrices have been used in the initialization of the Kalman Filter for batch and fed-batch operation:

$$\boldsymbol{Q} = \begin{bmatrix} 0.0001 & 0 & 0 \\ 0 & 10 & 0.01 \\ 0 & 0.01 & 0.0001 \end{bmatrix} \quad ; \quad \boldsymbol{S} = \begin{bmatrix} 0.01 \end{bmatrix} \tag{9.34}$$

$$\boldsymbol{P}_{0|0} = \begin{bmatrix} 0.001 & 0 & 0.001 \\ 0 & 0.01 & 0.001 \\ 0 & 0.001 & 0.01 \end{bmatrix} \quad ; \quad \boldsymbol{x}_{0|0} = \begin{bmatrix} 0 \\ 0 \\ W_0 \end{bmatrix} \tag{9.35}$$

The choice of values in $\boldsymbol{Q}$ indicates that most of the information is carried in the state ΔW corresponding $\boldsymbol{Q}(2,2) = 10$. There is a small cross correlation between ΔW and W as illustrated by $\boldsymbol{Q}(2,3) = \boldsymbol{Q}(3,2) = 0.01$. $\boldsymbol{S} = 0.01$ means that the measurement of the weight is accurate, and signifies that the residual formed by $W - \hat{W}$ is relevant in updating the filter states, $\boldsymbol{x}$.

The Kalman Filter is reinitialized at the change to continuous operation at sampling point l. This is done since the dynamics of the process instantly changes as ejection of culture broth is initiated. The following matrices have been found to provide a good description of the change to as well as throughout continuous operation:

$$\boldsymbol{Q} = \begin{bmatrix} 1 & -0.1 & -1 \\ -0.1 & 1 & 10 \\ -1 & 10 & 0.0001 \end{bmatrix} \quad ; \quad \boldsymbol{S} = \begin{bmatrix} 400 \end{bmatrix} \tag{9.36}$$

$$\boldsymbol{P}_{l|l} = \begin{bmatrix} 30 & 30 & -30 \\ 30 & 30 & 1 \\ -30 & 1 & 130 \end{bmatrix} \quad ; \quad \boldsymbol{x}_{l|l} = \begin{bmatrix} x^2_{l-1|l-1} \\ x^1_{l-1|l-1} \\ x^3_{l-1|l-1} \end{bmatrix} \tag{9.37}$$

Compared to the value of $\boldsymbol{S}$ used during fed-batch operation (equation 9.34) the value of $\boldsymbol{S}$ listed in equation 9.36 is 40000 times larger. This means that a smaller amount of the residual formed by W-$\hat{W}$ is relevant in updating the filter states, $\boldsymbol{x}$. The values in $\boldsymbol{Q}$ reflect the structure of the model during continuous operation. Most of the covariance is carried in $\boldsymbol{Q}(2,3)$ and $\boldsymbol{Q}(3,2)$, which accounts for the correlation between ΔW and W.

A very important feature in the reinitialization of the Kalman Filter are the values of $\boldsymbol{x}_{l|l}$ in equation 9.37. The new states are initialized with the values of the last state prior to the change in operation, however the last state of $x^1_{l-1|l-1}$ is used to initialize the new state of $x^2_{l|l}$, while the last state of $x^2_{l-1|l-1}$ is used to initialize the new state of and $x^1_{l|l}$. This serves the purpose of reflecting how the change to continuous operation is characterized by the sudden ejection of culture broth *i.e.* a sudden and large value of F_{out} at the expense of ΔW.

The Kalman Filter then provides the filtered estimates:

$$\begin{bmatrix} F^m_{out,k} \\ \Delta W \\ W_k \end{bmatrix} = \begin{bmatrix} x^1_{k|k} \\ x^2_{k|k} \\ x^3_{k|k} \end{bmatrix} \tag{9.38}$$

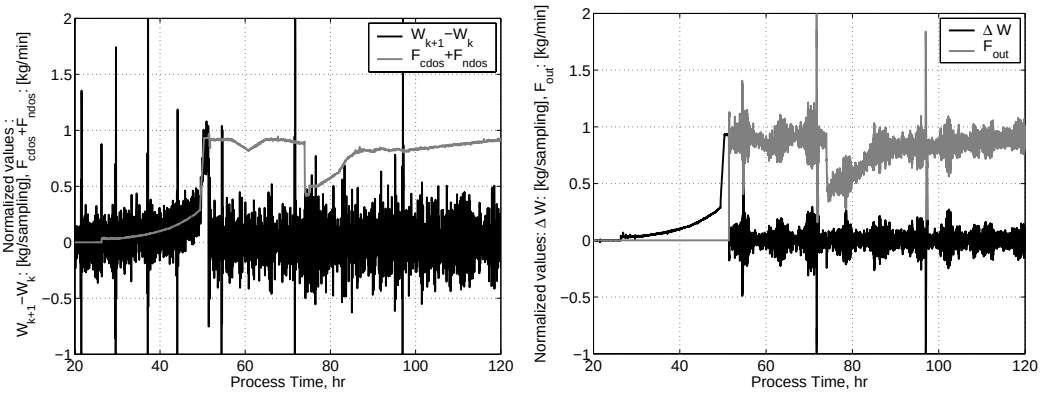
which is used to determine:

$$\left(\frac{\Delta V}{\Delta t} \right)_k = \frac{1}{\rho} \frac{\Delta W_k}{\Delta t_k} \tag{9.39}$$

9.1.11.3 Performance of Kalman Filter

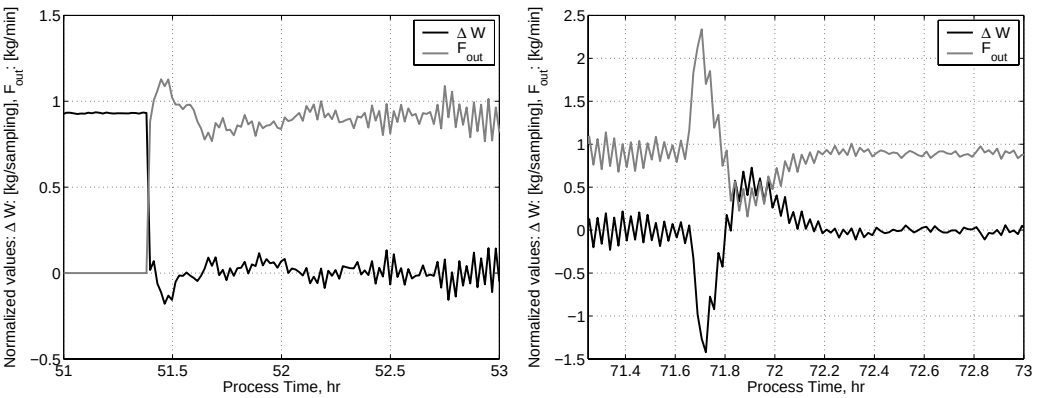
Figure 9.5 illustrates the effects of using a Kalman Filter to estimate ΔW and F_{out}. Roughly speaking the combined information of the two signals in figure 9.5(a) have been distributed between the estimates of ΔW and F_{out} in figure 9.5(b). Despite the fact that the two signals in figure 9.5(b) seem to be rather noisy, the fluctuations are the result of using the ejection strategy discussed earlier.

Figure 9.6 highlights two interesting time periods during the cultivation. Figure 9.6(a) illustrates the change from fed-batch to continuous operation at 51.4 hours, where the effect of the reinitialization of the filter states can be seen as the sudden



(a) Original data (b) Filtered data

Figure 9.5. Performance of Kalman Filter. (a): Data sequences showing W_{k+1}-W_k (black line) and $F_{cdos} + F_{ndos}$ (grey line), where Butterworth filtered values have been used. (b): Kalman Filter estimates of ΔW and F_{out} for the raw data sequences. Data in both plots are from fed-batch and continuous operation in cultivation MTS01.



(a) Onset of continuous operation (b) Sampling of broth

Figure 9.6.

change in the two signals. Due to the strategy used the estimations are instantly updated and no time delay due to the filter is seen. Figure 9.6(b) shows the response in the estimates of ΔW and F_{out} to sampling of culture broth. The sampling occurs from 71.67 to 71.72 hours corresponding to 3 min.[3] It is interesting to see that the response to the sampling of broth is a small overshoot and a settling time of approximately 0.5 hours to settle to the state prior to the sampling. This is a consequence of the parameter choice in the tuning matrices $\boldsymbol{S}$ and $\boldsymbol{Q}$, where a trade-off was made between a fast response time and attenuation of the fluctuations due to the ejection mechanism.

9.1.12 Summary

The previous sections have provided a description of the methods and materials used in order to provide intrinsic information of the process in question, the tools to analyze this information and the framework for developing models, while taking into account the process dynamics introduced by the configuration of the equipment used at the industrial pilot plant.

The information, tools and framework are now set for the more specific modelling of the process dynamics related to the growth, intrinsic flux distribution and product formation of an industrial recombinant strain of *S. cerevisiae*, which is the topic of the following section.

[3]The explanation for the relative long sampling time is that prior to sampling of the broth for analytical analysis, a small sample is taken for an at-line measurement of pH, which can take 1-2 min.

9.2 Soft Sensor for Conversion Rates of Biomass and Acetate

In the introduction to this chapter it was mentioned how quantitative information on key variables in industrial cultivation processes is often not available. It is not only the concentration of the cells, often referred to as the biomass concentration, that is relevant to monitor and control, but also the level of activity *i.e.* whether the cells are dead or alive, producing or not producing the desired product.

Whereas a number of analytical technologies have emerged for the online monitoring of the biomass concentration in the past two decades (for a review see Olsson and Nielsen (1997)), the activity of the biomass is very difficult to measure. Complicating the problem even more is the use of complex substrates in the cultivation industry, which limits the number of eligible technologies for online measurement of the biomass concentration to a few and rather expensive instruments.

This section will illustrate the construction of a simple model to provide online estimates of the conversion rates and concentrations of biomass and acetate. The importance of estimating the biomass concentration has been discussed above, while the conversion of acetate will be shown later to have an important role in the onset of the fermentative metabolism during aerobic growth with glucose as limiting substrate. Experiences obtained in previous investigations will be presented and discussed, and followed by the formulation of a simple threshold model describing the transition between two scenarios: purely oxidative growth and oxido-reductive growth. At the end of this section an algorithm for the calculations behind the model used in the soft sensor is presented.

9.2.1 Model Construction

The system considered is illustrated in figure 9.7. A number of species are assumed to describe the behavior of the system. These species are listed in table 9.2, although glycerol will not be considered here.

Based on the success and experience reported in chapter 6, the same approach was applied here, using another recombinant strain, but resulted in surprising observations, mainly because the strain produced and consumed large amounts of acetic acid under certain operating conditions, as will be presented in section 9.5.1.

In chapter 8 the effects influencing the ammonia flow rate during balanced growth were investigated and modelled. There it was assumed that only cellular activities related to the formation of biomass contributed to the proton flux balance. As a result a large discrepancy between measured and estimated ammonia flow rates could be observed during process upsets. Analysis of the composition of the culture broth revealed significant concentrations of acetate to be present during the time periods where the large discrepancies between measured and estimated ammonia flow rates were observed. The elevated ammonia flow rate can then be interpreted to be a result of the combined outflow of protons from the cells due to growth and production of acetic acid by the cells followed by excretion into the broth. Excretion of acetic acid during conditions of glucose repression is reported to occur as passive diffusion of the undissociated form of the acid (see chapter 4). The excretion of the acid results in an acidification of the broth, which is countered by increased flow of

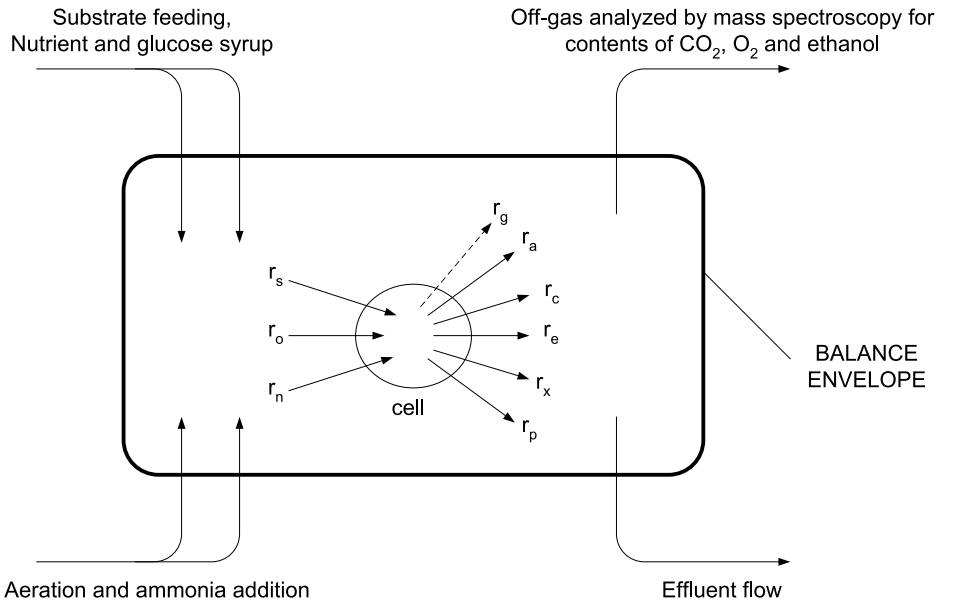


Figure 9.7. Balance envelope of bioreactor and cell. Symbols are listed in table 9.2. Adapted with modifications from Herwig *et al.* (2001).

ammonia to maintain a constant pH of the culture broth.

The model framework proposed by Lei (2001) could easily be extended to construct a soft sensor for online estimation of both biomass and acetate concentrations in the bioreactor. By using a balance envelope similar to the one shown in figure 9.7 a proton flux balance can be formulated. Contributions to the macroscopic flux balance are feed, aeration, off-gas and effluent streams to and from the bioreactor. The ammonia requirement, r_n, to neutralize the combined effect of production of biomass, r_x, and conversion of acetate, r_a, is found by subtracting the effects of other

Table 9.2. An overview of the species considered for the description of biomass and acetate formation. The *source* indicates where the species enter, appear or are measured in the system, while *information* reflects whether the species are measured online or offline. The last column indicates, if the species are considered to be known or unknown in the model. The unit of r_i is mole/L/hr. *: Ethanol in the broth is estimated using the measurements of ethanol in the offgas (see section 9.1.4).

Species	Symbol	Source	Information	Known/Unknown
Glucose	r_s	liquid feed	Offline	Known
Glucose	-	broth	Offline	Unknown (= 0)
Ethanol	-	offgas	Online	Known
Ethanol	r_e	broth	Offline	Known*
CO_2	r_c	offgas	Online	Known
O_2	r_o	offgas	Online	Known
Ammonia	r_n	gaseous feed	Online	Known
Biomass	r_x	broth	Offline	Unknown
Acetate	r_a	broth	Offline	Unknown
Glycerol	r_g	broth	Offline	Unknown

contributions to the proton flux balance from the measured ammonia flow rate[4]:

$$r_n = \alpha_x r_x + \alpha_a r_a = \frac{1}{V}\left(F_{NH_3} + \beta_e \cdot F_{out}^m + \beta_b \cdot \frac{\Delta W}{\Delta t} \right) \tag{9.40}$$

The parameters β_e and β_b are the buffer equivalents of the effluent (e) and culture broth (b) respectively, and were identified and estimated in chapter 8; the values are listed in table 9.3. Equation 9.40 results from a proton flux balance for neutralization of acidification due to metabolic activities, where r_i is the volumetric conversion rate of component i. The two α parameters in equation 9.40 represents the contributions to the proton flux balance from the conversion of biomass (α_x) and acetate (α_a). Assuming a constant concentration of extracellular nitrogen components during pure oxidative continuous cultivation (no conversion of acetate), a nitrogen balance yields that α_x is equal to the elemental nitrogen content of the biomass, cx, which for this strain is found to be $cx = \alpha_x = 0.146$ as mentioned in section 9.1.2 (see chapter 7 for details). The parameter α_a can be interpreted as the amount (mole) of ammonia required to neutralize the effects of excretion of one C-mole acetic acid into the broth. Considering only the acidification effect, 0.5 mole ammonia would neutralize the effect of one C-mole acetic acid, so a value of $\alpha_a = 0.5$ is used as an initial guess. The choice of value for α_a is not obvious, since a number of phenomena are related to the production and excretion of acetate including ATP and NAD(P)H balances.

Table 9.3. Estimated parameters in equation 9.40. The column *St.dev* lists the standard deviation of the estimated parameter in percentage of the parameter value.

Parameter	Estimate	St.dev	Unit
β_e	$1.6 \cdot 10^{-3}$	$\pm$ 45%	mmole NH_3/g eff
β_b	$1.5 \cdot 10^{-3}$	$\pm$ 54%	mmole NH_3/g broth

In chapter 8 the details of a model describing the proton flux balance during balanced oxidative growth was described *i.e.* when no acetate was converted. The validity of that model was demonstrated by accurate estimation of ammonia flow rates during fed-batch and continuous operation, where balanced oxidative growth conditions were known to exist. During operations where balanced oxidative growth did not exist, discrepancies between estimated and measured ammonia flow rates were observed. Based on these observations it was discussed how the model of the proton flux balance can be applied for monitoring of deviations from balanced oxidative growth conditions by comparing the volumetric flow rate of ammonia (r_n) to the estimated volumetric flow rate of ammonia ($\hat{r}_n$). It was suggested that detection limits should be based on measurements of the specific ammonia flow rate , $q_n = r_n/c_x$ (c_x being the biomass concentration) and should be $e_n = q_n - \hat{q}_n = \pm$ 1.5 mmole NH_3/mole biomass/hr. The detection limits were introduced to account for unmodelled effects such as deviations from the assumption of balanced growth or the dynamics of the pH control loop, regulating the flow rate of ammonia to the system.

The detection limit can be used as a threshold value to distinguish between the

[4]For a discussion on how to estimate $\frac{\Delta W}{\Delta t}$ and F_{out}, refer to section 9.1.10

two scenarios:

$$\begin{aligned} &\text{Scenario 1:} \quad r_n = \alpha_x r_x \\ &\text{Scenario 2:} \quad r_n = \alpha_x r_x + \alpha_a r_a \end{aligned}$$

The use of such a threshold will render the model inappropriate to describe a small continuous production of acetate within the detection limits, however the model in its current form, is not suitable for a more accurate description of the ammonia demand to facilitate a more precise modeling of the acetate production This was discussed in chapter 8. An improved description can be obtained by addressing some of the model assumptions and by providing and including a better description of changes of the elemental composition of the biomass or even to changes in the macromolecular (protein, carbohydrate, lipid *etc.*) distribution of the yeast.

A carbon flux balance of the system can be set up:

$$r_s = r_c + r_e + r_a + r_x \tag{9.41}$$

Assuming α_x and α_a to be constant and known parameters, the two equations 9.40 and 9.41 contains only 2 unknown variables, which therefore can be estimated:

$$r_a = \frac{1}{1 - \alpha_a/\alpha_x}(r_s - r_c - r_e - r_n/\alpha_x) \tag{9.42}$$

and subsequently the conversion of biomass can be determined by use of equation 9.41.

It is furthermore assumed that no degradation of biomass occurs *i.e.* cells dying or in other ways becoming inactive, which means that $r_x \geq 0$. Negative values of the acetate production r_a can appear, indicating uptake and consumption of acetate as passive diffusion across the cell wall of acetic acid, a topic reviewed in chapter 4.

Combining equations 9.40, 9.41 and 9.42 and the constraints on r_x and r_a mentioned above, with mass flux balances and approximated integration of these (see section 9.1.10 for further details), it becomes possible to estimate concentration trajectories of the involved components. Further constraints can be included to ensure that the estimated concentrations are positive or zero. From this a soft sensor for the determination of biomass and acetate conversion rates and concentrations is constructed, which is based on the assumptions listed in table 9.4. The algorithm of the biomass and acetate soft sensor is listed in the box entitled Algorithm 9.1.

Table 9.4. Assumptions behind the soft sensor for determination of biomass and acetate conversion rates (equations 9.40, 9.41 and 9.42). The assumptions are listed according to priority with the most important assumption at the top.

1	Asynchronous balanced growth
2	Constant biomass composition
3	Proton balance described by equation 9.40
4	Constant values of α_x and α_a
5	Threshold value can be used to distinguish between balanced growth conditions
6	Consumption of acetate and ethanol contribute to formation of biomass
7	Transition from production to consumption of acetate and ethanol is instantaneous and vice versa
8	Formation of biomass: $r_x \geq 0$
9	Concentrations are always ≥ 0
10	Only fluxes of acetate, ammonia, biomass, ethanol, glucose, CO_2 and O_2 are significant
11	Additive model structure can be used
12	Negligible offset or bias in measurements of ammonia flow rate
13	No significant variation in (constant) buffer capacity of the broth
14	No significant effect on proton balance from feeding of nutrient substrate

Algorithm 9.1
Soft Sensor for Biomass and Acetate Conversion, BAC

At sampling point k

Calculate $\frac{\Delta V}{\Delta t}|_k$ and $F_{out,k}$ using the Kalman Filter algorithm

Calculate $r_{c,k} = \mathrm{CER}_k/V_k$ and $r_{o,k} = \mathrm{OUR}_k/V_k$

Calculate $r_{s,k} = c^m_{cdos} F^m_{cdos,k}/M_{glu}/V_k$

Calculate $r_{e,k} = 1/120\left(\frac{\Delta c^g_{e,k}}{\Delta t_k} + \left(F_{out,k}/V_k + 1/V_k \frac{\Delta V}{\Delta t}|_k\right) c^g_{e,k}\right)/M_{eth}$

Calculate $r_{n,k} = \frac{1}{V_k}\left(F_{NH_3,k} - \rho\beta_e F_{out,k} + \rho\beta_b \frac{\Delta V}{\Delta t}|_k\right)$

Calculate $\hat{r}_{n,k} = \alpha \cdot c^m_{glu,cdos} \cdot \frac{Y_{sx}}{M_{glu}} F^m_{cdos,k}$

Check if $|r_{a,k} - \hat{r}_{a,k}|/c_{x,k-1} < 0.0015$ then $r_{a,k} = 0$

else $r_{a,k} = \frac{1}{1-\alpha_a/\alpha_x}(r_{s,k} - r_{c,k} - r_{e,k} - r_{n,k}/\alpha_x)$

Check if $r_{a,k} < 0$ and $c_{a,k-1} \leq 0$ then $r_{a,k} = 0$

Calculate $r_{x,k} = r_{s,k} - r_{c,k} - r_{e,k} - r_{a,k}$

Check if $r_{x,k} < 0$ then $r_{x,k} = 0$

Calculate $c_{a,k} = c_{a,k-1} + \left[M_{ace} \cdot r_{a,k} - \left(F_{out,k}/V_k + 1/V_k \frac{\Delta V}{\Delta t}|_k\right) \cdot c_{a,k-1}\right] \cdot (t_k - t_{k-1})$

Calculate $c_{x,k} = c_{x,k-1} + \left[M_{DW} \cdot r_{x,k} - \left(F_{out,k}/V_k + 1/V_k \frac{\Delta V}{\Delta t}|_k\right) \cdot c_{x,k-1}\right] \cdot (t_k - t_{k-1})$

9.2.2 Summary

This section has presented a model for online estimation of the conversion rates of biomass and acetate as well as their concentrations in the culture broth. The model is based on a simple extension of the methodology presented by Lei (2001) for online estimation of the biomass concentration and tested in chapter 6 on process data from a production site. It was presented how the simple extension was required

in order to account for a large unforeseen production of acetate during the onset of oxido-reductive growth.

With estimates of the biomass concentration available online, it is now possible to evaluate specific uptake and conversion rates of a number of species. Such information can be compiled in a new model in order to provide an online estimate of the intrinsic metabolic flux distribution in the cells. In this way aspects of the physiology of the microorganism can be used for the online interpretation of process data.

9.3 Soft Sensor for Intrinsic Metabolic Fluxes

This section will illustrate the construction of an intrinsic metabolic flux model to provide online estimates of fluxes of the most relevant metabolic pathways related to growth during aerobic conditions. The model structure and model parameters are discussed followed by providing an algorithm for the calculations behind the model at the end of the section.

9.3.1 Model Construction

To elucidate how the intrinsic metabolic fluxes are distributed, a simple metabolic model is proposed, as schematically illustrated in figure 9.8. The stoichiometry of the different reactions are presented in table 9.5, where it can be seen that a number of physiological parameters are required to be determined or assumed.

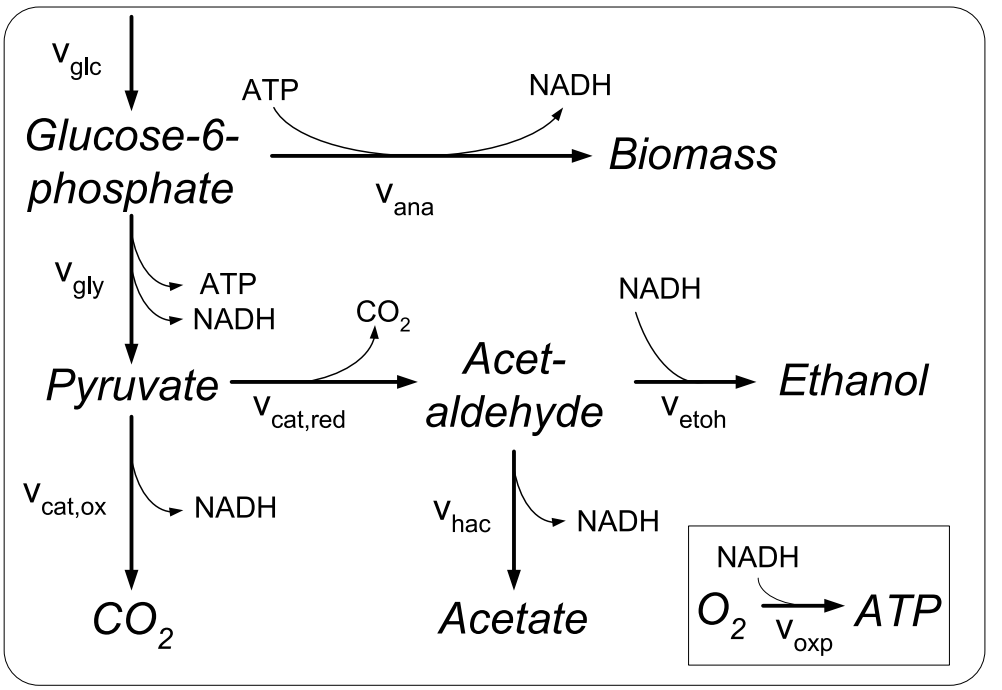


Figure 9.8. Simple model for metabolic flux analysis. Subscript i indicates the reaction, while v_i represents the specific intracellular flux via reaction i. Glucose from the extracellular medium is taken up through reaction *glc*.

In the metabolic flux model the glucose taken up from the extracellular medium

is converted into glucose-6-phosphate by reaction *glc*. Glucose-6-phosphate is then converted into biomass by reaction *ana*, representing lumped anabolic reactions, and into pyruvate by reaction *gly*, representing lumped glycolytic reactions. Pyruvate can either be dissimilated into carbon dioxide in the tricarboxylic acid (TCA) cycle, also referred to as the oxidative catabolism represented by reaction cat, ox, or converted into acetaldehyde by the reductive catabolism, reaction cat, red. Acetaldehyde can either be oxidized to acetate/acetic acid by reaction *hac* or reduced to ethanol by reaction *etoh*. In addition, the activity of the oxidative phosphorylation for the generation of ATP by oxidation of NADH by oxygen has been included, reaction *oxp*.

Table 9.5. The stoichiometry in [C-mole/C-mole] of the reaction in the proposed intrinsic metabolic flux model of figure 9.8. Most of the stoichiometric coefficients have been found in biochemical textbooks (Stryer (1995); Walker (1998)). A number of parameters are dependent on the microorganism or strain and listed below with references to their source and are discussed in the text. Glucose-6-phosphate is referred to as CH_2O (int).

Anabolism		
ana	:	$X + \gamma CO_2 + Y_{xNADH} X_{NADH} - Y_{xATP} X_{ATP} - (1+\gamma) CH_2O - \gamma_N NH_3 = 0$
Catabolism		
glc	:	$CH_2O(int) - CH_2O(ext) = 0$
gly	:	$CH_{4/3}O + 1/3\, X_{ATP} + 1/3\, X_{NADH} - CH_2O_{int} = 0$
cat, ox	:	$CO_2 + 1/3\, X_{ATP} + 5/3\, X_{NADH} - CH_{4/3}O = 0$
cat, red	:	$2/3\, CH_2O_{1/2} + 1/3\, CO_2 - CH_{4/3}O = 0$
etoh	:	$CH_3O_{1/2} - 1/2\, X_{NADH} - CH_2O_{1/2} = 0$
hac	:	$CH_2O + 1/2\, X_{NADH} - CH_2O_{1/2} = 0$
Oxidative phosphorylation		
oxp	:	$P/O\, X_{ATP} - X_{NADH} - 1/2\, O_2 = 0$
Parameters		
X	:	$CH_{1.82}O_{0.576}N_{0.146}$ $M_{DW} = 27.1$g/C-mole (f_r: 7.3 w%)
γ	:	0.17 (calculated from Y_{xNADH})
γ_N	:	is equal to $cx = 0.146$ (see chapter 7)
P/O	:	1.20 (van Gulik and Heijnen (1995))
Y_{xNADH}	:	0.23 (Duboc *et al.* (1998); Verduyn *et al.* (1990))
Y_{xATP}	:	1.7 (calculated in section 9.5.2)

In a number of the reactions in the metabolic flux model the cofactors ATP and NADH are involved, as indicated in figure 9.8. When investigating steady state behavior of continuous cultivations it is assumed that the concentrations of ATP and NADH are constant and therefore can be used as constraints in the system of equations (Lei (2001); Vanrolleghem *et al.* (1996); van Gulik and Heijnen (1995); Herwig and von Stockar (2002)(only NADH)). In this work such an approach has not been used since the metabolic flux model is used to study the response of the culture to dynamic changes in the system. By not enforcing constraint on the concentration of ATP and NADH, information on the variation of the cofactor levels can be used as an indication of how much the metabolic energy and reductive requirements for cellular activities change during the varying process conditions of fed- batch and

Table 9.6. Equations behind the intrinsic metabolic flux model. Specific conversion rates of biomass (q_x) and acetate (q_a) as well as the biomass concentration (c_x) are calculated using the soft sensor for biomass and acetate conversions (BAC) presented in section 9.2, algorithm 9.1. Glucose-6-phosphate is referred to as CH_2O (int), and the intracellular concentration of glucose-6-phosphate, pyruvate and acetaldehyde are assumed to be quasistationary. $max(q_i, 0)$ means that only positive values of q_i is allowed to affect the calculation of a quantity. The values of the parameters γ, P/O, Y_{xATP} and Y_{xNADH} are listed in table 9.5.

Specific conversion rates			**Origin**
Biomass	:	$q_x = r_x/c_x$	r_x and c_x from BAC
Acetate	:	$q_a = r_a/c_x$	r_a from BAC
Glucose	:	$q_s = r_s/c_x$	$r_s = c^m_{cdos} F^m_{cdos}/M_{glu}/V$
Ethanol	:	$q_e = r_e/c_x$	r_e from equation 9.2
CO_2	:	$q_c = \mathrm{CER}/c_x$	CER from equation 7.3
O_2	:	$q_o = \mathrm{OUR}/c_x$	OUR from equation 7.2
Quasistationary flux balances			
CH_2O (int)	:	$0 = -v_{glc} + (1+\gamma)v_{ana} + v_{gly}$	
Pyruvate	:	$0 = -v_{gly} + v_{cat,ox} + v_{cat,red}$	
Acetaldehyde	:	$0 = -v_{cat,red} + 3/2 v_{etoh} + 3/2 v_{hac}$	
Intrinsic metabolic fluxes			
ana	:	$v_{ana} = q_x$	
glc	:	$v_{glc} = q_s$	
gly	:	$v_{gly} = q_s - (1+\gamma)q_x$	
cat, ox	:	$v_{cat,ox} = q_c - \gamma q_x - 1/2(max(q_a,0) + max(q_e,0))$	
cat, red	:	$v_{cat,red} = 3/2(max(q_a,0) + max(q_e,0))$	
$etoh$	:	$v_{etoh} = q_e$	
hac	:	$v_{hac} = q_a$	
oxp	:	$v_{oxp} = 2q_o$	
Cofactor fluxes			
ATP	:	$v_{atp} = P/O v_{oxp} + 1/3\, v_{gly} + 1/3\, v_{cat,ox} - Y_{xATP} v_{ana}$	
NAD(P)H	:	$v_{nadh} = -v_{oxp} + 5/3\, v_{cat,ox} + 1/3\, v_{gly} + Y_{xNADH} v_{ana} - 1/2\, v_{etoh} + 1/2\, v_{hac}$	

continuous operation as well as during process disturbances.

It should be noted that only two of the reactions in the metabolic flux model are allowed to be reversible, namely the reactions to ethanol and acetate. When either of these are negative, they do not contribute to the estimation of the flux through the reductive catabolism, $v_{cat,red}$, as the reaction, pyruvate decarboxylase catalyzing the conversion of pyruvate to acetaldehyde is irreversible (Pronk *et al.*, 1996). Instead the fluxes v_{hac} and v_{etoh} contribute directly to the flux of biomass formation v_{ana}, representing growth on acetate and ethanol. These special cases are not shown in figure 9.8; they have been left out to keep the figure simple, as in most cases the contribution from these fluxes to the overall biomass formation rate is negligible. No time delay, also referred to as lag, to the change between different metabolisms have been included in the model.

For the recombinant strain used in this project no information is available on the parameters listed at the bottom of table 9.5. In chapter 7 the biomass composi-

tion was estimated based upon measurements of the elemental composition combined with macroscopic mass balances of carbon, nitrogen, oxygen and hydrogen: $CH_{1.82}O_{0.576}N_{0.146}$ (f_r: 7.3 w%). γ and Y_{xNADH} are highly correlated through the redox balance of v_{ana}: $\gamma = (\kappa_x + 2Y_{xNADH})/4 - 1$, with $\kappa_x = 4.23$ based on the estimated elemental composition of the strain described in section 7, Y_{xNADH} can be determined from anaerobic experiments as described by Duboc *et al.* (1998) using data from Verduyn *et al.* (1990), however such experiments have neither been carried out in this work, nor by others with this strain, hence the value reported by Duboc *et al.* (1998) is also being used here: $Y_{xNADH} = 0.23$. This leads to $\gamma = 0.17$. The P/O ratio accounts for the net output of moles ATP produced per mole oxygen atom consumed in oxidative phosphorylation, v_{oxp}. A constant value of P/O = 1.20 is used as reported by van Gulik and Heijnen (1995) leaving the Y_{xATP} to be determined by setting up an ATP balance using the glucose and oxygen consumption rates to estimate the fluxes v_{ana} and v_{gly} under the assumption of purely oxidative growth ($v_{cat,red} = 0$ and $v_{cat,ox} = v_{gly}$):

$$0 = -Y_{xATP} \cdot v_{ana} + (2 \cdot v_{gly} + Y_{xNaDH} \cdot v_{ana}) \cdot P/O + 2/3 \cdot v_{gly} \tag{9.43}$$

where it has been used that $NADH$ fuels the oxidative phosphorylation, v_{oxp}. The factor 2/3 is the overall ATP produced in the glycolysis (v_{gly}) and TCA cycle ($v_{cat,ox}$) when converting one C-mole glucose to CO_2.

$$q_o = 2 \cdot v_{gly} + Y_{xNaDH} \cdot v_{ana} \tag{9.44}$$

$$q_s = (1 + \gamma) \cdot v_{ana} + v_{gly} \tag{9.45}$$

where the factor 2 in the equation 9.44 is the overall NADH produced in the glycolysis (v_{gly}) and TCA cycle ($v_{cat,ox}$) when converting one C-mole glucose to CO_2.

Using equations 9.44 and 9.45 to estimate v_{gly} and v_{ana} yields:

$$v_{gly} = 1 - (1 + \gamma) \cdot \frac{1 - Y_{so}}{1 + \gamma - 1/2 \cdot Y_{xNADH}} \cdot q_s \tag{9.46}$$

$$v_{ana} = \frac{1 - Y_{so}}{1 + \gamma - 1/2 \cdot Y_{xNADH}} \cdot q_s \tag{9.47}$$

where $Y_{so} = \frac{q_o}{q_s}$.

Y_{xATP} can be determined using equation 9.43 with either measured fluxes of v_{gly} and v_{ana} or estimated fluxes using equations 9.46 and 9.47.

The individual fluxes illustrated in figure 9.8 are calculated based on flux balances using specific conversion rates (q_i) of a number of species (i), which are either measured or estimated. Furthermore it is assumed that the concentrations of the pools of glucose-6-phosphate, pyruvate and acetaldehyde are quasistationary, meaning that they do not change significantly over time. The equations behind the model of intrinsic fluxes are listed in table 9.6. The assumptions made for the construction of the model leading to the soft sensor of intrinsic metabolic fluxes are listed in table 9.7, while the algorithm is presented in the box entitled Algorithm 9.2.

Table 9.7. Assumptions behind the soft sensor for intrinsic metabolic fluxes. The assumptions are listed according to priority with the most important assumption at the top.

1	Asynchronous balanced growth
2	Constant biomass composition
3	Constant values of parameters γ, P/O, Y_{xATP} and Y_{xNADH}
4	Quasistationary intracellular concentrations of glucose-6-phosphate, pyruvate and acetaldehyde
5	Conversion rates of biomass and acetate can be calculated using algorithm 9.1 presented in section 9.2
6	Consumption of acetate and ethanol contribute to formation of biomass
7	Transition from production to consumption of acetate and ethanol is instantaneous and vice versa
8	Formation of biomass: $r_x \geq 0$
9	Only fluxes of acetate, ammonia, biomass, ethanol, glucose, CO_2 and O_2 are relevant
10	Negligible offset or bias in measurements of flow rate

Algorithm 9.2
Soft Sensor for Intrinsic Metabolic Fluxes, IMF

At sampling point k

From Algorithm 9.1 obtain: $r_{i,k} = [r_{c,k}; r_{o,k}; r_{s,k}; r_{e,k}; r_{a,k}; r_{x,k}]$ and $c_{x,k}$

Calculate $q_{i,k} = r_{i,k}/c_{x,k}$

Calculate $v_{ana} = q_x$

Calculate $v_{glc} = q_s$

Calculate $v_{gly} = q_s - (1+\gamma)q_x$

Calculate $v_{cat,ox} = q_c - \gamma q_x - 1/2(max(q_a, 0) + max(q_e, 0))$

Calculate $v_{cat,red} = 3/2(max(q_a, 0) + max(q_e, 0))$

Calculate $v_{etoh} = q_e$

Calculate $v_{hac} = q_a$

Calculate $v_{oxp} = 2q_o$

Calculate $v_{atp} = P/Ov_{oxp} + 1/3\, v_{gly} + 1/3\, v_{cat,ox} - Y_{xATP}v_{ana}$

Calculate $v_{nadh} = -v_{oxp} + 5/3\, v_{cat,ox} + 1/3\, v_{gly} + Y_{xNADH}v_{ana} - 1/2\, v_{etoh} + 1/2\, v_{hac}$

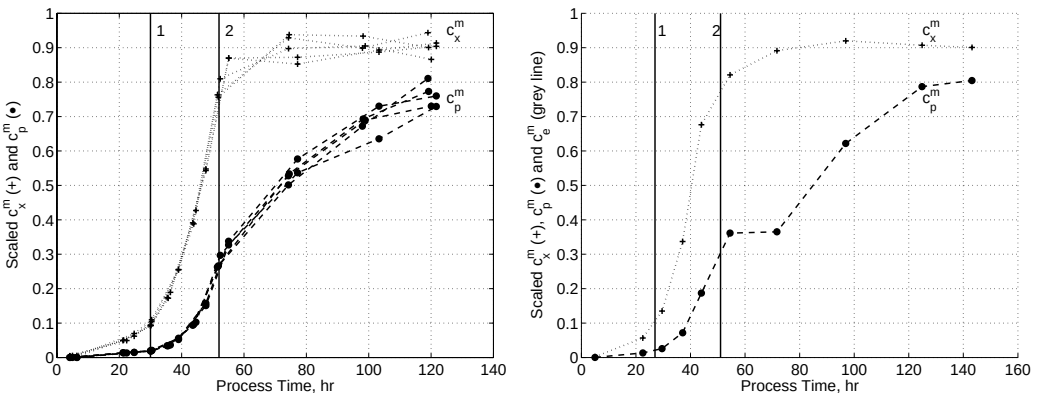
9.3.2 Summary

This section has presented a model for online estimation of the intrinsic flux distribution between some of the major metabolic pathways active during aerobic grotwh on glucose. The model is based on a simple metabolic network and applies online process data in combination with the outputs from the soft sensor in section 9.2. A number of physiological parameters are needed for the model. Some of these parameters were calculated from the available information of the system, while the remaining parameters were taken from literature.

9.4 Soft Sensor for Insulin Precursor Production Rate

The two soft sensors presented above both address issues related to the growth of the microorganism. From an industrial production point of view, this information might not seem as relevant as an actual description of the productivity of the desired product. Therefore a third soft sensor has been constructed for this specific purpose, and this is the topic of the current section.

The expression system of the insulin precursor is known to be linked to the glycolytic activity and is constitutive (Kjeldsen, 2000), so with information on the biomass production rate predicted by the soft sensor presented in section 9.2, only the specific productivity of the insulin precursor needs to be modeled. An indication that the specific productivity is not constant throughout the process is obtained from figure 9.9(a) showing scaled trajectories of the insulin precursor and biomass concentrations obtained from offline measurements in normal cultivations. It is important and surprising to notice that the specific productivity is not constant throughout the process.



(a) Normal cultivations (b) MTS02

Figure 9.9. Scaled biomass concentration (c_x^m) and insulin precursor concentration (c_p^m). (a) 5 normal cultivations: MTS05, MTS06, MTS07, MTS11 and MTS12. (b) Cultivation MTS02 were ethanol and acetate where observed around the start of the continuous operation. The vertical lines illustrate the phase changes from batch to fed-batch (1) to continuous operation (2). During longer cultivations the scaled insulin precursor concentration is known to level off at a value between 0.75 and 0.85. The growth rate is maintained at a constant value during both fed-batch and continuous operation, and is identical for all the batches.

From figure 9.9(a) it appears that the trajectories of scaled biomass and insulin precursor concentrations are not parallel and therefore the production rates are not proportional. This is best seen during continuous operation, where the biomass concentration reaches a constant level at a scaled value of 0.9 at the onset of continuous operation, while the insulin precursor concentration has not yet reached a constant level after 120 hours of cultivation. From longer continuous cultivations it is known that scaled stationary insulin precursor concentrations between 0.75 and 0.85 are

normally obtained. These findings suggested that the specific production rate of the insulin precursor is not constant throughout the process despite the fact that the expression system of the insulin precursor was known to be controlled by a constitutive promoter (Kjeldsen, 2000). In the construction of an expression system for the recombinant insulin precursor, a promoter of the glycolytic enzyme triose phosphate isomerase (*tpi*) has been used in front of the insulin precursor gene sequence (Kjeldsen *et al.* (2001); see chapter 5 for a detailed description).

The observations indicated that from the start of the fed-batch operation more than 80 hours of aerobic cultivation is needed to obtain the high level of specific productivity, which corresponds to more than 5 doubling times. This is a rather interesting observation, when compared to the expression rate of the insulin precursor reported by Kjeldsen *et al.* (2001) and reviewed in section 5.2.1. Using pulse chase experiments on the insulin precursor with metabolic labelling with [^{35}S]cysteine for 2.5 min followed by use of unlabelled cysteine medium the first labelled insulin precursor appeared in the supernatant 2-4 min after the pulse, and the majority of the secreted IP appeared in the supernatant within 15 min after the pulse, leading to a $t_{1/2}$ in the range of 5-10 min. On the other hand analysis of the intracellular retention of the insulin precursor showed that following 30 min after a metabolic labelling for 2.5 min, approximately 30% of the labelled insulin precursor was still present as processed intracellular insulin precursor, primarily in vacuoles, and was not secreted into the culture broth. This indicated that two different intracellular routes for the insulin precursor are present in the late secretory pathway, and that secretion may reflect saturation of a sorting mechanism due to over-expression of the insulin precursor or that secretion occurs in competition with intracellular retention. This can to some extent explain the slow build up of the specific productivity, however with the expression rates of the pulse chase experiments reported above in the range of minutes and compared to the 80 hours needed to reach the high level of specific productivity, it appears that something else has an influence on the specific productivity. It is currently not known which physiological mechanism could be responsible for the extended time period needed to reach the maximum specific productivity and the findings highlights that it is not a trivial matter to tune the expression system for production of the recombinant product.

To model the observed trajectories of the insulin precursor concentrations during fed-batch and continuous operation up until 150 hours, it is assumed that the following boundary conditions apply for specific productivity:

B1 The specific productivity at the start of fed-batch operation is low and therefore assumed to be $q^m_{p,t_{fb}} = 0$

B2 The specific productivity obtained after 120 hours[5] of combined fed-batch and continuous operation, is proportional to the specific growth rate q_x (or μ) with the proportionality constant α_p (unit: (g-product·kg-broth)/(g-biomass·L-broth)).

The reason for using the specific growth rate rather than the specific glucose uptake rate is supported by the observation made in figure 9.9(b), where the trajectories of the biomass and product concentrations are shown for cultivation MTS02. In this experiment oxido- reductive conditions occurred around the change to continuous

[5]Approximately equal to 150 hours of total cultivation with a batch period of $\approx$ 30 hours.

operation as shown in figure 9.10 (the observations in cultivation MTS02 will be discussed in more detail below). During the period of oxido-reductive growth the specific glucose uptake rate is high, while the specific growth rate is low. Assuming the specific productivity of the product to be proportional to the specific growth rate seems therefore to be a better choice. The model construction is separated in two parts. In the first part only the observations from the normal cultivation in figure 9.9(a) are considered, while in the second part, the observations during oxido-reductive growth or presence of acetate are included.

9.4.1 Modeling the Adaptation of the Specific Productivity

To describe the behavior between the above stated boundary conditions, a first order dynamic model for the adaptation of the specific productivity is proposed:

$$\frac{dq_p^m}{dt} = \frac{1}{\tau}(\alpha_p q_x - q_p^m) \tag{9.48}$$

assuming α_p to be constant, and μ^m (q_x^m) to change only a little $\approx$ constant [6], the equation above can be solved for $t > t_{fb}$ with the assumption $q_{p,t_{fb}}^m = 0$ to yield:

$$q_{p,k}^m = \frac{r_{p,k}^m}{c_{x,k}^m} = \alpha_p q_{x,k}\big(1 - exp\big(\frac{-(t_k - t_{fb})}{\tau}\big)\big) \tag{9.49}$$

where t_{fb} is the starting time of the fed-batch operation, k is the sample number, while τ is the characteristic time for the adaptation to the new specific productivity $\alpha_p q_x$. The value of τ reflects the rate at which the expression system of the microorganism adapts to the new growth conditions during aerobic growth when feeding glucose as the limiting substrate.

Using a constant value of τ in equation 9.49 renders the adaptation of the specific productivity independent of how the microorganism is grown *i.e.* which feeding strategy has been applied. It is therefore desirable to make τ dependent on a variable that is representive of the growth conditions in order to account for different, but similar feeding strategies. In the construction of an expression system for the recombinant insulin precursor, a promoter of the glycolytic enzyme triose phosphate isomerase (*tpi*) has been used in front of the insulin precursor gene sequence (Kjeldsen *et al.*, 2001). *tpi* is an enzyme in the glycolysis and it has therefore been assumed that:

1. the transcription rate of the *tpi* promotor is proportional to the specific growth rate q_x
2. the transcription rate of the *tpi* promotor is proportional to the specific production rate of the insulin precursor

This means that at low dilution rates where the specific growth rate is low, the transcription rate of the *tpi* promotor is low and so is the expression of the insulin precursor. It is further assumed that 3) at lower expression rate the adaptation of q_p

[6] Except for small perturbations to the substrate feed rate μ^m was constant during fed-batch and continuous operation for the cultivations shown in figure 9.9(a).

to the new specific productivity ($\alpha_p q_x$) becomes slower. The final step is to assume that 4) the characteristic time for adaptation, τ_k, is inversely proportional to the specific growth rate q_x:

$$\tau_k = \psi / q_{x,k} \tag{9.50}$$

where ψ is a proportionality constant.

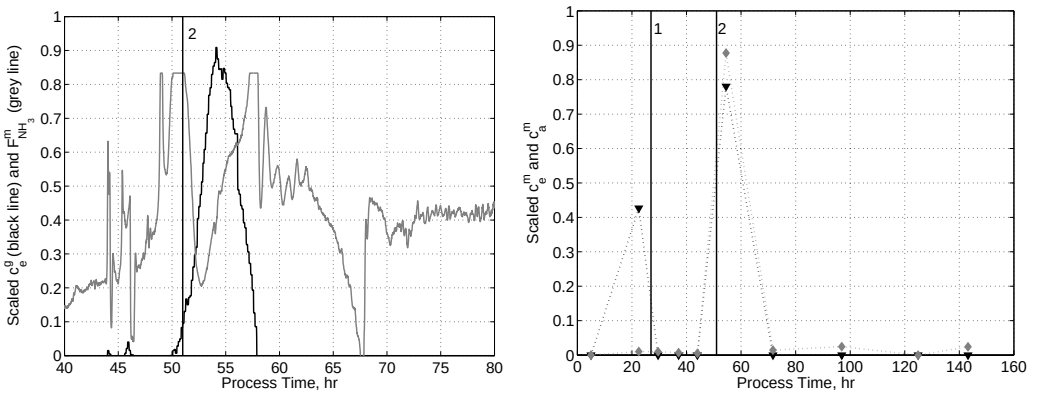
9.4.2 Modeling the Influence of Oxido-reductive Growth on the Specific Productivity

While figure 9.9(a) illustrated the behavior in cultivations that were not showing signs of oxido-reductive growth *i.e.* formation of ethanol or acetate, figure 9.9(b) illustrates the observations made in cultivation MTS02, where significant amounts of ethanol were observed both in the online measurements of the offgas from the bioreactor (figure 9.10(a)) and in the offline analysis of the culture broth using HPLC (figure 9.10(b)) between 50 and 57 hours. Acetate was also observed by the offline analysis at 54.5 hours (figure 9.10(b)). The investigations related to the construction of the soft sensor for biomass and acetate conversions in section 9.2 had used that large fluctuations in the ammonia flow rate signal could be related to the conversion of acetate. Such fluctuations were seen in the ammonia flow rate between 48 and 68 hours in figure 9.10(a), indicating that acetate was present in the culture broth. An interpretation of these observations is that during conversion of acetate, the production of the insulin precursor is slowed or stopped, and sped up or restarted as oxidative growth is reassumed at 68 hours.

The reason for using the conversion of acetate for modeling rather than *e.g.* conversion or presence of ethanol, is due to the observations reported earlier that the estimated conversion of acetate provides a better indication of changes to the growth condition *i.e.* oxidative versus oxido-reductive growth.

Only a very simple model is suggested to account for the effects of acetate on the specific productivity, since the data available for model construction is rather limited and only related to problems occurring at the change from fed-batch to continuous operation. A very simple way to account for the presence of acetate is by delaying the adaptation to the new specific productivity ($\alpha_p q_x$), when acetate is present in the culture broth. From figure 9.9(b) it could be noted that the specific productivity rate seemed to be constant during process upsets *i.e.* the first order behavior in equation 9.48 was replaced by a zero order behavior. In other words, as acetate production is estimated above the threshold mentioned in section 9.2 (1.5 mmole NH_3/mole biomass/hr) a zero order description is used for q_p^m *i.e.* the value of q_p^m is kept constant during the production of acetate. A zero order behavior can be obtained using equation 9.49 by letting t_{fb} increase in parallel with t during the time when acetate is being produced.

While the first order behavior of equation 9.49 can be interpreted as a slow buildup of capacity or filling up *storage* for producing and expressing the insulin precursor, the zero order behavior can be interpreted as a stop of the buildup of the capacity or storage, so the current maximum level of the specific productivity is not increased any further. It could be relevant to include a term for the degradation of capacity, however this has not been attempted in this work.



(a) Online signal

(b) Offline signal

Figure 9.10. Cultivation MTS02. (a) Scaled online signals of ethanol concentration in off-gas (c_e^g, black line) and ammonia flow rate ($F_{NH_3}^m$, grey line) around the onset of continuous operation. (b) Scaled offline measurements of ethanol (c_e^m) and acetate (c_a^m). The vertical lines illustrate the phases change from batch to fed-batch (1) to continuous operation (2)

Combining equations 9.49 and 9.50 a model for describing the adaptation of the insulin production rate becomes:

$$r_{p,k}^m = q_{p,k}^m c_{x,k}^m = \alpha_p r_{x,k}\big(1 - exp\big(\frac{-(t_k - t_{fb})q_{x,k}}{\psi}\big)\big) \tag{9.51}$$

t_k is the time passed from the beginning of the cultivation, r_x, q_x and c_x^m can be obtained from algorithm 9.1 and by combining equation 9.51 with a macroscopic mass balance of the system such as equation 9.8, a soft sensor for insulin precursor production and concentration can be constructed. The algorithm for such a soft sensor is listed in the box entitled Algorithm 9.3. Assumptions used for the construction of the soft sensor are listed in table 9.8.

Algorithm 9.3
Soft Sensor for Insulin Precursor Production Rate and Concentration, IPP

At sampling point k

From Kalman Filter obtain: $F_{out,k}$ and V_k

From Algorithm 9.2 obtain: $q_{x,k}$, $q_{a,k}$ and $c_{x,k}$

Check if $q_{a,k} > 0$ then $t_{fb,k} = t_{fb,k-1} + (t_k - t_{k-1})$

else $t_{fb,k} = t_{fb,k-1}$

Calculate $r_{p,k}^m = \alpha_p r_{x,k}\big(1 - exp\big(\frac{-(t_k - t_{fb,k})q_{x,k}}{\psi}\big)\big)$

Calculate $c_{p,k}^m = c_{p,k-1}^m + \Big[r_{p,k}^m - \Big(F_{out,k}/V_k + 1/V_k \frac{\Delta V}{\Delta t}|_k\Big) \cdot c_{p,k-1}^m\Big] \cdot (t_k - t_{k-1})$

9.4.3 Summary

This section has provided a model for the online estimation of the productivity and concentration of the insulin precursor. The model includes a term for describing

Table 9.8. Assumptions behind the soft sensor for insulin precursor production rate and concentration. The assumptions are listed according to priority with the most important assumption at the top.

1	Asynchronous balanced growth
2	First order model of adaptation valid
3	Characteristic adaptation time τ inversely proportional to the specific growth rate q_x
4	At lower expression rate the adaptation of q_p^m to the new specific productivity becomes slower
5	The transcription rate of the *tpi* promotor is proportional to the specific production rate of the insulin precursor
6	The transcription rate of the *tpi* promotor is proportional to the specific growth rate q_x
7	Specific productivity at the start of fed-batch operation $q^m_{p,t_fb} = 0$
8	Zero order model of adaptation valid when acetate is present in culture broth

the time varying characteristics of the product concentration profile shown in figure 9.9(a), since it was shown that the specific productivity of the product is not proportional to the specific growth rate, although this was expected from the knowledge of the genetic modifications applied to the strain. A simple adjustment of the model during conversion of acetate was suggested in order to describe the observed variation around the change to continuous operation in cultivation MTS02 (figure 9.9(b)).

9.5 Results

The purpose of developing the soft sensors presented in the previous section is to provide tools to improve interpretation of process data. Such tools can be used in the offline analysis of cultivations that for some reason have shown inferior performance or been exposed to process variations or disturbances resulting in interesting responses. An example of this could be planned attempts to improve process performance. A second and very interesting application area for the tools is online monitoring and control. Implementing the soft sensors in a process control system can provide more detailed analysis of the state of the process. This can lead to a totally new way of approaching operator-based control of cultivations, but can also be extended to a more automatized control of the cultivation process in order to enhance its performance.

Using the three soft sensors developed above, a number of interesting observations appear and many of these will be illustrated below. First, a value of α_a will be estimated and the soft sensor in algorithm 9.1, referred to as the *BAC* soft sensor, will be validated on data from two cultivations with focus on the ability of the *BAC* soft sensor to estimate the trajectories of biomass and acetate concentrations. Secondly, the conversion rates of biomass and acetate from the *BAC* soft sensor are used in the proposed soft sensor in algorithm 9.2, referred to as the *IMF* soft sensor, for estimation of intrinsic metabolic fluxes. Data from four cultivations are investigated, analyzed and discussed. Thirdly, results using the soft sensor in algorithm 9.3, referred to as the *IPP* soft sensor, for description of the specific production rate of the insulin precursor is presented and evaluated.

All the data presented and discussed has been normalized.

9.5.1 Modeling Biomass and Acetate Conversion

Data from three cultivations are used in this section to elucidate the performance of the *BAC* soft sensor. Cultivation MTS01 is used as a modeling batch for the identification of α_a, while cultivations MTS05 and MTV03 are used as validation batches.

It should be noted that the acetate conversion rate is estimated using the ammonia flow rate, which serves the purpose of maintaining a constant culture pH. The pH-controller has been tuned to ensure good control throughout the operation regimes of the process. The pH-controller has not been tuned for fast variations, *e.g.* during synchronized growth. Therefore the pH can vary with $\pm$ 0.4 pH-units during such variations, and hence affect the estimation of acetate conversion. In most cases the pH is back at the desired set point within 0.5 hours after the onset of the variation, resulting in a small delay in the estimated acetate conversion rate and concentration.

9.5.1.1 Selection of α_a

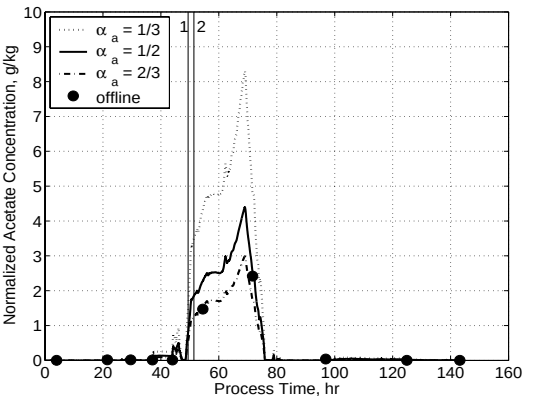
The results obtained using the *BAC* soft sensor on data from cultivation MTS01 are shown in figure 9.11(a+b), where three different values of α_a has been used (α_a was presented in equation 9.40, page 144) . Comparing the estimated trajectories of the acetate concentration in figure 9.11(a) with the offline measurements of acetate, it is difficult to assess which of the values $\alpha_a = 1/2$ or $\alpha_a = 2/3$ provides the better description. The offline measurements of the biomass concentration in figure 9.11(b) are best described using $\alpha_a = 1/3$, where the two other values of α_a result in too large estimations of the biomass concentration. Here it should be noted that significant amounts of glycerol has been observed in the offline measurement at 72 hours as shown in figure 9.11(d)[7], which will have an influence on both the estimation of biomass and acetate conversions. The effect of this influence would be a lowering of the biomass concentration to account for glucose being used for formation of glycerol in equation 9.41, and at the same time an increase in the acetate concentration to balance the ammonia demand in equation 9.40 used in the *BAC* soft sensor. Unfortunately it has not been possible to propose a model for the conversion of glycerol, allowing the estimation of the trajectory of glycerol concentration.

Based on the above observations and the assumption that the sole influence of production of acetate on the ammonia demand is for neutralization of the acidification as discussed in section 9.2, the value $\alpha_a = 1/2$ will be used in the following investigations. This value provides reasonable descriptions of both acetate and biomass conversions.

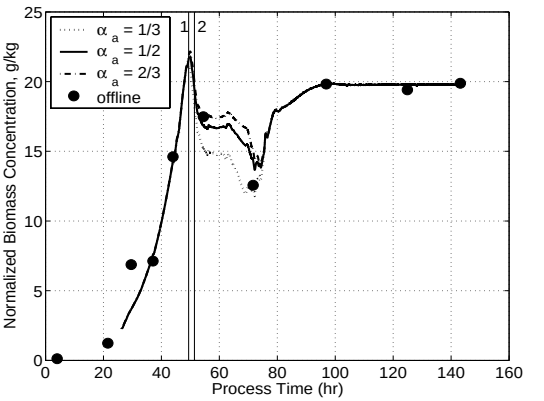
9.5.1.2 MTS01:

In this cultivation a faulty pump in the cooling circuit of the bioreactor resulted in a slow increase in the temperature of the broth of 3°C from 37 to 45 hours, at which point the cooling circuit was working properly again and the broth temperature was back at its set point at 46 hours. It is not known to what extent this deviation

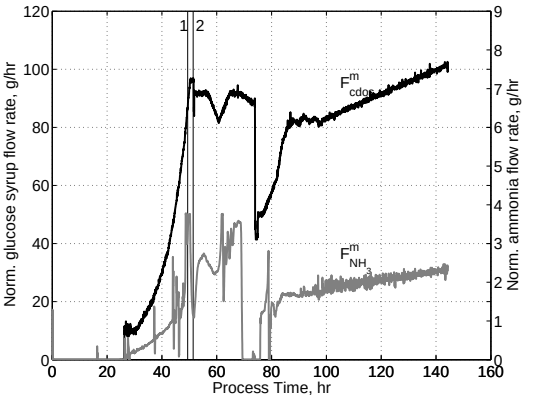
[7] Approximately the same concentration (C-mole) as acetate in the offline sample



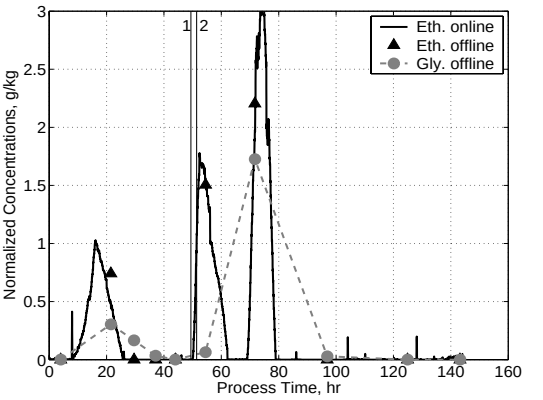
(a) Normalized acetate concentration



(b) Normalized biomass concentration



(c) Normalized flow rates of glucose syrup (black line) and ammonia (grey line)



(d) Normalized glycerol (•, dashed line: linear interpolation) and ethanol (▲, full line) concentrations

Figure 9.11. Estimated trajectories of normalized concentrations of acetate (a) and biomass (b) as a function of different values of α_a in equation 9.42. Trajectories of normalized measured values of glucose syrup feeding rate and ammonia flow rate (c) and glycerol and ethanol concentrations (d). The ethanol concentration in the broth is estimated from measurement of the ethanol concentration in the off-gas. Offline measurements of acetate (a), biomass (b), glycerol and ethanol (d) concentrations have been included for comparison. The offline glycerol measurements are connected by a hatched line. The vertical lines labelled 1 and 2 indicate the onset of nutrient substrate feeding and transition from fed-batch to continuous operation, respectively. The dissolved oxygen tension (DOT) is above 20% during the whole experiment. Data from cultivation MTS01. (See also figures 8.13 and 8.17 for details on performance regarding estimation of ammonia flow rate and indication of the onset of oxidoreductive growth, respectively.)

in temperature affects the metabolism. In figure 9.11 the trajectories obtained using $\alpha_a = 1/2$ show how a large amount of acetate is being formed just prior to the switch from fed-batch to continuous operation at 51 hours. The biomass concentration is 20% higher than the value obtained in the stationary phase of the continuous operation ($c^m_{x,stat} \approx 20$ g/kg)[8]. As the biomass concentration in the fed-batch phase increases to 15% above $c^m_{x,stat}$, significant production of acetate occurs from 48.5 hours, which initially is not accompanied by production of ethanol (figure 9.11(d)). From 50 to 52.5 hours ethanol is being formed, causing a drop in biomass concentration below $c^m_{x,stat}$ in combination with the initiation of continuous operation and a continuous decrease in biomass for the next 20 hours.

After the initial production, ethanol is slowly being diluted from the bioreactor (figure 9.11(d)), however figure 9.11(a) shows that acetate is estimated to continuously being produced for almost 20 hours (48.5 to 68 hours). At 69 hours a large production of ethanol appears as the acetate concentration drops rapidly. The estimated biomass concentration is not as low as the measured offline value, due to the aforementioned undescribed role of glycerol, however it is clearly seen that the biomass concentration is much lower than the expected value of $c_{x,stat}$ at this point[9]. By instantly lowering the substrate feeding rate to 50% of the value prior to ethanol production (figure 9.11(c)) and slowly increasing the dilution rate back to the normal rate over 30 hours, the process gets back on track as indicated by acetate, ethanol and biomass measurements.

For the last 50 hours of the cultivation (100-150 hours) the two substrate feeding rates[10] are slowly and simultaneously being increased as seen in figure 9.11(c). This is done to investigate the response of the process to increased dilution rates and to get an indication of what the critical growth rate[11] is during the growth conditions of the continuous operation. The system is seen to respond to this slow change without problems.

9.5.1.3 MTS05:

Whereas figure 9.11 demonstrated the performance of the model in a situation with extensive amounts of acetate formed, figure 9.12 illustrates the performance of the BAC soft sensor in a scenario, where no ethanol is measured in the offgas during the fed-batch and continuous operation of the cultivation. No significant amounts of acetate are reported to appear when evaluating the offline measurements. The estimated trajectories of acetate and biomass are in good agreement with the offline measurements. The changes in the acetate concentration seen between 45 to 55 hours are primarily due to oscillations due to synchronized growth, which can be observed in the ammonia flow rate in figure 9.12(c). It is interesting to note that the feeding strategy does not lead to an overshoot in the biomass concentration,

[8] Normalized value

[9] During operation it was expected that biomass concentration could be determined by using a constant yield on the glucose fed to the system, since no ethanol was present in the offgas and the appearance of acetate or glycerol was not anticipated.

[10] In continuous operation glucose syrup and nutrient substrate are fed in a ratio of 1:2 as discussed in section 9.1.8.1.

[11] Here the critical growth rate refers to the growth rate above which ethanol is continuously formed.

compared to the overshoot of 20% in MTS01 (figure 9.11(b)). The feeding strategy in cultivation MTS05 seems to have been better balanced than in cultivation MTS01.

9.5.1.4 MTV03:

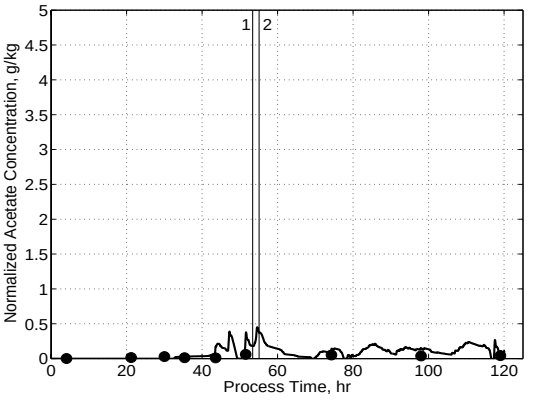
In cultivation MTV03 two process disturbances occurred, one unplanned and one planned, and the estimated trajectories of acetate and biomass concentrations are shown in figure 9.13. At the change from fed-batch to continuous operation at approximately 55 hours it can be seen from figure 9.13(a) that large amounts of acetate has accumulated, and the biomass concentration is 10% higher than during steady state conditions. Some ethanol is observed in figure 9.13(d), but at its maximum the ethanol concentration is less than 50% (C-mole/C-mole) of the acetate concentration.

From figure 9.13(c) it can be seen that the feeding rate following the change from fed-batch to continuous operation is far from ideal. The changes in feeding rate just before 60 hours are a response to the ethanol observed in figure 9.13(d). Using a ramp to slowly increase the substrate feed rate back to the normal value and thereby getting the process back on track; this is obtained at 68 hours. As the process seems to settle into stable mode at 80 hrs, a step-up in the substrate feed rate is conducted, elevating for 3 hours the dilution rate to 150% compared to the value prior to the step, followed by lowering of the dilution rate for 1.5 hours to a level of 60% of the value prior to the initial step, after which the dilution rate again is set to the original dilution rate.

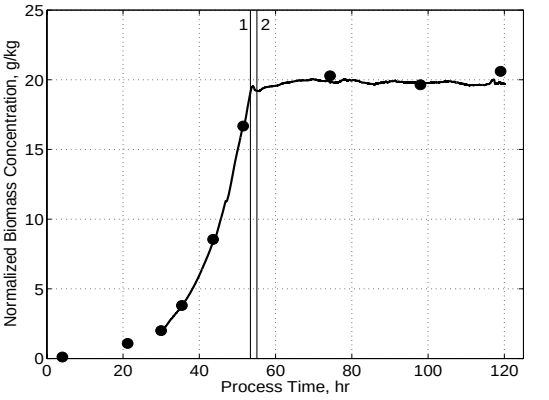
It is interesting to observe the estimated trajectory obtained using the BAC soft sensor after the peak in acetate and ethanol concentrations at 60 hours. Whereas the ethanol is quickly consumed, acetate is only slowly being removed from the culture broth, both by dilution and consumption. By close inspection of figure 9.13(a) (not shown here) the estimated acetate concentration can be seen to fluctuate from 70 to 80 hours. An explanation for this can be found by inspection of the CO_2 emission rate (CER) and ammonia flow rate shown in figure 9.14(a). These signals indicate that synchronized growth is observed from 70 hours and up until the step in dilution rate.

The step change introduced at 80 hours is much more severe than *e.g.* the ramp used at the end of cultivation MTS01 in figure 9.11(c) or the situations observed during the fed-batch phases of cultivations MTS01, MTS05 and MTV03. This is seen in the response of ethanol to the sudden increase in dilution rate, which occurs more or less instantaneously as the step is introduced. The estimated trajectory of the acetate concentration shows a significant peak as a result of the step change in the dilution rate. However the offline measurement of acetate at 83 hours is only 1/8 of the estimated value at the same time. As the system settles after the dilution rate is brought back to its nominal value, synchronized growth seems to reappear at 90 hours, leading to a slight overestimation of the acetate concentration for the remaining part of the experiment (see figure 9.14(b)).

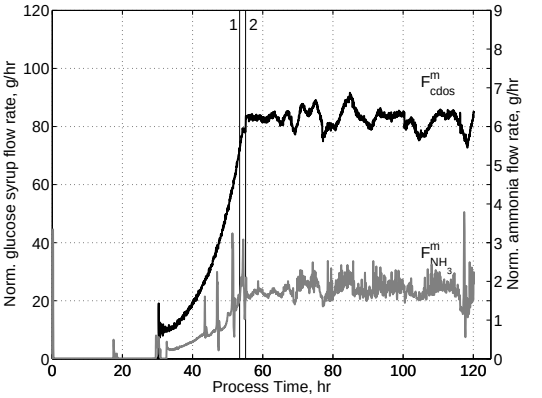
In figure 9.13(d) it appears that the concentration of glycerol is more or less constant and repeatedly observed in significant amounts (> 0.1 of the normalizing concentration). From inspection of data from other cultivations it appears that the offline samples analyzed by HPLC in December 2003 all show a low but constant



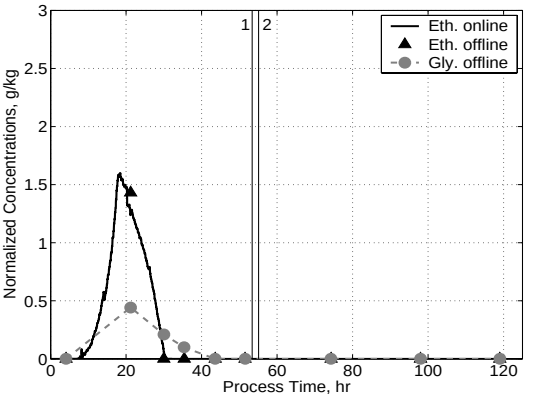
(a) Normalized acetate concentration



(b) Normalized biomass concentration

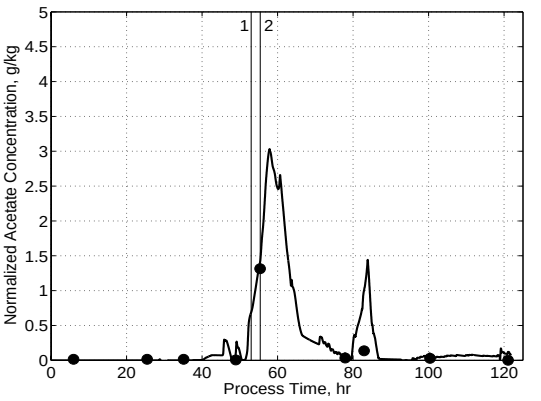


(c) Normalized flow rates of glucose syrup (black line) and ammonia (grey line)

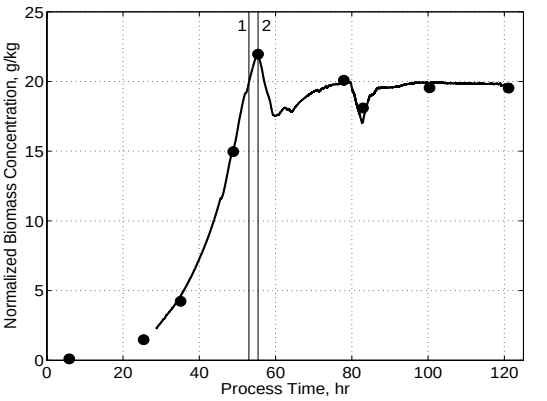


(d) Normalized glycerol (•, dashed line: linear interpolation) and ethanol (▲, full line) concentrations

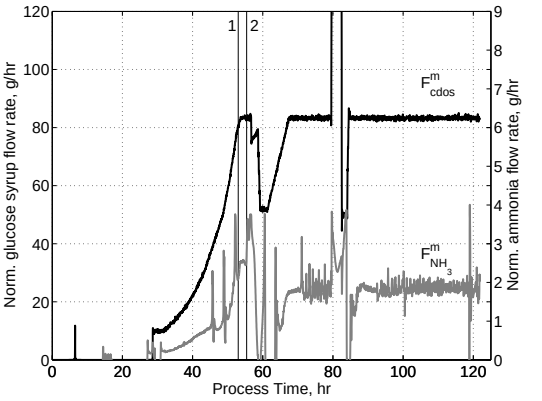
Figure 9.12. Estimated trajectories of normalized concentrations of acetate (a) and biomass (b) using $\alpha_a = 1/2$. Trajectories of normalized measured values of glucose syrup feeding rate and ammonia flow rate (c) and glycerol and ethanol concentrations (d). The ethanol concentration in the broth is estimated from measurement of the ethanol concentration in the off-gas. Offline measurements of acetate (a), biomass (b), glycerol and ethanol (d) concentrations have been included for comparison. The variations seen in the glucose syrup feeding rate are planned perturbations (see section 9.1.8.2). The vertical lines labelled 1 and 2 indicate the onset of nutrient substrate feeding and transition from fed-batch to continuous operation, respectively. The dissolved oxygen tension (DOT) is above 20% during the whole experiment. Data from cultivation MTS05. (See also figures 8.12 and 8.17 for details on performance regarding estimation of ammonia flow rate and indication of the onset of oxidoreductive growth, respectively.)



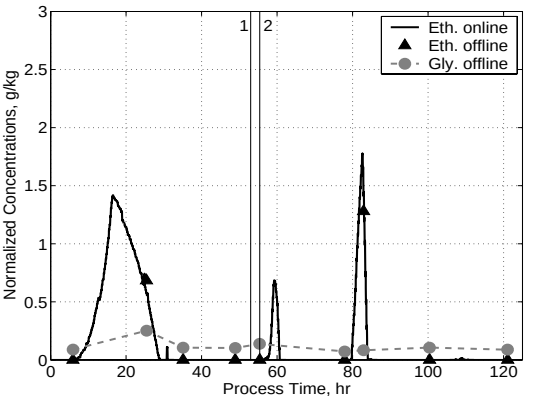
(a) Normalized acetate concentration



(b) Normalized biomass concentration

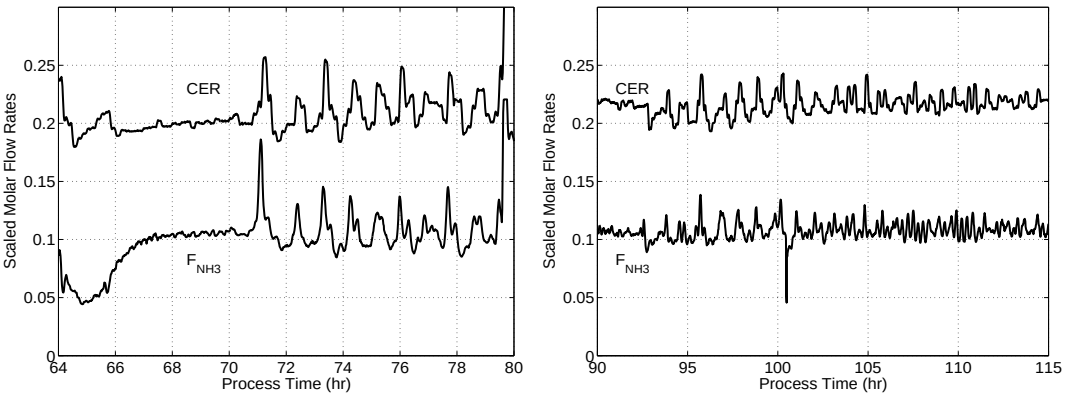


(c) Normalized flow rates of glucose syrup (black line) and ammonia (grey line)



(d) Normalized glycerol (•, dashed line: linear interpolation) and ethanol (▲, full line) concentrations

Figure 9.13. Estimated trajectories of normalized concentrations of acetate (a) and biomass (b). Trajectories of normalized measured values of glucose syrup feeding rate and ammonia flow rate (c) and glycerol and ethanol concentrations (d). The ethanol concentration in the broth is estimated from measurement of the ethanol concentration in the off-gas. Offline measurements of acetate (a), biomass (b), glycerol and ethanol (d) concentrations have been included for comparison. The vertical lines labelled 1 and 2 indicate the onset of nutrient substrate feeding and transition from fed- batch to continuous operation, respectively. The dissolved oxygen tension (DOT) is above 20% during the whole experiment. Data from cultivation MTV03.



(a) CER and $F^m_{NH_3}$, 64-80 hours

(b) CER and $F^m_{NH_3}$, 90-115 hours

Figure 9.14. Trajectories of normalized carbon emission rate (CER) and ammonia flow rate ($F^m_{NH_3}$) in the time periods 64-80 hours and 90-115 hours showing synchronized growth. Data from cultivation MTV03.

presence of glycerol. It remains to be tested if these observations are simply measurement errors or actually reflect presence of glycerol, since none of the cultivations analyzed by HPLC in September 2003 showed these tendencies.

9.5.1.5 Summary and Remarks

From the three examples shown it can be seen that the *BAC* soft sensor is able to explain the observed fluctuations in the ammonia flow rate as a response to production of acetate. In addition to the estimation of the acetate conversion rate, the model also provides an estimation of the biomass production rate. The two rates can then be used to estimate the concentrations of acetate and biomass in the bioreactor using equation 9.8.

The *BAC* soft sensor provides a good indication of the biomass concentration, while the estimation of the acetate concentration is more qualitative. It is however quite interesting to note that the recombinant strain of *S. cerevisiae* used in this work, show a remarkably large production of acetate during certain operating conditions. Acetate production is related to the fermentative metabolism, and while other strains of *S. cerevisiae* (Lei *et al.* (2003);Postma *et al.* (1989b)) are seen to produce large amounts of ethanol as a result of activation of the fermentative metabolism, this does not seem to be the case with this strain. The production of acetate sets in before ethanol is being detected in the off-gas.

9.5.2 Metabolic Flux Distributions

In the previous section it was demonstrated that the *BAC* soft sensor provides a good indication of the biomass conversion rate and concentration, while the estimation of the acetate conversion rate and concentration is more qualitative. With an online estimator of the biomass concentration available it is possible to perform a more quantitative analysis of the specific conversion rates of the different metabolites and subsequently analyse the distribution of intrinsic fluxes.

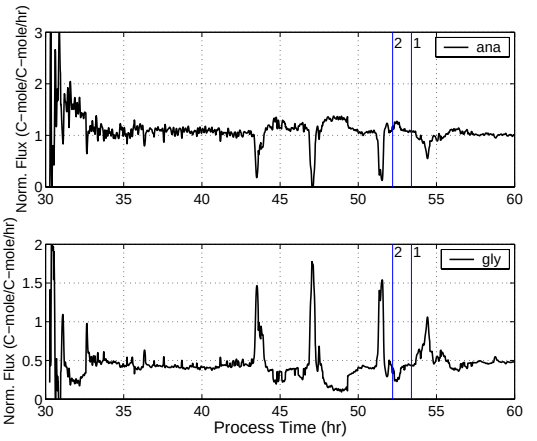
The distribution of intrinsic metabolic fluxes is analyzed using the *IMF* soft sensor presented in figure 9.8 and algorithm 9.2. The *IMF* soft sensor is used to elucidate how the flux is distributed between the major metabolic pathways. The distribution of fluxes onto the major metabolic pathways can provide an understanding of the mechanisms governing the metabolic activities of the cell. This discipline has extensively been used in steady state investigations of an wide range of wild type and engineered microorganisms (Vanrolleghem *et al.* (1996); Stückrath *et al.* (2002)), and a few studies have used metabolic flux models for dynamic investigations (*e.g.* Herwig and von Stockar (2002), Lei *et al.* (2003) and Lei *et al.* (2004)).

9.5.2.1 MTS05 Fed-batch operation:

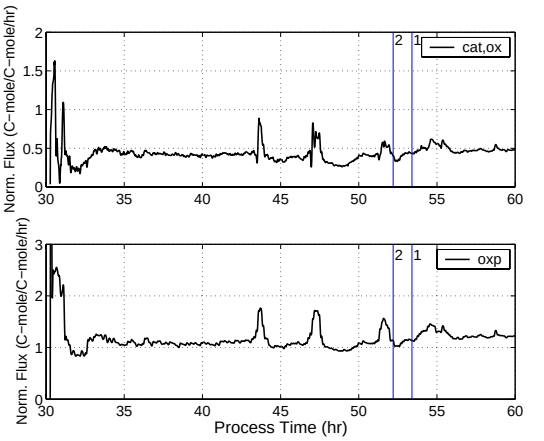
This cultivation represents one of the experiments conducted without significant amounts of acetate being produced. Figure 9.15 shows the estimated intrinsic fluxes during fed-batch operation estimated using the *IMF* soft sensor. Three fluctuations in the fluxes can be seen at 43, 47 and 51 hours as well as in the process variables shown in figures 9.15(e) and 9.15(f). From the trajectory of the glucose syrup feeding rate in figure 9.15(f), it can be seen that it is not the feed rate that is responsible for the fluctuations, since the signal is rather smooth. The behavior can be explained by synchronization of cellular activities, resulting in synchronized growth. A fourth less obvious fluctuation appear at 54 hours. This last fluctuation does not have the same shape as the previous three and may be triggered or disturbed by the onset of nutrient substrate feeding at 53.4 hours (line 1).

The intrinsic metabolic fluxes in figures 9.15 (a-d) can be seen to vary around a constant level during fed-batch operation, except in the vicinity for the fluctuations caused by synchronized growth. Figure 9.15(a) illustrates how glucose taken up by the cell as reaction *glc* in figure 9.8 is being divided between the anabolic, *ana*, and catabolic pathway (the glycolysis, *gly*), while figure 9.15(b) shows the oxidative catabolic flux, cat, ox and the flux through the oxidative phosphorylation, *oxp*. The normalized activity of the oxidative phosphorylation has a value of 1.0-1.2 mole/mole/hr. Fluxes of acetate and ethanol are shown in figure 9.15(c). No indications of ethanol production is observed, while production of acetate can be seen to occur during the oscillations, while consumption occurs between these.

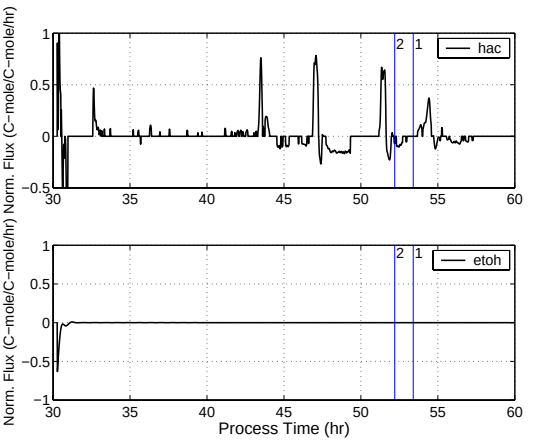
As discussed in section 9.3, the amount of CO_2 produced during anabolic activities (γ) has not been determined in this work. By assuming $Y_{xNADH} = 0.23$ as reported by Duboc *et al.* (1998) a value of $\gamma = 0.17$ is obtained. These assumptions have an effect on the estimated glycolytic and oxidative catabolic flux as well as the conversion rate of ATP, whereas all the other fluxes are not influenced by the values of γ and Y_{xNADH}. Figure 9.15(d) show the estimated specific conversion rate of NADH based on the aforementioned assumptions. The figure reflects the qualitative



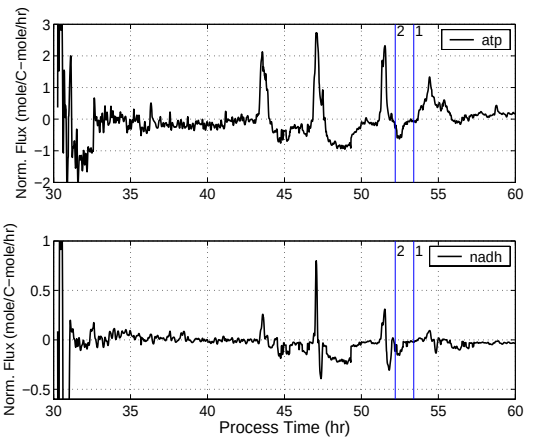
(a) Anabolic (top) and glycolytic flux (bottom)



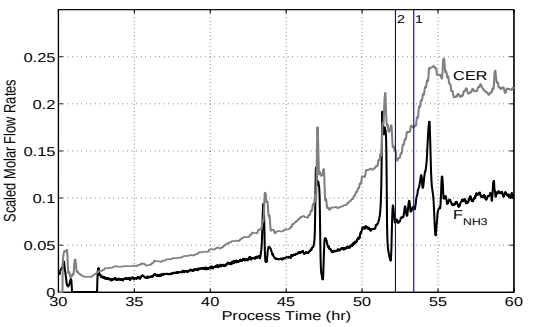
(b) Oxidative catabolic flux (top) and flux through oxidative phosphorylation



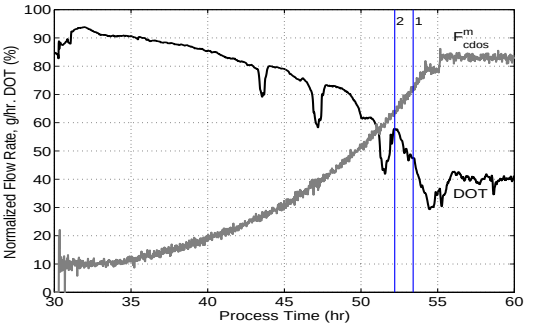
(c) Specific conversion rate of acetic acid (top) and ethanol (bottom)



(d) Specific conversion rate of ATP and NADH



(e) CER (grey line) and ammonia flow rate (F_{NH_3}, black line)



(f) DOT (black line) and glucose syrup flow rate (F^m_{cdos}, grey line)

Figure 9.15. Estimation of intrinsic flux distributions onto the major metabolic pathways (a-d). Data from fed-batch operation of cultivation MTS05. For comparison with process variables: (e) CER and ammonia flow rates and (f) DOT and glucose syrup flow rates. Vertical line 1 indicates the time at which feeding of nutrient substrate is initiated, while line 2 indicates the time at which continuous operation is initiated. (Data from continuous operation of MTS05 is reported in figure 9.16, while results using the *BAC* soft sensor are shown in figure 9.12.)

behavior of the specific conversion rate of NADH and it can be seen that production and consumption is balanced most of the time, except during synchronized growth.

The specific conversion rate of ATP is also shown in figure 9.15(d). For this estimation an additional assumption has been used, namely that the P/O ratio = 1.2 and constant. An estimate of Y_{xATP} can be obtained by combining equation 9.43 with data from continuous operation of 7 different cultivations listed in table 7.4, page 91 in chapter 7. A value of $Y_{xATP} = 1.7$ has been determined, which corresponds well with the observations by Lei *et al.* (2003) reporting a value of $Y_{xATP} = 1.8$-2 at steady state conditions during oxidative growth. Here it should be noted that the determination of Y_{xATP} using equation 9.43 is sensitive to the assumed values of Y_{xNaDH}, the P/O ratio and γ (equations 9.46 and 9.47). Since none of these parameters have been determined for the strain investigated in this study, it is unlikely that the reported value of Y_{xATP} is correct; however it serves the purpose of closing the ATP-balance. To obtain more accurate values of the 4 parameters mentioned above, further and more specific studies should be conducted to clarify this.

From figure 9.15(d) it can be seen that the specific conversion rate of ATP is more severely affected by the synchronized growth than is the case of NADH, and in general larger variations are seen in the conversion rates of ATP than in NADH.

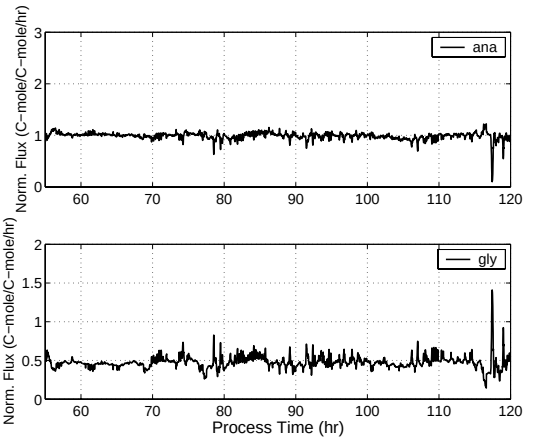
9.5.2.2 MTS05 Continuous operation:

During continuous operation perturbations are introduced in the substrate feed rates as shown in figure 9.17(f), meaning that no true steady state conditions are attempted to be achieved. The resulting distribution of intrinsic metabolic fluxes is shown in figure 9.16. The figure illustrates that the perturbations introduced in the substrate feed rate (see figure 9.12(c)), do not result in any significant changes to the flux distribution, except for for periods of acetate production. One exception to this is seen at 117.3 hours where acetate suddenly appears to be formed in large amounts, having an effect on both the anabolic and glycolytic flux. Otherwise figure 9.16(d) illustrates how the ATP and NADH balances seem to be closing throughout the continuous operation, supporting the assumption that $Y_{xATP} = 1.7$, and that the other physiological parameters Y_{xNADH}, P/O ratio and γ in table 9.5 can be assumed constant.

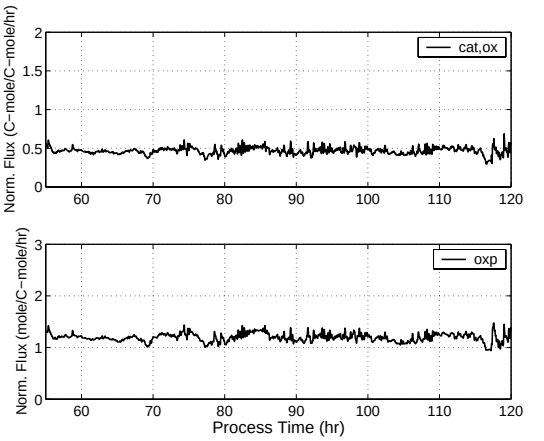
9.5.2.3 MTS01 Fed-batch operation:

Cultivation MTS01 was characterized by a number of situations where ethanol was being produced, and offline measurements indicated that both acetate and glycerol were present in the culture broth in large amounts. The distribution of fluxes in the fed-batch phase is shown in figure 9.17. Up until 45 hours the process is comparable to the results shown for MTS05 in figure 9.15, also showing fluctuations at 37 and 44 hours indicating synchronization of growth activities.

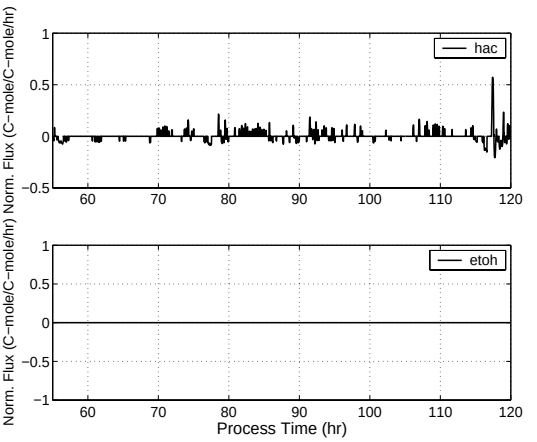
At 48.5 hours (line 1) an upset in the fluxes is again observed. This time it does not appear to be synchronized growth triggering an oscillation, since the steepness of the rise in most of the fluxes is almost 2.5 times lower than observed during synchronized growth, combined with no observation of significant variations in the oxidative phosphorylation (figure 9.17(b)). Acetate is being formed from 48.5 to



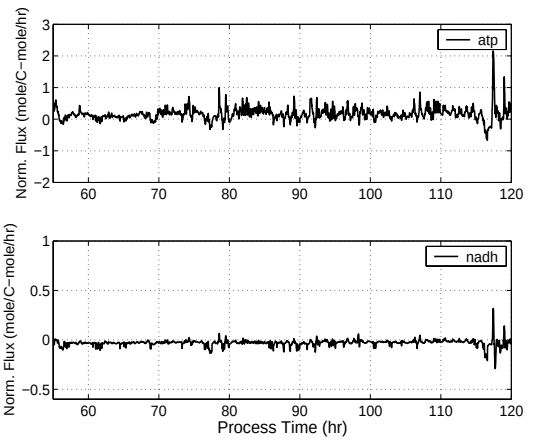
(a) Anabolic (top) and glycolytic flux (bottom)



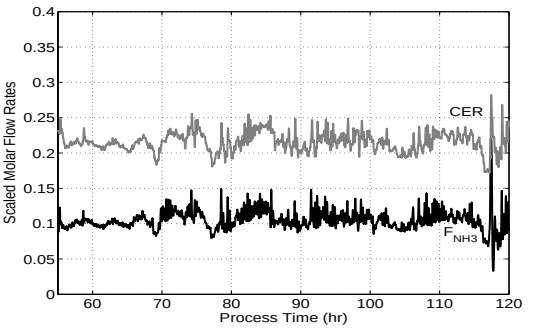
(b) Oxidative catabolic flux (top) and flux through oxidative phosphorylation



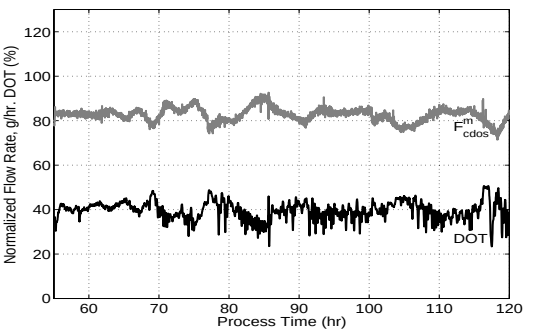
(c) Specific conversion rate of acetic acid (top) and ethanol (bottom)



(d) Specific conversion rate of ATP and NADH

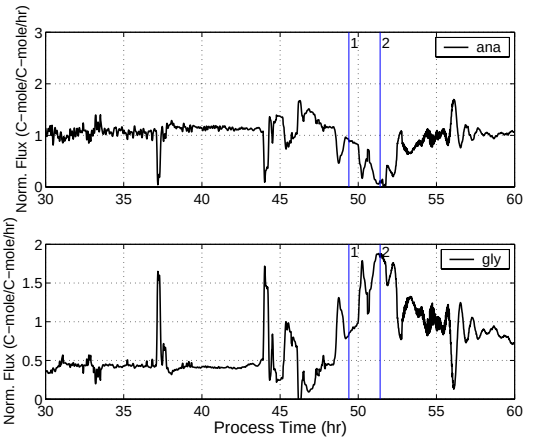


(e) CER (grey line) and ammonia flow rate (F_{NH_3}, black line)

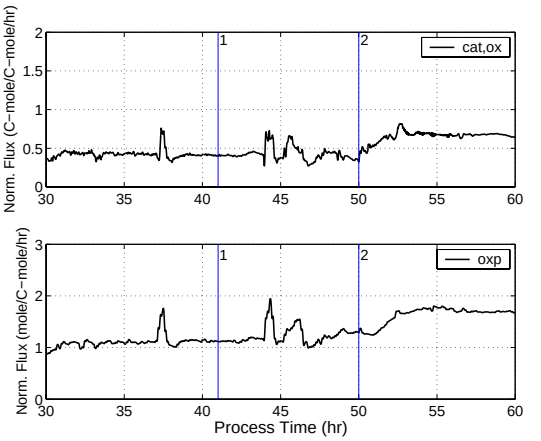


(f) DOT (black line) and glucose syrup flow rate (F^m_{cdos}, grey line)

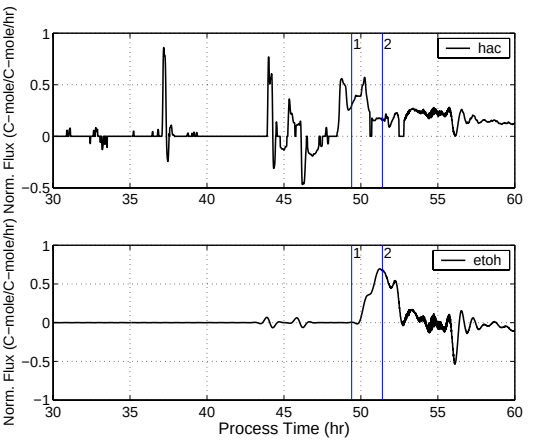
Figure 9.16. Estimation of intrinsic flux distributions between the major metabolic pathways (a-d). Data from continuous operation of cultivation MTS05. For comparison with process variables: (e) CER and ammonia flow rates and (f) DOT and glucose syrup flow rates. (Data from fed-batch operation of MTS05 is reported in figure 9.15, while results using the BAC soft sensor are shown in figure 9.12.)



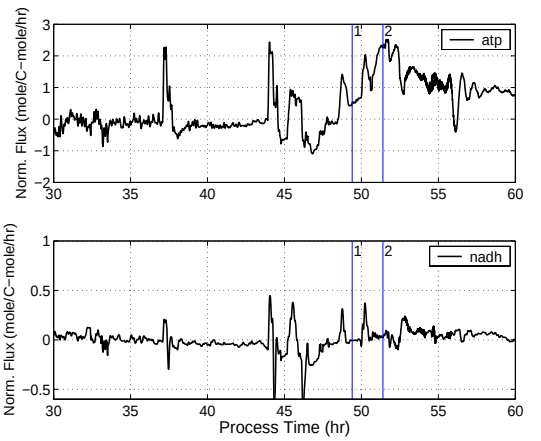
(a) Anabolic (top) and glycolytic flux (bottom)



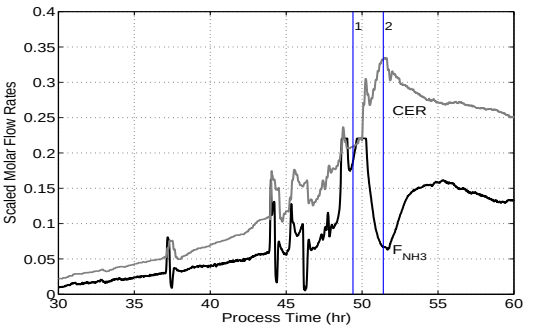
(b) Oxidative catabolic flux (top) and flux through oxidative phosphorylation



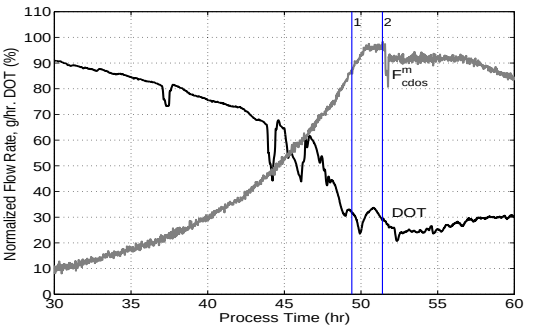
(c) Specific conversion rate of acetic acid (top) and ethanol (bottom)



(d) Specific conversion rate of ATP and NADH



(e) CER (grey line) and ammonia flow rate (F_{NH_3}, black line)



(f) DOT (black line) and glucose syrup flow rate (F^m_{cdos}, grey line)

Figure 9.17. Estimation of intrinsic flux distributions onto the major metabolic pathways (a-d). Data from fed-batch operation of cultivation MTS01. For comparison with process variables: (e) CER and ammonia flow rates and (f) DOT and glucose syrup flow rates. Vertical line 1 indicates the time at which feeding of nutrient substrate is initiated, while line 2 indicates the time at which continuous operation is initiated. (Data from continuous operation of MTS01 is reported in figure 9.18, while results using the *BAC* soft sensor are shown in figure 9.11.)

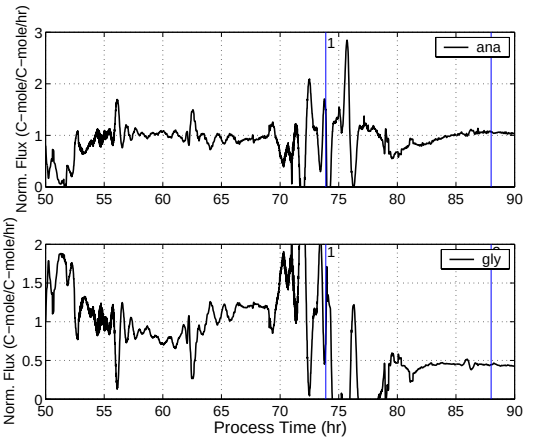
50.5 hours and again continuously from 52.8 to 60 hours, while a small amount of ethanol is formed between 50 and 52.5 hours. The trajectory of the ammonia flow rate in figure 9.17(e) illustrates how the maximum flow rate of ammonia is reached for 48.7 to 49.0 hours and again from 49.7 to 50.2 hours. Calculated variations in the fluxes are a combination of 1) metabolic changes and 2) changes of the process conditions since at 49.4 hours feeding of the nutrient substrate is initiated and at 51.3 hours continuous operation of the bioreactor is commenced (line 2). It is also interesting to note the change in the flux through the oxidative catabolism occurring at the onset of ethanol formation (50 hours). A sharp increase in the activity of the oxidative catabolism is seen in figure 9.17(b), reaching a level around 50% above the activity prior to the upset. Also the activity of the oxidative phosphorylation is affected by the process upset and slowly increases to a level around 50% above (55 hours) the activity prior to the upset. The NADH balance shown in figure 9.17(d) indicates that an imbalance exists between the reactions producing and consuming NADH. As the activity of the oxidative phosphorylation is increased the NADH balance is reestablished around zero at 55 hours. At the upset it can be observed that a net production of ATP is estimated by the model. It is not obvious why this large production rate of ATP appears, but an explanation could be found in the assumption of constant values of Y_{xNADH} or P/O ratio.

The above analysis neither provides any clear indication of why the sudden production of acetate occurs at 48.5 hours, nor why ethanol is being produced at 50 hours. However it is interesting to note at the onset of ethanol production, a sudden jump in the oxidative catabolism flux appears, while this is not the case during production of acetate at 48.5 hours, despite the fact that these two metabolic products are closely positioned in the fermentative pathway from pyruvate both being products of conversion of acetaldehyde. It is not apparent whether the production of ethanol is a consequence of or the reason for the elevated activity of the oxidative catabolism in figure 9.17(b).

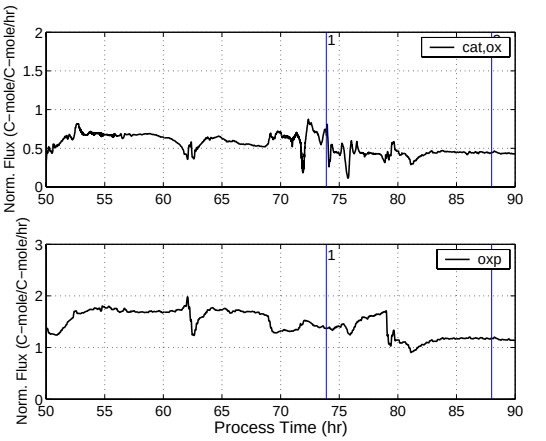
9.5.2.4 MTS01 Continuous operation:

In the continuous phase of the MTS01 cultivations interesting behavior was observed as shown and discussed in section 9.5.1 and shown in figure 9.11. Figure 9.18 illustrates the estimated flux distribution from 50 to 90 hours; it should be borne in mind that glycerol was observed in offline sampling at 72 hours as shown in figure 9.11(d), but this effect has not been modeled.

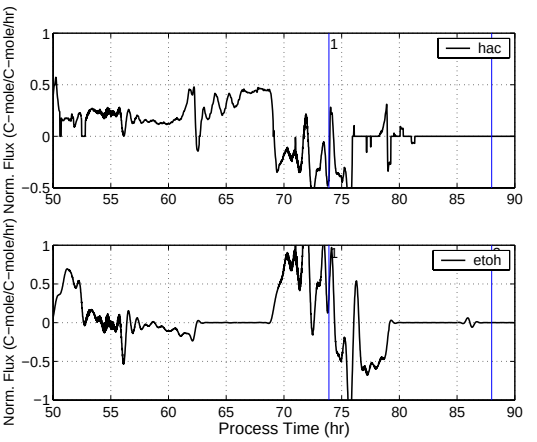
Just after 60 hours a change in the flux distribution is observed, where the acetate flux in figure 9.18(c) is increased and so is the activity of the oxidative phosphorylation (figure 9.18(b)) as well as the glycolytic flux (figure 9.18(a)). At the peak the maximum ammonia flow rate is reached as seen in figure 9.18(e), lasting for 0.5 hours after which a large drop in the acetate flux rate is observed at 62 hours, combined with a drop in the activity of oxidative phosphorylation and the glycolytic flux. It is also interesting to note the fluctuation in the DOT signal at 62 hours shown in figure 9.18(f). Oscillations around a constantly increasing acetate flux is observed between 63 and 66 hours, having effect on the estimation of some of the other fluxes (anabolic, glycolytic and ATP fluxes), while neither the flux through the oxidative phosphorylation nor the oxidative catabolism, both independent of the estimated



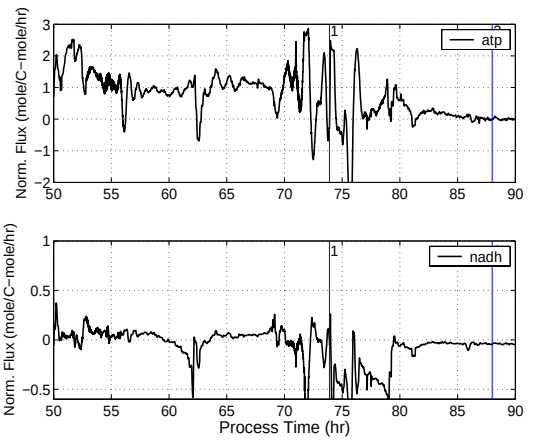
(a) Anabolic (top) and glycolytic flux (bottom)



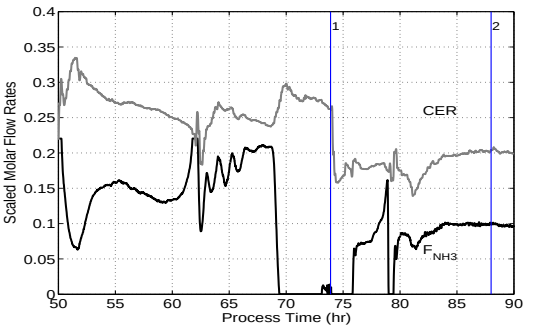
(b) Oxidative catabolic flux (top) and flux through oxidative phosphorylation



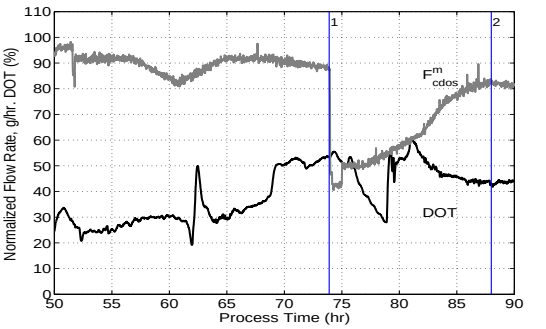
(c) Specific conversion rate of acetic acid (top) and ethanol (bottom)



(d) Specific conversion rate of ATP and NADH



(e) CER (grey line) and ammonia flow rate (F_{NH_3}, black line)



(f) DOT (black line) and glucose syrup flow rate (F^m_{cdos}, grey line)

Figure 9.18. Estimation of intrinsic flux distributions onto the major metabolic pathways (a-d). Data from continuous operation of cultivation MTS01. For comparison with process variables: (e) CER and ammonia flow rates and (f) DOT and glucose syrup flow rates. Vertical line 1 indicates the time at which feeding is stopped due to large amounts of ethanol being present in the off-gas signal (see figure 9.11), while line 2 indicates the time at which the process is back on track in continuous operation. (Data from fed-batch operation of MTS01 is reported in figure 9.17, while results using the *BAC* soft sensor are shown in figure 9.11.)

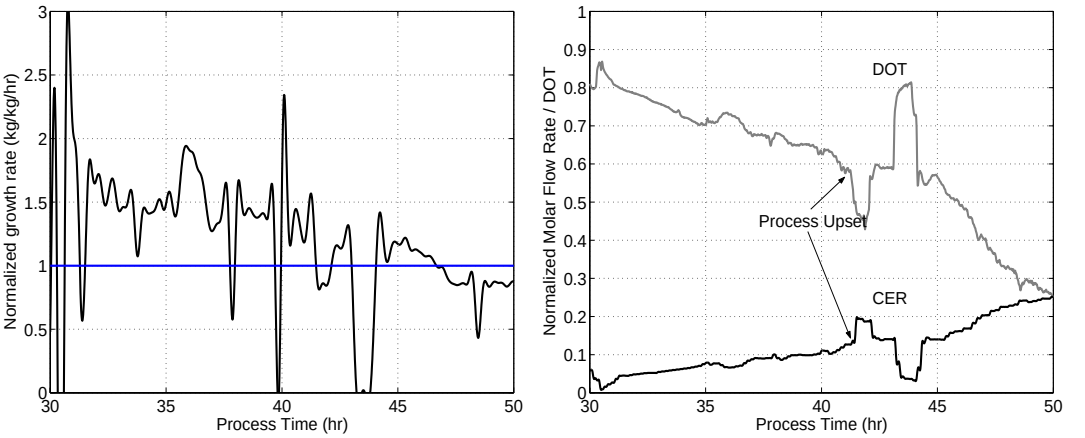
acetate flux, can be seen to be affected by these variations. It is not known what underlying phenomenon is responsible for the observed behavior, however it is speculated that the fluctuations are a response to the change in glucose feed flow rate, which according to figure 9.18(f) occurs at 61 hours.

From 63 to 68.8 hours a constant increase in the acetate flux (figure 9.18(c)) can be observed, followed by a very sharp and steep decrease into a state with acetate consumption. This transition is not triggered by exceeding the maximum feed rate of ammonia or due to changes in other process parameters (pH, aeration, stirring *etc*). At the same time ethanol starts being produced, while the activity of the oxidative phosphorylation drops significantly and the flux through the oxidative catabolism increases (figure 9.18(b)). The NADH balance (figure 9.18(d)) show that the drop in the oxidative phosphorylation results in a too small capacity in the reoxidation of NADH leading to a surplus in NADH. This increase in NADH could be the reason for the activation of glycerol production (see figure 9.11(d) on page 160) in order to restore the redox potential of the cells. The anabolic flux can be seen to slowly increase from 63 to 74 hours, which is misleading since the observations in figure 9.11 indicated that the biomass production rate was overestimated, even before the ethanol production set in. In general a lower and decreasing biomass concentration would result in larger fluxes, so rather than the observed slow decrease in the activity of the oxidative phosphorylation up until the drastic drop at 68.8 hours, the erroneous biomass estimation could hide that a maximum in the capacity of the oxidative phosphorylation is reached. From figure 9.11 it can be observed that the biomass concentration estimated using the BAC soft sensor is larger than the off-line measurement at 71 hours. The estimated value is 20% above the measured biomass concentration. Multiplying the estimated activity of the oxidative phosphorylation at 68.8 hours by 1.2, yields a normalized activity of the oxidative phosphorylation close to 2.0 mole/C-mole/hr, which is comparable to the highest values of the activity of the oxidative phosphorylation for any of the conducted experiments. Assuming that a maximum activity of the oxidative phosphorylation is being achieved, the consequence of such a maximum would be an increase in the NADH concentration and the cell reacting to this by activating mechanisms to reoxidize NADH, namely the ethanol and glycerol pathways. The drop in the activity of the oxidative phosphorylation at 68 hours indicates that repression of this system occurs under conditions of ethanol and/or glycerol production. This observation is supported by the small drop in the activity of the oxidative phosphorylation at 50 to 52.5 hours, where ethanol and possibly glycerol are being produced.

From 68.8 up until 74 hours ethanol is continuously being produced, at which point the feed rates are lowered with 50% in an attempt to get the process back to a state of balanced, purely oxidative growth. Up until the lowering of the feed rates the acetate flux is seen to be negative, indicating consumption of acetate. During this time period no ammonia is fed to the system, which is the reason for the more or less constant consumption rate of acetate. The pH increases with 0.2 pH- unit and is back at its set point at 71 hours (not shown here).

After the feed rate is lowered (line 1 figure 9.18), a slow increase in the feeding rate is begun to restore the standard conditions at 88 hours. Ethanol is being consumed from 74 to 79 hours, and the NADH balance shown a reoxidation of NADH coinciding with an increasing activity of the oxidative phosphorylation. The

apparent negative flux of NADH in this period might be due to consumption of the glycerol formed earlier. As the last amounts of ethanol is being consumed, acetate is seen to be formed. A similar scenario was also reported by Pons *et al.* (1986) during batch operation. The acetate is quickly being consumed shortly after the ethanol is exhausted. From 82 hours and onwards the system can be seen to have been reestablished in a mode of balanced oxidative growth.



(a) Normalized estimated growth rate

(b) CER and DOT

Figure 9.19. Normalized estimated growth rate, carbon emission rate (CER) and dissolved oxygen tension (DOT) in the fed-batch phase of cultivation MTE02. (a) The growth rate under normal operating conditions is indicated by the horizontal line at 1.0. (b) The stop of ammonia addition is indicated with arrows. Flow rates of ammonia and glucose syrup can be seen in figures 9.20(e) and 9.20(f)

9.5.2.5 MTE02 fed-batch operation:

This cultivation was carried out in order to test if it was possible to apply a more aggressive feeding profile in the fed-batch phase, since earlier investigations had indicated this to be possible. At the onset of fed-batch operation a feeding profile was imposed corresponding to a required growth rate for oxidative growth of twice the growth rate during standard fed-batch operation. As the operation proceeded, the feeding profile was adjusted such that as the continuous phase was initiated, the growth rate would be slightly lower than during normal operating conditions, or in other words a more or less constant deccelation of the required growth rate as illustrated in figure 9.19(a). Between 41 and 44 hours a series of process upsets occurred. First the ammonia supply system broke down from 41 to 43.5 hours as seen in figure 9.20(e), and to avoid the growth being nitrogen limited, the feed rate of the glucose syrup was stopped from 43 to 44 hours (figure 9.20(f)) after which the system quickly got back on track.

Figure 9.20 shows the estimated flux distributions in the fed-batch phase. A very high anabolic flux is obtained in the beginning of the phase, slowly decreasing for the duration of the phase, except during the aforementioned process upsets (figure 9.20(a)). The activity of the oxidative phosphorylation can be seen to slowly increase

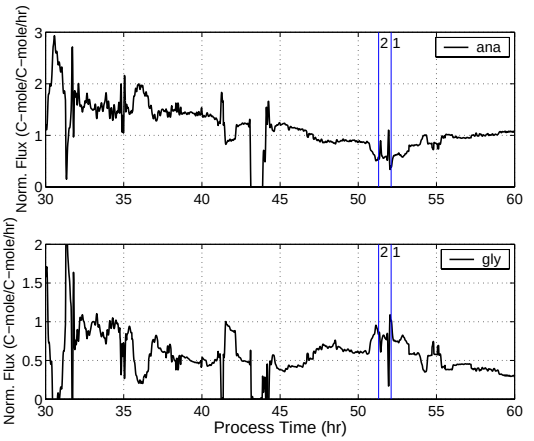
from 30 to 35 hours in a response to an increased need for reoxidation of NADH, illustrated in figures 9.20(b) and 9.20(d) respectively. In this time period acetate is constantly being produced, up until the time point at which the activity of the oxidative phosphorylation reaches its maximum at 34.5 hours. Again a maximum value of the activity of the oxidative phosphorylation close to 2.0 mole/C-mole/hr can be seen as was observed in cultivation MTS01. In this case however, the activity of the oxidative phosphorylation is not repressed.

As the ammonia supply system breaks down a peculiar response is seen in the estimation of the flux distributions. Whereas most of the other fluxes are being influenced by the erroneous estimation of the anabolic and acetate fluxes, the activity of the oxidative phosphorylation is independent of this. It is interesting to observe that the response to the process upset is a fast increase in the activity of the oxidative phosphorylation, an observation supported by measurements of the dissolved oxygen tension and the carbon emission rate in figure 9.19(b). The dissolved oxygen tension can be seen to decrease rapidly as the ammonia flow is stopped, and after an hour returning to a value, which is similar to the level before the process upset. The variation in the CER signal also indicates that for some reason increased metabolic activity appears as a consequence of the stop of the ammonia flow. It is not until the substrate feeding is stopped from 43 to 44 hours that indications of decreasing cellular activity is seen in the two signals.

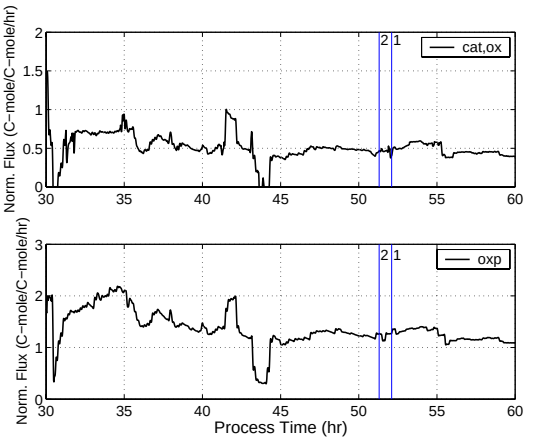
After the process is back on track at 45 hours the flux distribution is similar to the distribution prior to the process upset. From 47 hours and until 53 hours it appears that acetate is being produced, which however is not supported by offline measurements of the culture broth at 48.5 and 55 hours (not shown here). It is not known whether this discrepancy can be an artifact due to the process disturbances from 41 to 45 hours.

9.5.2.6 MTE02 Continuous operation:

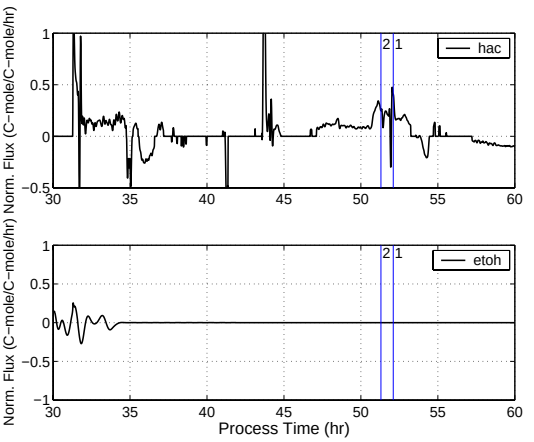
Following the variation in the fed-batch phase of cultivation MTE02, a constant feeding strategy is maintained for more than 20 hours up to 80 hours. From 65 to 80 hours oscillations can be seen in many of the signals shown in figure 9.21, indicating synchronization of cellular activities. The oscillations are however rather small and not very ordered, giving an impression that a number of subpopulations exist and oscillate with different periods. Figure 9.21 shows the estimated flux distribution from 60 to 123 hours where the process is terminated. As the ramp in the feeding rates is initiated the oscillations from synchronized growth can be seen to disappear, and until 113 hours the process proceeds without any incident. At 113.5 hours acetate starts being continuously produced (figure 9.21(c)), having an effect on the estimated anabolic and glycolytic fluxes (figure 9.21(a)) and also on the ATP and NADH balances (figure 9.21(d)). At 115 hours ethanol starts being formed, followed by a drop in the activity of the oxidative phosphorylation (peaking at 1.8 mole/C-mole/hr) indicating a repression and a drop in the acetate formation rate. At 116.8 hours the feeding rates are set back to values prior to the initiation of the ramp and a derepression of the oxidative phosphorylation (peaking at 1.9 mole/C-mole/hr) quickly leads to a conversion of the accumulated ethanol and apparently also accumulated NADH. As the process is terminated the system has not settled



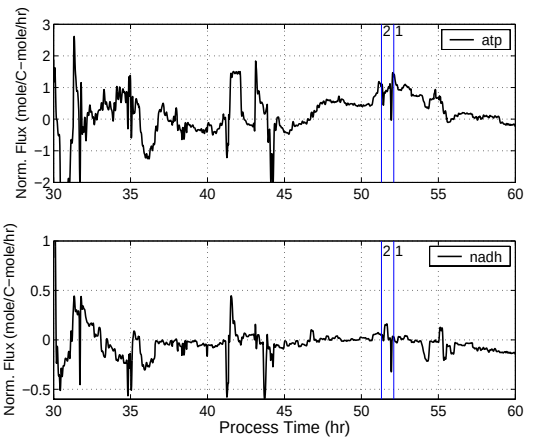
(a) Anabolic (top) and glycolytic flux (bottom)



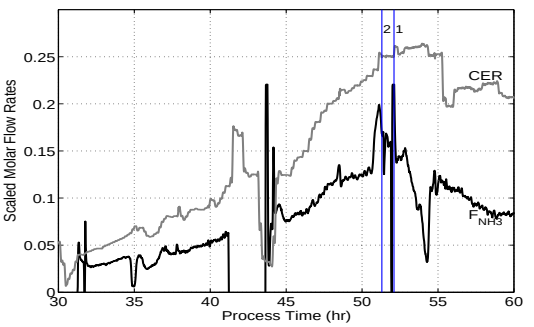
(b) Oxidative catabolic flux (top) and flux through oxidative phosphorylation



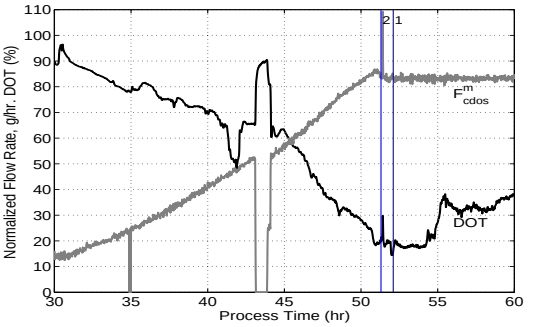
(c) Specific conversion rate of acetic acid (top) and ethanol (bottom)



(d) Specific conversion rate of ATP and NADH

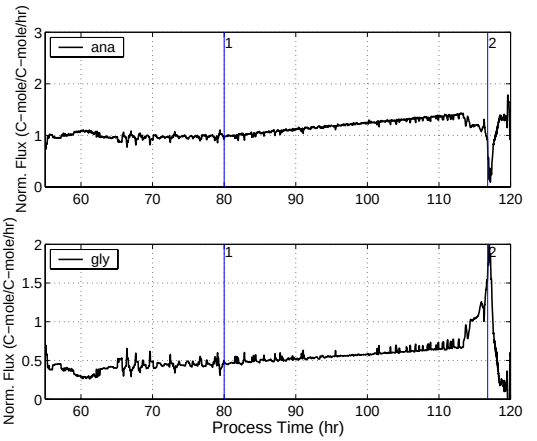


(e) CER (grey line) and ammonia flow rate (F_{NH_3}, black line)

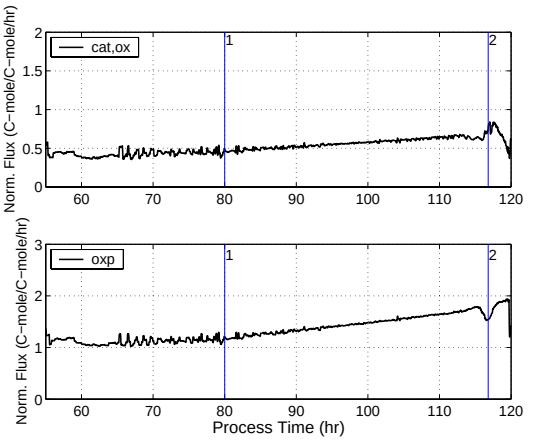


(f) DOT (black line) and glucose syrup flow rate (F^m_{cdos}, grey line)

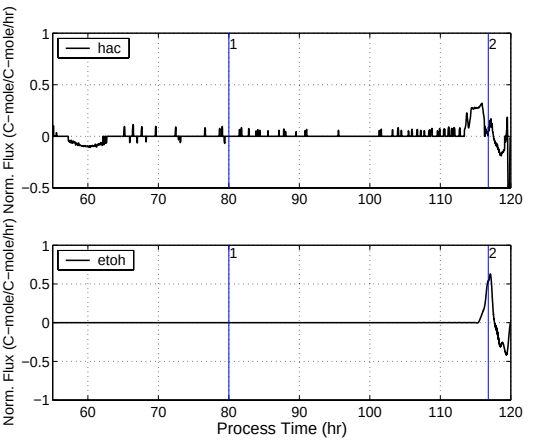
Figure 9.20. Estimation of intrinsic flux distributions onto the major metabolic pathways (a-d). Data from fed-batch operation of cultivation MTE02. For comparison with process variables: (e) CER and ammonia flow rates and (f) DOT and glucose syrup flow rates. Vertical line 1 indicates the time at which feeding of nutrient substrate is initiated, while line 2 indicates the time at which continuous operation is initiated. Note that the flux estimations between 41 and 45 hours are questionable, since a number of process upsets occurred in this time period. (Data from continuous operation of MTE02 is reported in figure 9.21)



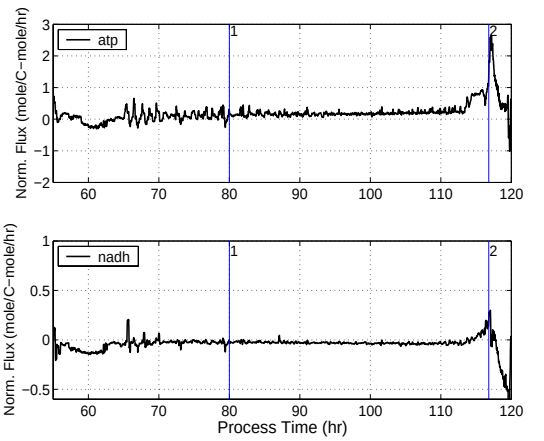
(a) Anabolic (top) and glycolytic flux (bottom)



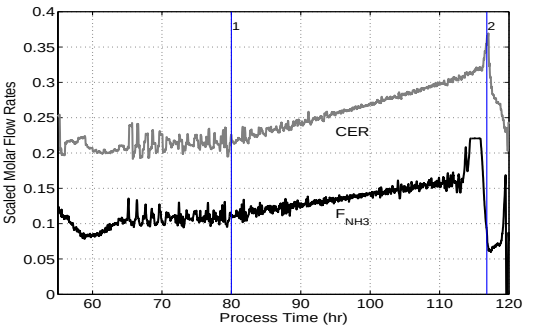
(b) Oxidative catabolic flux (top) and flux through oxidative phosphorylation



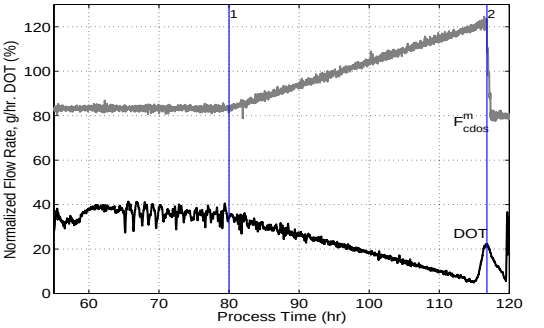
(c) Specific conversion rate of acetic acid (top) and ethanol (bottom)



(d) Specific conversion rate of ATP and NADH



(e) CER (grey line) and ammonia flow rate (F_{NH_3}, black line)



(f) DOT (black line) and glucose syrup flow rate (F^m_{cdos}, grey line)

Figure 9.21. Estimation of intrinsic flux distributions onto the major metabolic pathways. Data from continuos operation of cultivation MTE02. For comparison with process variables: (e) CER and ammonia flow rates and (f) DOT and glucose syrup flow rates. Vertical line 1 indicates the time at which a constant increase (a ramp) in the feeding of substrates is initiated, while line 2 indicates the time at which the feed rate is adjusted back to the original value before the initiation of the ramp. (Data from fed-batch operation of MTE02 is reported in figure 9.20)

yet.

It is interesting to note how the *IFM* soft sensor estimates that formation of acetate begins almost 2 hours before the formation of ethanol at 115 hours. After the shift-down it is also interesting to note that the activity of the oxidative phosphorylation increases even as ethanol reportedly still is being produced in the time period from 116.8 to 117.7 hours.

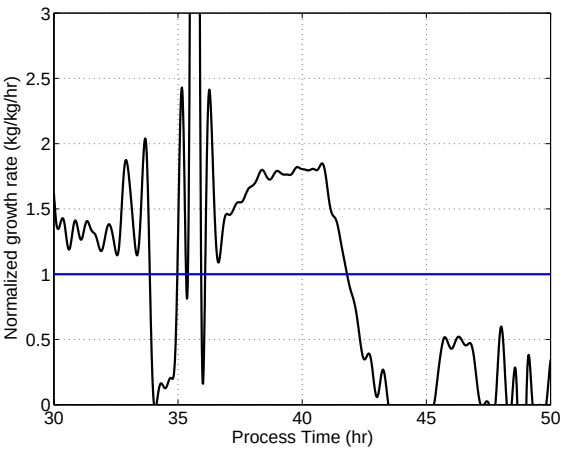


Figure 9.22. Normalized estimated growth rate in the fed-batch phase of cultivation MTF02. The growth rate under normal operating conditions is indicated by the horizontal line at 1.0. Between 34 and 36 hours initialization problems of the closed loop controller led to upsets in the substrate feed rate in this time period. See figure 9.23 for flow rates of glucose syrup and ammonia as well as CER and DOT signals in the fed-batch operation.

MTF02 Fed-batch operation: In cultivation MTF02 a control loop was closed between a measurement of reducible components in the offgas from the bioreactor and the substrate feed rates (the nutrient feeding was inactive during batch and fed-batch operation), see section 9.1.8.4 for more detail. At the time when the cultivation was conducted, the role of acetate as described above had not yet been elucidated. It was therefore expected that ethanol would be a good indicator for approaching the critical growth rate as reported by Lei *et al.* (2003). Using a small value of ethanol concentration as set point to the controller it was thought possible to maintain balanced oxidative growth throughout the fed-batch phase, as formation of ethanol would lead to a decreased feeding rate of the substrate. The set point has been changed a number of times during the cultivation, corresponding to ethanol concentrations 0.1 - 0.3 g/L.

Figure 9.22 illustrates the estimated normalized growth rate and the normalized feeding rate of glucose syrup during the fed-batch operation. Between 34 and 36 hours controller problems led to upsets in the substrate feed rate during this time period. Aggressive feeding, as dictated by the controller, at almost 2 kg/kg/hr from 36 to 40.8 hours is followed by a sharply decreasing growth rate from 41 to 45 hours, after which the feeding of glucose syrup is stopped. An attempt to restart the system at 47.5 hours is not successful, due to problems with the Figaro sensor[12].

[12]In general the data from the Figaro sensor showed indications of ethanol being absorbed

The variations in the feed rate are due to the combined actions of the controller in response to the process system. Initially the concentration of ethanol in the offgas is below the controller set point (0.1 g/L) and the substrate feeding rate is being increased to a maximum value at 41 hours. A cellular response is triggered and suddenly a very rapid increase in the ethanol concentration leads to a decrease in the substrate feed rate, however the control actions are not sufficient to get the cells to stop producing large amounts of ethanol, resulting in a feeding rate of zero from 46 to 50 hours, after which oxidative growth condition is reestablished and continuous operation begun at 50 hours.

The estimated flux distributions through the fed-batch and a part of the continuous phase is shown in figure 9.23. Focusing on the time period from 36 to 40.8 hours, it appears that acetate is continuously being produced in significant amounts as seen in figure 9.23(c). During the same time period the activity of the oxidative phosphorylation can be seen to go through a maximum of 2.2 mole/C-mole/hr at 39 hours, while the NADH conversion rate seems to be in balance in figures 9.23(b) and 9.23(d), respectively. At 40.8 hours the concentration of ethanol in the offgas exceeds the set point and the substrate feed rate is decreased. Acetate production continues as does the decrease in the activity of the oxidative phosphorylation. A sudden decrease can be seen in the anabolic flux, combined with a slight increase in the glycolytic flux. At 42 hours production of ethanol is measured in the offgas by the standard method using mass spectroscopy. This is coupled to a fast decrease in the glycolytic and acetate fluxes as well as the activity of the oxidative phosphorylation, while the anabolic flux continues to decrease.

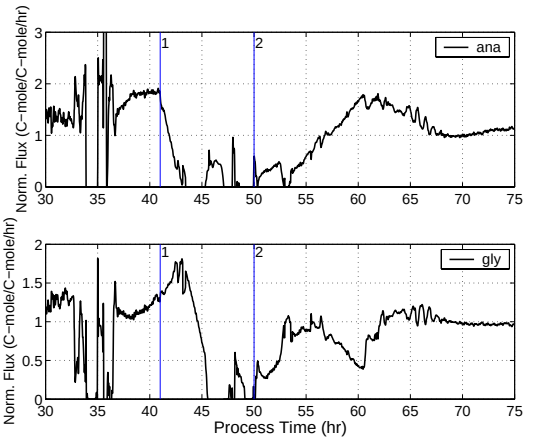
In figure 9.23(d) it is interesting to see the changes to the ATP and NADH balances during the time span from 40 to 45 hours. The production and conversion of the two components is almost balanced at 40.8 hours, after which a fast change occurs first in the ATP balance and a little later in the NADH balance. The feed rate is continuously lowered until 46 hours. While ethanol is being consumed from 45.2 to 47.2 hours, acetate is not consumed until ethanol has disappeared (figure 9.23(c)). At 50 hours the process is restarted and quickly gets back on track as shown in figure 9.22 and figure 9.23.

During continuous operation the activity of the oxidative phosphorylation reaches a constant level around a normalized value of 1.9 moles/moles/hr at 60 hours and remains there for 15 hours. Compared to the observations prior to ethanol formation in the continuous operation of cultivation MTS01, the level reached in cultivation MTF02 is approximately 10% higher than the peak values in cultivation MTS01.

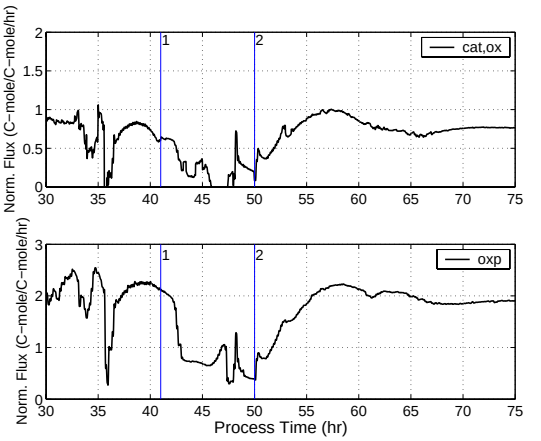
9.5.2.7 Summary and Remarks

The examples from the cultivations MTS01, MTS05, MTE02 and MTF02, show how the *IFM* soft sensor is able to provide a deeper insight into the balancing of intrinsic

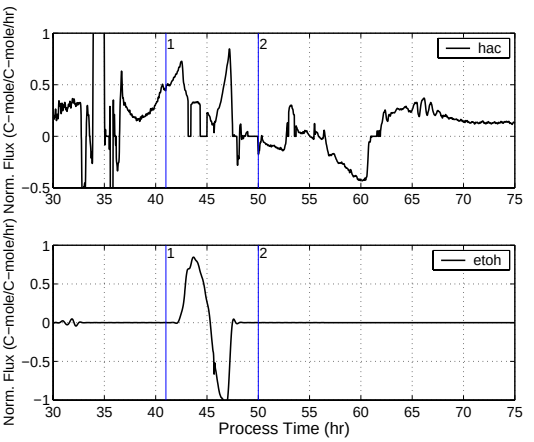
somewhere between the gas outlet of the bioreactor and the position of the Figaro sensor in the exhaust line. The exhaust line was equipped with heat tracing for half of the distance between the gas outlet and the Figaro sensor. In situations where large concentrations of ethanol had been observed in the broth, followed by total consumption and/or dilution, ethanol was for an extended period of time (hours) still being measured by the Figaro sensor in what appeared to be artifacts from slow release of the condensed ethanol in the exhaust line. It was not possible to determine the reason for the problem.



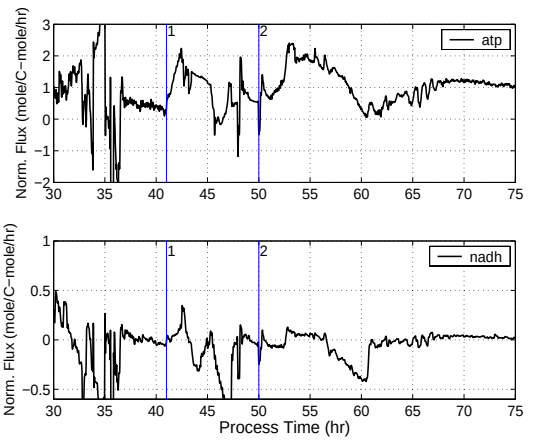
(a) Anabolic (top) and glycolytic flux (bottom)



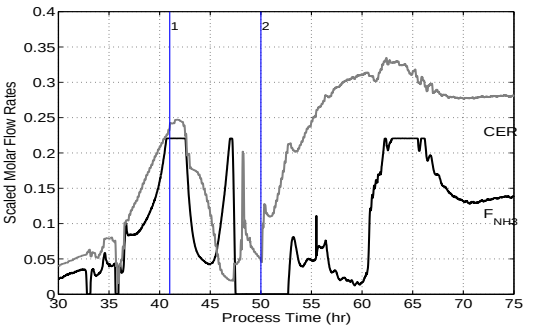
(b) Oxidative catabolic flux (top) and flux through oxidative phosphorylation



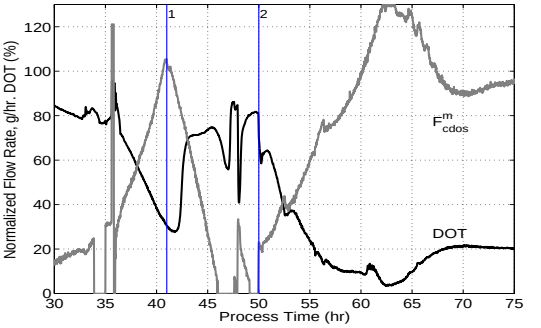
(c) Specific conversion rate of acetic acid (top) and ethanol (bottom)



(d) Specific conversion rate of ATP and NADH



(e) CER (grey line) and ammonia flow rate (F_{NH_3}, black line)



(f) DOT (black line) and glucose syrup flow rate (F^m_{cdos}, grey line)

Figure 9.23. Estimation of intrinsic flux distributions onto the major metabolic pathways. Data from fed-batch operation of cultivation MTF02. For comparison with process variables: (e) CER and ammonia flow rates and (f) DOT and glucose syrup flow rates. Vertical line 1 indicates the time at which feeding of nutrient substrate is initiated, while line 2 indicates the time at which continuous operation is initiated. Note that the flux estimations between 34 and 36 hours are questionable, due to initialization problems of the closed loop controller between the measurement of ethanol and the substrate feed rate during this time period.

fluxes during different process conditions and more interesting during process upsets or disturbances. The estimated flux distributions should be seen as an additional tool for interpretation of experiments, providing information that is based on a microscopic rather than macroscopic approach to interpreting process data.

9.5.3 Modeling Insulin Precursor Production Rate

In section 9.4 a first-order model was suggested for the description of the specific production rate of the insulin precursor, the product:

$$r^m_{p,k} = \alpha_p r_{x,k}\big(1 - exp\big(\frac{-(t_k - t_{fb})q_{x,k}}{\psi}\big)\big) \tag{9.52}$$

The model accounted for the slow adaptation of the specific productivity observed in figure 9.9(a) and a simple adjustment of the model made was used to describe the observed variation related to conversion of acetate. This model was implemented in a soft sensor, the *IPP* soft sensor presented in algorithm 9.3.

The parameters of the model, ψ and α_p, are determined by non-linear least squares regression using data from cultivation MTS07, with minimization of the relative error e between offline measurements and estimates of the insulin precursor concentration, c_p and $\hat{c}_p$ at sampling number i:

$$e_i = \hat{c}_{p,i} - c_{p,i} \tag{9.53}$$

The resulting normalized parameters are:

$$[\psi_p, \alpha_p] = [2.66, 0.059] \tag{9.54}$$

$r^m_p = \alpha_p r_{x,k}$ is then used in an expression as equation 9.8 to obtain the trajectory of the product concentration. To evaluate the performance of the estimations, the root mean square error (RMSEC/V)[13] is used:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{c}^m_{p,i} - c^m_{p,i})^2} \tag{9.55}$$

Using the *IPP* soft sensor with the parameters in equation 9.54 on cultivation MTS07: $\text{RMSEC}_{MTS07} = 0.003$ compared to concentrations ranging between normalized values from 0.004 to 0.2 g/kg.

The performance of the *IPP* soft sensor on validation data is shown in figure 9.24 as estimator number 2 (Est2, dashed grey line) with corresponding specific production rate (q_p, dashed grey line); all cultivations carried out without any process upsets or disturbances. In the plots are also shown the performance of another estimator (1), which represents the performance when the correction for conversion of acetate is not used. From the figure it is seen that the *IPP* soft sensor, estimator 2, gives a reasonable description of the offline measurements of the product concentration in validation data from cultivation MTS05, MTS11 and MTS12 in figures 9.24(b-d), respectively. Since no significant amount of acetate is produced or consumed during the fed-batch and continuous operation of these three cultivations, the performance of estimator 1 and estimator 2 is almost identical.

In figure 9.25, the performance of the two estimators is presented for four cultivations, where process upsets or disturbances were present. Cultivation MTS02, shown in figure 9.25(a) was used to motivate the model adjustment to account for

[13] C: calibration and V: validation.

the presence of acetate in the broth, and is therefore referred to as a modeling batch. The effect of the model adjustment is best seen in the plot showing the ratio between the volumetric production rate of the product and the volumetric production rate of biomass, shown also in figures 9.24 and 9.25. Here it is seen that for cultivation MTS02 in figure 9.25(a) (bottom) a discrepancy between the estimators is most easily observed between 60 and 80 hours, although the trajectories in figure 9.25(a) (top) indicate that the discrepancy already appears at the onset of nutrient substrate feeding (vertical line 1). The adjustment of the model to account for the presence of acetate in the broth (estimator 2) can be seen in figure 9.25(a) (top) to capture the variations in the offline measurements of the product concentration rather well. The RMSECs of the two estimators also highlight the superior performance of estimator 2.

Similar conclusions are made when observing the performance of the two estimators in cultivations MTS01 and MTV03, shown in figures 9.25(b) and 9.25(c), respectively. In figure 9.25(b) a discrepancy between the signal of estimator 2 and the offline measurement at approximately 97 hours can be observed. This discrepancy could be related to the formation and presence of glycerol reported in section 9.5.1.2, figure 9.11(d), a source of variation that has not been included in the model construction.

For the data from cultivation MTV03, shown in figure 9.25(c) a good description of the offline measurements of the product concentration is obtained. The signals of the two estimators are also rather similar.

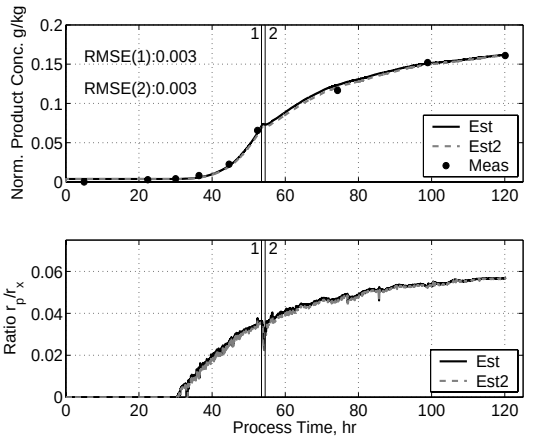
Cultivation MTE02 in figure 9.25(d) shows a rather peculiar behavior of the estimators during fed-batch and continuous operation. In this cultivation an aggressive feeding strategy was applied in the fed-batch phase. From the observed difference between the two estimator in figure 9.25(d) (top) it appears that acetate was estimated to be present in the broth during the fed-batch operation, although offline analysis of the culture broth indicated the acetate levels to be below a normalized concentration of 0.1 g/kg throughout the cultivation. It is furthermore surprising to see that the offline measurements of the product concentration in the continuous operation at 79 and 104 hours are lower than anticipated by both estimators. There is no apparent explanation for these observations.

It was argued in section 9.4.2 that estimated acetate production rather than ethanol production should be used to model the influence of oxido-reductive growth on the observed variations on the specific productivity. To demonstrate the consequences of using the ethanol production ($q_{e,k} > 0$) in algorithm 9.3 rather than acetate production ($q_{a,k} > 0$) calculation have been performed for cultivation MTS01 and MTE02 and these are shown in figure 9.26. A slightly inferior performance is seen in the estimation of the insulin precursor concentration profile in cultivation MTS01 using the ethanol production as indication of oxido-reductive growth, while no significant improvement in the performance in cultivation MTE02 was seen.

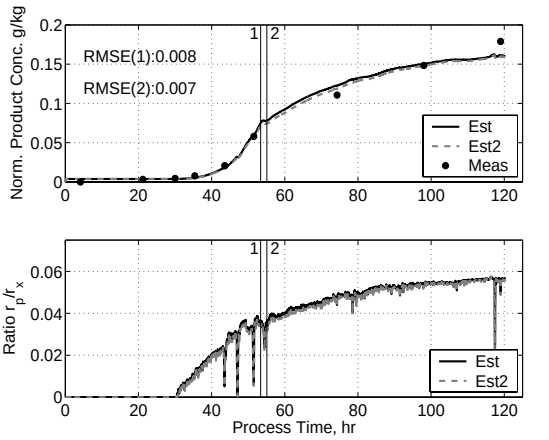
9.5.3.1 Summary and Remarks

In this section it has been demonstrated how a simple soft sensor can be used for providing online estimates of the production rate and concentration of the insulin precursor during a cultivation. In normal standard cultivations a good description is

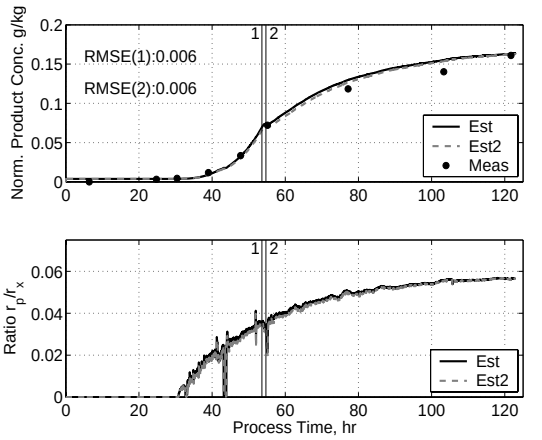
obtained, while in cultivations experiencing disturbances or upsets, the estimation of the product concentration trajectories is reasonable. In one case, cultivation MTE02, a rather weak performance is observed, which can not be explained. Using the ethanol conversion rather than acetate conversion to model the influence of oxido- reductive growth on the development of the specific productivity were shown to provide an inferior description of the offline measurements of the insulin precursor in cultivation MTS01, while the performance in cultivation MTE02 was not improved significantly, compared to the results obtained using acetate conversion as indicator for the growth condition.



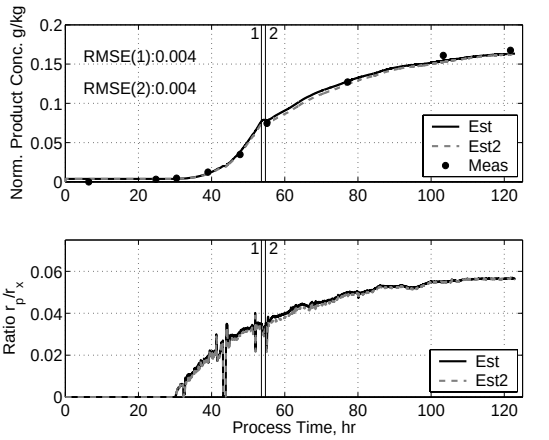
(a) Modeling data set, MTS07.



(b) Validation data set, MTS05.

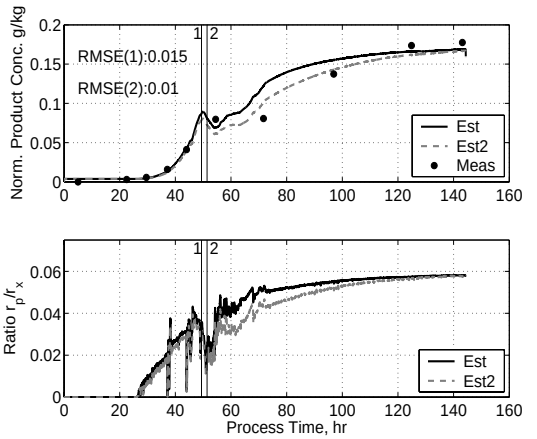


(c) Validation data set, MTS11.

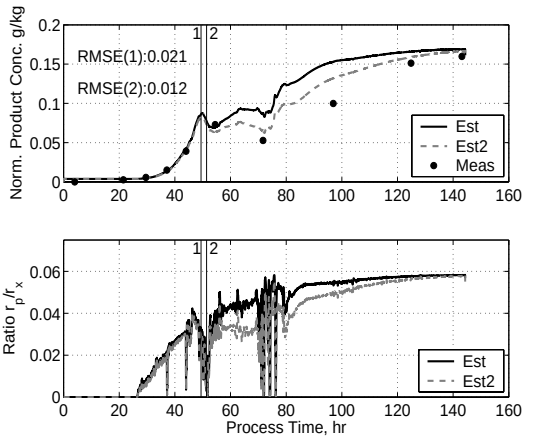


(d) Validation data set, MTS12.

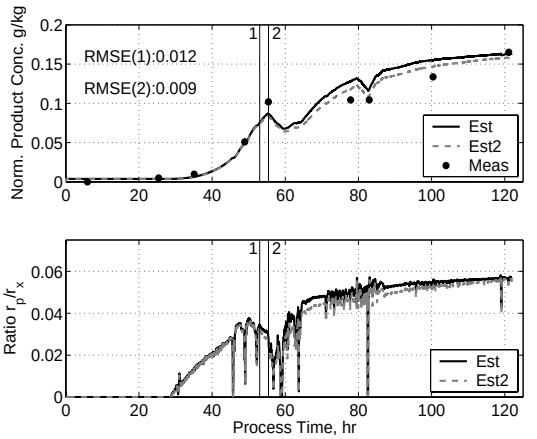
Figure 9.24. Estimation of normalized product concentration (top) and the ratio between the volumetric production rates of product versus biomass (r_p/r_x) (bottom) for cultivations MTS07 (a, modeling), MTS05 (b, validation data), MTS11 (c, validation data), MTS12 (d, validation data). All cultivations were carried out without process disturbances or upsets. The results of two estimators are shown. Estimator 1 (full black line) is based on soft sensor algorithm 9.3 without correction for conversion of acetate, while estimator 2 (dashed grey line) is based on soft sensor algorithm 9.3. •: normalized offline measurements of insulin precursor concentration. RMSE for the estimators are given as RMSE(1) and RMSE(2), respectively. Vertical line 1 indicates the time at which feeding of nutrient substrate is initiated, while line 2 indicates the time at which continuous operation is initiated.



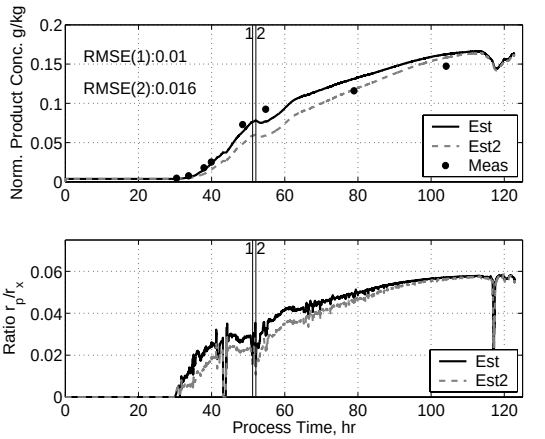
(a) Modeling data set, MTS02.



(b) Validation data set, MTS01.

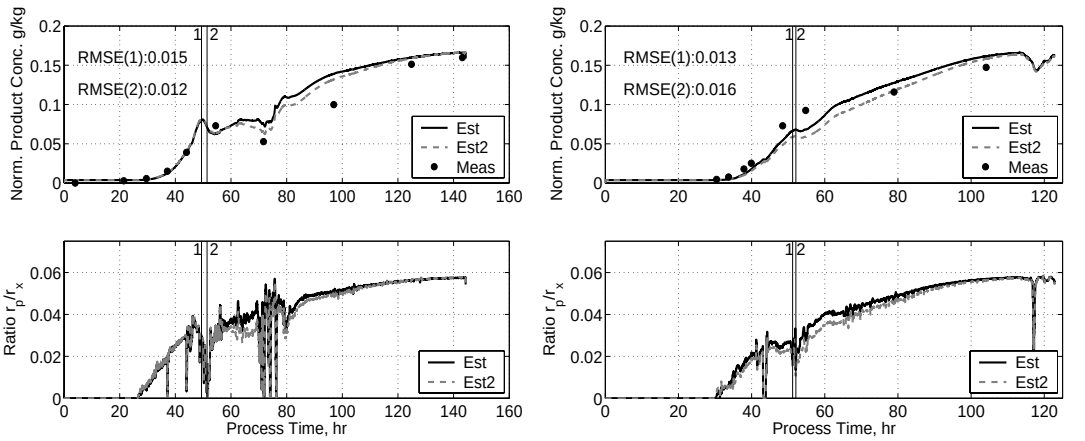


(c) Validation data set, MTV03.



(d) Validation data set, MTE02.

Figure 9.25. Estimation of normalized product concentration (top) and the ratio between the volumetric production rates of product versus biomass (r_p/r_x) (bottom) for cultivations MTS02 (a, modeling), MTS01 (b, validation data), MTV03 (d, validation data), MTE02 (e, validation data). Process disturbances or upsets occurred in these cultivations. The results of two estimators are shown. Estimator 1 (full black line) is based on soft sensor algorithm 9.3 without correction for conversion of acetate, while estimator 2 (dashed grey line) is based on soft sensor algorithm 9.3. •: normalized offline measurements of insulin precursor concentration. RMSE for the estimators are given as RMSE(1) and RMSE(2), respectively. Vertical line 1 indicates the time at which feeding of nutrient substrate is initiated, while line 2 indicates the time at which continuous operation is initiated.



(a) MTS01 (b) MTE02

Figure 9.26. Estimation of normalized product concentration (top) and the ratio between the volumetric production rates of product versus biomass (r_p/r_x) (bottom) for cultivations MTS01 (a) and MTE02 (b), where ethanol production (estimator 1, full black line) and acetate production (estimator, dashed grey line) have been used as an indication for a constant specific productivity. Process disturbances or upsets occurred in these cultivations. •: normalized offline measurements of insulin precursor concentration. RMSE for the estimators are given as RMSE(1) and RMSE(2), respectively. Vertical line 1 indicates the time at which feeding of nutrient substrate is initiated, while line 2 indicates the time at which continuous operation is initiated.

9.6 Discussion

The previous section presented the application of three soft sensors, where it was demonstrated how each of the soft sensors provided new and relevant information on the state of the cultivation. The primary focus in this section will be on the type of information that the models will be able to provide for analysis of the cultivation process and what insight the soft sensors provide into the regulation of the metabolism of the genetically engineered organism. Furthermore the aspects of using the soft sensors in an online application for monitoring and control of cultivation processes will also be discussed.

9.6.1 The Use of Soft Sensors

In the introduction to this chapter it was mentioned how the frequent measurement of key variables *e.g.* once an hour, is seldom carried out in industrial cultivations. This is contradictory to the recognized importance of these measurements for evaluation of process performance and to obtain leads for possible improvement of productivity, robustness or process economy. To some extent the development and introduction of more advanced process analytical technologies for process instrumentation *e.g.* near-infra red spectroscopy, GC-MS instruments[14] or online HPLC, is able to provide information on the key variables. Assuming that such advanced analytical technologies are introduced at the pilot plant or even at the production site, models/soft sensors are still needed in order to interpret the large amounts of data available from these types of instruments and to related them to the process variables that can be manipulated and thereby control and guide the cultivation process towards enhanced performance. Since process data and process dynamics are highly multivariate due to the combination of a complex metabolic network with complex mass transport in the culture broth, multivariate models are needed to capture the import changes in the process conditions. In other words, soft sensors are used to acquire more knowledge from the available process data and therefore constitue an important tool to improve and optimize the use of process information and hence enhance process performance.

Application of soft sensors can take two directions: **semi-hardware** and **innovative**. The **semi-hardware** direction represents an application, where the soft sensor is seen as an instrument continuously providing the same type of information. The name refers to the similarity to hard sensors *e.g.* a pH electrode, which always provides the same type of information as long as it is calibrated and applied within its application range. Focus for this type of application of a soft sensor is production, where robustness and stable performance is important; both online process monitoring and offline process analysis can be the areas of application. The **innovative** direction refers to the use of soft sensors in environments where process development is in focus *i.e.* laboratory and pilot plant. Here the performance of the soft sensor is continuously challenged in order to gain deeper process knowledge and understanding of the mechanisms controlling the microorganism. Again both online process monitoring and offline process analysis can be the areas of application. The semi-hardware and innovative directions highlights the approach to interpretation

[14]GC-MS: gas chromatography-mass spectroscopy

of the information provided by soft sensors:

- Knowledge confirmation. The soft sensor provides signals that are in line with the expected state of the process. This is either evaluated by experience or by comparison to some kind of measurement *e.g.* delayed offline analytical measurements.

- Knowledge rejection. The soft sensor provides signals that are in conflict with the expected state of the process. More information is needed in order to explain what is going on and construct a better model for description of the underlying phenomena.

9.6.2 Model Mismatch

It is straight forward to interpret and evaluate data from soft sensors during periods where a process behaves normally and as expected *i.e.* during knowledge confirmation. It is far more difficult, but at the same time more interesting, to interpret odd looking observations in the signals from the soft sensors or discrepancies between measured and estimated data. All three proposed soft sensors have shortcomings as was also demonstrated and discussed in the result section above (section 9.5).

9.6.2.1 BAC Soft Sensor

In the result section of the BAC soft sensor (section 9.5.1) it was discussed how the observed presence of glycerol had an influence on the estimations of biomass and acetate concentrations, since conversion of glycerol influenced the elemental mass flux balance of carbon. Despite its relevance, it was not found possible to adjust or expand the model to account for the conversion of glycerol, since not enough information was available. It was attempted to use the balance of reduction equivalents (the κ balance) for the estimation of the glycerol conversion rate. This corresponds to assuming the NADH flux balance in the IFM soft sensor to be constant. This extension of the model behind the BAC soft sensor did not prove to be a suitable solution, primarily since the variations in the NADH flux balance did not only originate from the glycerol production. The indication that the extended model did not perform well was best seen from the influence that estimated glycerol production had on estimated production rates of biomass and acetate, as the production rates of the three species were related through the carbon mass flux balance. The estimated concentration profiles of biomass and acetate did not fit the offline measurements.

A second model discrepancy related to the BAC soft sensor, is the use of the threshold value in order to diffentiate between purely oxidative growth and oxido-reductive growth. The reason for including the threshold was unmodeled changes to the biomass composition occurring at the change from batch to fed-batch operation and complex behavior of cellular growth, synchronized growth of subpopulations combined with the transport of weak acids across the plasma membrane under varying process conditions removed from steady state and equilibrium conditions.

9.6.2.2 *IFM* Soft Sensor

In the figures shown in the result section (section 9.5.2), numerous odd fluctuations and variations in the estimated fluxes could be observed, where especially the ATP and NADH conversion rates showed many variations that were difficult to explain. Here it is important to bear in mind the model assumptions presented in chapter 9 and in the materials and methods section (section 9.1), and some of these model assumptions will be discussed below. Before entering the discussion it is important to note that the purpose of the models are primarily to detect that unwanted process behavior occur *e.g.* formation of acetate, and only secondarily to diagnose the reason for the unwanted behavior, *e.g.* a high glucose uptake rate. In other words it is the behavior in cultivation MTS05 (figures 9.15 and 9.16) and in the continuous operation of cultivation MTE02 up until the onset of acetate formation (figure 9.21), that the model should detect as being normal.

It was assumed that process conditions sustained growth where biomass formation and conversion of acetate were the only physiological effects with influence on the proton balance. During process upsets this scenario might not be valid, an example might be that starvation in one or more substrates occurs that can have effects not only during the process upset but also as the process gets back on track. Such a scenario was observed in cultivation MTF02 between 34 and 35 hours and again between 45 and 50 hours where the substrate feeding was stopped for different reasons. Most of the fluxes in figure 9.23 show peculiar behavior during these time periods, and it has not been attempted to explain the observed patterns. Furthermore the need for a minimal threshold value on the acetate conversion rate highlights that assuming constant values of α_x and α_a might not fully capture the actual dynamics of the system.

In the metabolic flux model in table 9.6 assumptions were made on the parameters Y_{xNADH}, P/O and γ_N, as well as the parameters deducted from these namely γ and Y_{xATP}. First it was assumed that all of these parameters were constant throughout the process operation and secondly that values reported from studies of another strain of *S. cerevisiae* could be adopted. One of the consequences of the assumptions above is that the quantitative distribution of the glucose flux into the anabolism and catabolism may be erroneous as Y_{xNADH} and γ influences this distribution, however the qualitative fluctuations in the fluxes through these two major pathways provide information as to how the microorganism reacts to changing growth conditions. The variations in the specific conversion rate of ATP and NADH are also consequences of the assumptions on the parameters. In most of the examples shown in section 9.5.2 the NADH balance appears to close rather well, whereas the ATP balance shows much more variation, indicating either that the parameters related to the specific conversion rate of ATP might be more susceptible to variation or that a component with significant effect on the ATP balance has not been included in the model. In general the information that these two balances provides is primarily how well the model is able to describe the observed behavior. When the specific conversion rates show constant, large discrepancies from being balanced, this indicates that the model estimates should be used with caution.

9.6.2.3 *IPP* Soft Sensor

In case of the results shown for the *IPP* soft sensor in section 9.5.3 it was mentioned that in cultivation MTE02 shown in figure 9.25(d) a discrepancy between the estimated trajectory and the offline measurement of the insulin precursor concentration could be seen. This indicates that the proposed model provides a reasonable description of the productivity during growth conditions which are close to the standard conditions, while other effects need to be included if a more comprehensive model is required. To investigate this, more offline measurements are needed.

From the comparison of the performance of *IPP* soft sensors using ethanol or acetate production as an indicator for oxido-reductive growth in figure 9.26 showed that a slightly better performance is obtained using the acetate production.

9.6.3 Model Describing Acetate Production

The results of chapter 8 indicated that discrepancies between the anticipated and observed ammonia consumption rate appeared just before or during a process upset. In section 9.5.1 it was demonstrated that the large variations observed in the flow rate of ammonia to the bioreactor could be explained by a conversion of acetate. A parameter, α_a, was used to relate the moles of protons necessary to neutralize 1 C-mole of acetate. A value of $\alpha_a = 1/2$ was found to reasonably well describe the observed trends in the offline measurements of acetate. This value of α_a corresponded well with the observations by Casal *et al.* (1996) that during growth on glucose, weak acid transporters in the plasma membrane are repressed and only undissociated acetic acid can diffuse across the membrane. Acetic acid dissociates in the abiotic phase, because the extracellular pH is higher than the pK_a of the weak acid. For each C-mole of acetic acid that dissociates, 1/2 mole of protons are released requiring an equal amount of ammonia to neutralize the acidifying effect of the protons. It remains to be investigated to what extent the closeness of the operational cultivation pH to the pK_a of acetic acid has an effect on the value of α_a. Such an investigation should be combined with a determination of the variations to the elemental composition of the biomass in the different operating regimes of the cultivation.

As described in section 9.2 a threshold of a normalized value of ± 1.5 mmole NH_3/C-mole biomass/hr of the specific ammonia flux (q_n) is used. The threshold value is used to distinguish between the two scenarios:

$$\begin{aligned} &\text{Scenario 1:} \quad r_n = \alpha_x r_x \\ &\text{Scenario 2:} \quad r_n = \alpha_x r_x + \alpha_a r_a \end{aligned}$$

This is used in order to minimize the influence from using a simple model to describe the rather complex behavior of cellular growth, synchronized growth of subpopulations combined with the transport of weak acids across the plasma membrane under varying process conditions removed from steady state and equilibrium conditions. The validation data indicates that the chosen value of α_a results in good descriptions of the offline measurements of acetate, however discrepancies can be seen to occur especially during synchronized growth as pointed out in case of cultivation MTV03 (figures 9.13(a) and 9.14).

Combining the model for acetate conversion with a simple carbon balance furthermore provided an estimate of the biomass production rate. As illustrated in the validation batches of section 9.5.1, this simple model gave a good description of the offline measurements of biomass concentrations. In the case of cultivation MTS01 in figure 9.11(b) it could be seen that choosing another value for α_a ($= 1/3$) would lead to a better description of the biomass concentration, but an inferior description of the acetate concentration. The reason for this discrepancy was explained by the observation that another metabolic product, namely glycerol, was present in significant amounts at the sampling close to 72 hours (see figure 9.11(d)). Taking glycerol conversion into account, combined with the uncertainty on the possible variation in the biomass composition as discussed above, the proposed model of equations 9.40-9.41 with $\alpha_a = 1/2$ and $\alpha_x = cx = 0.146$ provides a reasonable description of both the acetate and biomass conversion rates, evaluated using offline measurements of these components in validation batches MTS05 and MTV03.

9.6.4 Analysis of Intrinsic Metabolic Fluxes

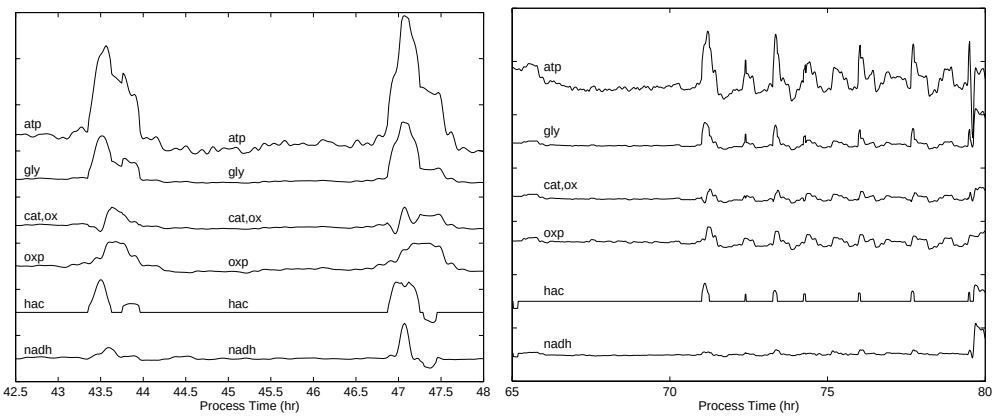
With the modeling of acetate and biomass conversion rates in place, a metabolic flux model was proposed to facilitate the analysis of the intrinsic flux distributions assuming constant pools of the intermediates: glucose-6-phosphate, pyruvate and acetaldehyde. The details of the model are discussed in section 9.3 along with assumptions on constant yield coefficients related to ATP and NADH balances, making it possible to evaluate significant fluctuations in these cofactors.

9.6.4.1 Synchronized Growth, Fast Dynamics

In the presentation of the results obtained using the metabolic flux model on data from cultivation MTS05 it is interesting to notice how synchronized growth occurs during fed-batch operation of the bioreactor, where the estimated intrinsic flux distribution changes during synchronized growth as can be seen in figure 9.15. The first of the synchronizations at 43 hours appear to be triggered by a temperature drop of 0.5°C lasting 10 min. The reason for this drop is the activities related to preparation for sampling. Under normal operation *i.e.* between sampling, a steam circuit ensures that the sampling valve is kept sterile. This setup also contributes energy (heat) to bioreactor; energy which is removed by the cooling jacket. In preparation for sampling of broth, the steam circuit is closed, and as the controller of the cooling system reacts the temperature decreases 0.5°C. The standard deviation in temperature during normal operation is 0.12°C, so even though a temperature variation of 0.5°C seems small it is still significantly different from the normal variations. Interestingly the second and third oscillation are not triggered by sampling preparation.

Surprisingly the model estimates that the anabolic flux decreases significantly as a consequence of synchronization of growth. To understand this behavior it should be borne in mind that the metabolic flux model has been constructed to facilitate monitoring of the dynamic behavior of the cultivation process at a macroscopic level, and is not sufficiently complex to account for the dynamics at the microscopic level of the cell cycle, which becomes apparent during synchronized growth. Münch (1992) describes how the intracellular concentration of carbohydrates *i.e.* trehalose

and glycogen, varies during the cell cycle. He further explains how trehalose is consumed during the S-phase (DNA-synthesis) coinciding with a peak in the CO_2 production rate as well as production of overflow metabolites, in his case ethanol and acetate. In the subsequent G2 (gap-2) and M (mitosis) phases, accumulation of trehalose is reported by Münch as the budding cell prepares to separate into mother and daughter cells. Taking these observations into account, the dips seen in the anabolic flux at 43, 47 and 51 hours are explained by the unmodeled effects from rapid utilization of storage trehalose leading to increased activity of the glycolysis. Furthermore these dips in the anabolic flux has a direct effect on the estimated ATP balance as can be seen from the ATP flux balance equation in table 9.6, and the large fluctuations seen in 9.15(d) are therefore in part due to the effects from trehalose utilization, and in part due to the fluctuations in the activity of the oxidative phosphorylation, the other major contributor to the ATP balance equation.



(a) MTS05 (b) MTV03

Figure 9.27. Intrinsic fluxes as estimated by the IMF soft sensor during periods of operation with synchronized growth in cultivation MTS05 (a) and MTV03 (b). The figures shown the activities of the glycolytic flux (gly), oxidative catabolism (cat,ox), oxidative phosphorylation (oxp) and acetate conversion (hac) as well as the reaction consuming and producing ATP (atp) and NADH (nadh). The distance between the ticks on the y-axis is one normalized unit, as was also used in *e.g.* figure 9.15.

In figure 9.27(a) the estimated intrinsic fluxes during synchronized growth in the fed-batch phase of cultivation MTS05 are shown. From the figure it can be seen that at the onset of the oscillation at 43.3 hours, the initial increase in the glycolytic flux appears to lead to an overflow at the pyruvate branch point and results in production of acetate. Only as the activity of the oxidative phosphorylation is slowly increased, so are the flux through the oxidative catabolism. The dynamic details of this picture is somewhat misleading since the original sampling rate of the data from gas analysis (ethanol, CO_2 and O_2) is every tenth minute, and these data sequences have been subsampled by the process control system to every minute using a zero order hold. This setup leads to an apparent delay in the gas signals, due to the difference in sampling frequencies. The ammonia flow rate used for the determination of the acetate flux is sample every minute and therefore the observed

response to dynamic changes is much faster in this signal than is the case for the data from the gas analysis. With these comments in mind it is still interesting to note how the activity of the oxidative phosphorylation appears to respond more slowly to the elevated glycolytic flux than the conversion rates of acetate and ATP, which can also be seen to be the case for the second oscillation at 46.8 hours.

It is interesting to note that no production of ethanol appeared during the cell cycle oscillations of cultivation MTS05 (see figure 9.15(c)), in contradiction to the observations reported by Münch (1992). This is an interesting observation, since this indicates that alcohol dehydrogenase, the enzyme catalyzing the reaction from acetaldehyde to ethanol, is not expressed under these conditions, since the primary product of the overflow metabolism at the pyruvate branch point is acetate rather than ethanol.

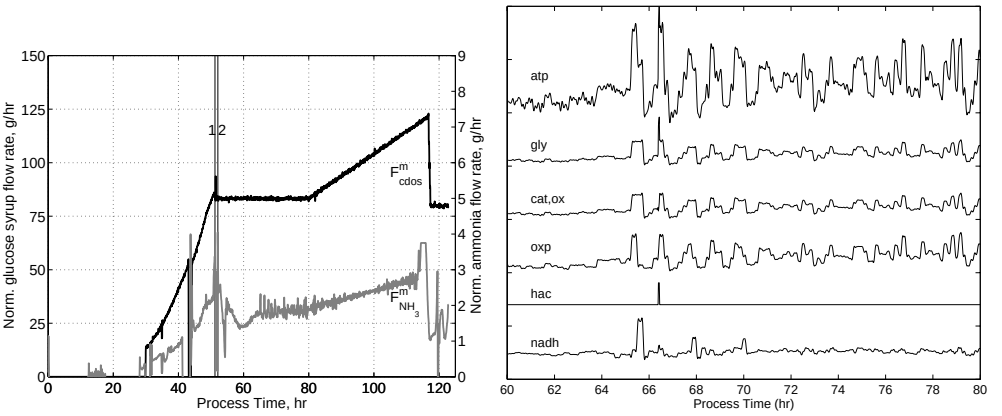
In cultivation MTV03 synchronized growth was reported to occur during continuous operation from 65 to 80 hours and the estimated intrinsic fluxes are shown in figure 9.27(b). This time the onset of synchronization is not related to the sampling procedure, as was the case in cultivation MTS05. As discussed in relation to figure 9.13, ethanol production occured around 60 hours, followed by an abrupt step-down in the dilution rate and a slowly increasing ramp in the dilution rate from 60 to 68 hours, after which continuous operation conditions had been restored. As seen in figure 9.27(b) oscillations suddenly appeared from 71 to 79.5 hours and ceased as a step-up of 50 % in the dilution rate was initiated[15]. During the oscillatory time period 9 peaks can be observed in the ATP balance and the glycolytic flux estimates which is supported by CER and F_{NH_3} measurements in figure 9.14. The peaks are positioned approximately at hours:

$$[71.2,\ 72.4,\ 73.4,\ 74.4,\ 75.3,\ 76.0,\ 76.9,\ 77.7,\ 78.7] \tag{9.56}$$

The peaks in cultivation MTV03 are smaller and more closely positioned than reported in MTS05 with 4 hours between each peak (see figure 9.27(a)). This indicates that what is observed is the effect of synchronization of cellular activities of a number of subpopulations with different sizes and maybe also with different oscillation periods. Prior to the first oscillation there does not seem to be any indications as to what might have triggered the synchronization of the growth, except for the step-down in dilution rate 11 hours before the first oscillation, the slowly increasing ramp over 8 hours and constant continuous operating conditions for 3 hours.

Synchronized growth appearing in the time period 65-80 hours of cultivation MTE02 was different from the synchronized growth seen in both cultivation MTS05 and MTV03. In the case of cultivation MTE02 no severe changes in the glucose syrup substrate feeding rate occurred except for a 1 hour stop of feeding between 43 and 44 hours due to problems with the ammonia supply system and the onset of feeding of nutrient substrate at 52 hours. Cultivation MTE02 was special in the sense that a more aggressive substrate feeding strategy was carried out in the fed-batch operation as discussed in section 9.5.2 in relation to figure 9.20. Figure 9.28(b) show how synchronization of growth appear at 65 hours and then continues,

[15]The variations following this step-up in dilution rate was discussed in section 9.5.1.4. Here it was also shown that oscillations reappeared at 93 hours and lasted for the rest of the cultivation (see figure 9.14(b))



(a) Normalized feeding rate of glucose syrup

(b) MTE02

Figure 9.28. The trajectory of normalized values of glucose syrup feeding rate (a) and a window from 60 to 80 hours of some of the intrinsic fluxes during a period of synchronized growth in cultivation MTE02 (b). Figure (b) shown the activities of the glycolytic flux (gly), oxidative catabolism (cat,ox), oxidative phosphorylation (oxp) and acetate conversion (hac) as well as the balancing of reaction consuming and producing ATP (atp) and NADH (nadh). The distance between the ticks on the y-axis is 1/3 normalized unit; note that this is different from both figure 9.27 and figure 9.21.

although less pronounced after 70 hours, until a ramp in the dilution rate is initiated at 80 hours. Compared to figures 9.27(a) and 9.27(b) the oscillations in figure 9.28(b) are much smaller (approximately 1/3 in amplitude), which indicates that it is much smaller subpopulations that have synchronized their cell cycles, while it appears that a number of subpopulations exists, as was also the case in cultivation MTV03. The peaks are positioned approximately at hours:

$$[65.4,\ 66.4,\ 67.7,\ 68.6,\ 69.6,\ 70.7,\ 71.7,\ 72.5,\ 73.7,\ 74.7,\ 75.6,\ 76.7,\ 77.6,\ 78.7] \tag{9.57}$$

Notice that after 70 hours the identification of the individual peaks becomes difficult.

9.6.4.2 Onset of Ethanol Formation, MTS01

Two periods of ethanol formation occurs after the batch phase in cultivation MTS01 as shown in figure 9.12(d). The two situations are rather different, although in both cases acetate formation is observed prior to the onset of ethanol formation. Focussing on the switch from fed-batch to continuous operation in cultivation MTS01, it was pointed out in section 9.5.2 that the cooling circuit of the bioreactor was out of control from 37 to 45 hours leading to a 3°C increase in the broth temperature, which however was quickly returned to the desired set point at 46 hours. It is not known what the possible effects could be of such temperature variations, however it is possible that these variations have had such an effect on the yeast that the subsequent acetate formation was a combined result of process history and process conditions. Similar responses were observed in three other cultivations (MTS02, MTS03 and MTS04) also experiencing problems in the cooling circuit.

An alternative explanation could be that the capacity of an important metabolic pathway was exceeded. Other investigators have pointed to a limited capacity (a bottleneck) of the enzyme pyruvate dehydrogenase catalyzing the reaction from pyruvate to AcetylCoA leading to an overflow at the pyruvate node in the catabolism (Pronk *et al.*, 1996) or a limited capacity in the oxidative phosphorylation or respiratory capacity (Barford and Hall (1979);Sonnleitner and Käppeli (1986); Sonnleitner and Hahnemann (1994); Herwig (2001)). From the data presented in figure 9.17 it appears that the activity of the oxidative phosphorylation and the oxidative catabolism are not significantly affected by the onset of acetate production, while the NADH balancing equation is only affected momentarily at the onset and otherwise remains in balance until the start of ethanol formation at 50 hours. Shortly after the production of ethanol begins the acetate formation is decreased and the result is only a very small flux through the reductive catabolism ($v_{etoh} + v_{hac}$), while a large flux through the oxidative catabolism is estimated leading to an accumulation of NADH from the activities of the TCA cycle. This would indicate that the maximum capacity of pyruvate dehydrogenase is not the bottleneck at this point, since the sudden increasing flux through the oxidative catabolism would not be possible otherwise. A much slower response at the onset of ethanol production was seen in the activity of the oxidative phosphorylation, and over a time period of 10 hours the activity of the oxidative phosphorylation is increased to a level that leads to reestablishment of the NADH balance.

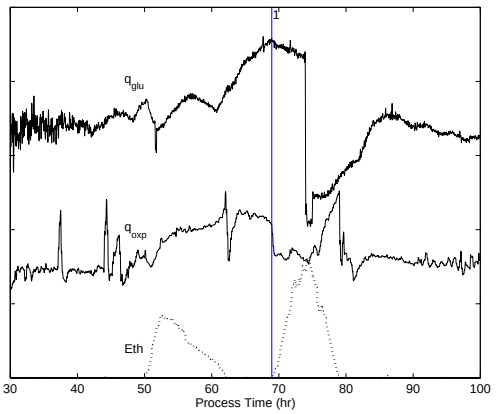


Figure 9.29. Scaled trajectories of the specific glucose uptake rate (q_{glu}=q_{ana}+q_{gly}), the specific activity of the oxidative phosphorylation (q_{oxp}) and the ethanol concentration in the offgas from the bioreactor (Eth). The vertical line (1) indicates the onset of the second ethanol production period. Note that during the time period 60-72 hours the estimated biomass concentration is larger than the observed offline measurements, which means that the specific rates are estimated too low during this time period. Glycerol was observed in an offline sample taken at 72 hours. Data from cultivation MTS01

In figure 9.17 it can be seen that at 48 hours a continuous production of acetate can again be seen to occur and this continues with some fluctuations up until 69 hours (figure 9.18), where ethanol formation starts. This time the mechanism responsible for the observed behavior seems to be different from the ethanol formation at 50 hours. This is highlighted by figure 9.29, where it can be seen how the specific glucose uptake rate (q_{glu}=v_{ana}+v_{gly}) increases continuously from 50 to 69 hours, assuming

insignificant accumulation of glucose in the culture broth. The dilution rate over this time period is more or less held at a constant level, however the production of acetate leads to a decreasing biomass concentration. The rise in specific glucose uptake rate as well as the activity of the oxidative phosphorylation is even more dramatic, keeping in mind that the estimated biomass concentration is larger than the values measured from offline samples as pointed out in section 9.5.1. From figure 9.29 it could appear that it is a maximum in the activity of the oxidative phosphorylation that triggers a metabolic response that leads to ethanol formation, however the peak at 79 hours as well as observations in cultivations MTE02 (figures 9.20 and 9.21) and MTF02 (figure 9.23) suggests that even higher specific activities of the oxidative phosphorylation can be obtained. An alternative explanation would be that it is the onset of ethanol formation that results in a repression of the respiratory capacity, which has also been reported by Lei *et al.* (2003), however at the ethanol formation at 50 hours a repression of the oxidative phosphorylation was not seen and in both cases the repression is not maintained as ethanol is being consumed from 51 to 62 hours and again from 74 to 79 hours, indicating that it is not the presence of ethanol, but some other phenomenon that is responsible for the observed repression at 69 hours. Again the flux through the oxidative catabolism can be seen to react to the production of ethanol and the observed increase in the flux indicates that it is neither a bottleneck in the form of a maximum capacity of pyruvate dehydrogenase nor the sudden ethanol formation that is the reason for the acetate production. At the beginning of the ethanol formation, acetate formation is estimated to cease and revert to a consumption of acetate instead as was shown in figure 9.18(c).

It has been assumed above that the proposed metabolic flux model is sufficient to describe the flux distribution, however in section 3.2, a more complex reaction network was presented. It is possible that acetate under certain conditions is further converted by AcetylCoA-synthetase into AcetylCoA, which is then transported into the mitochondria, where AcetylCoA enters into the TCA cycle and is dissimilated. The series of reactions from pyruvate through acetaldehyde, acetate, AcetylCoA and into the TCA cycle is referred to as the pyruvate dehydrogenase by-pass by Pronk *et al.* (1996), since the combined actions correspond to the function of pyruvate dehydrogenase. The observed variations in the acetate conversion may then be the consequence of the activity of AcetylCoA-synthetase, where repression and derepression is playing a key role in controlling the acetate concentration. With the data at hand it has not been possible to investigate this possible scenario in more detail.

9.6.4.3 Shorter Fed-batch Phase, MTE02

Despite the process upsets occurring during the fed-batch operation of cultivation MTE02 it was possible to use a more aggressive feeding strategy and shorten the fed-batch phase with up to 5 hours compared to cultivation MTS05 when taking the duration of the process upset into account. The applied feeding strategy challenged earlier observations from investigations of the critical growth rate in continuous operation, where it had been concluded that the critical growth rate occurred around a normalized value of 1.4 mole/mole/hr at a cell density corresponding to a normalized value of 20 g/kg (Internal Novo Nordisk Research Report). The results of

cultivation MTE02 showed that a normalized growth rate, μ (or q_{ana}), above 1.4 mole/mole/hr was used for more than 7 hours without ethanol being formed (figure 9.23, 33-40 hours).

During the fed-batch operation the normalized biomass concentration increases from 3 to 20 g/kg and at the same time the estimated normalized growth rate decreases from 1.8 to 1.0 mole/mole/hr (figure 9.20(a)). It is speculated that these observations could be interpreted as the critical growth rate for the onset of oxido-reductive growth being affected by the biomass concentration in high-density cultivations. This is an interesting hypothesis, since it challenges the current fed- batch strategy, suggesting that in the beginning of the fed-batch operation an aggressive feeding strategy will lead to a faster fed-batch phase. The observations in cultivation MTE02 also suggest that an even more aggressive feeding strategy can be used. Ideally investigations of a feeding strategy should be done using a producto-stat suggested by Lei *et al.* (2003). Lei used a control loop to maintain a constant ethanol concentration in the culture broth. For the recombinant strain used in this work, setting up a control structure to maintaining a constant acetate concentration could be used in order to investigate an optimal feeding strategy for a short and effective fed-batch operation.

9.6.4.4 Ramping Dilution Rate, MTE02

Towards the end of cultivation MTE02 a feeding strategy of a slowly increasing dilution rate was applied. Acetate production occurred 1.5 hours before ethanol formation was observed. The acetate production occurred at a specific growth rate, μ (or q_{ana}), of 1.4 mole/mole/hr, which corresponds to earlier observations in continuous operation as reported above. As ethanol starts to be formed the oxidative phosphorylation can be seen to decrease. The flux through the oxidative phosphorylation is not quite as high as was observed in MTS01, and the response to the lowering of the dilution rate following the formation of ethanol indicates that it is not a limitation in the oxidative phosphorylation that seems to be the explanation for the ethanol production, since the oxidative phosphorylation after the lowering of the dilution rate at 117 hours increases to even higher levels than prior to the ethanol formation.

9.6.4.5 Closed Loop Control, MTF02

From the experience gained in cultivation MTE02, closed loop control of the substrate feeding rate was applied in an attempt to obtain not only a shorter fed-batch phase, but also a more robust transition form fed-batch to continuous operation by letting a physiological response in the form of ethanol in the offgas be an indicator of the process state. Acetate was again seen in large quantities prior to formation of ethanol and prior to the proper functioning of the controller, and the activity of the oxidative phosphorylation decrease dramatically at the onset of ethanol formation as did the formation rate of acetate. From figure 9.23(c) it is interesting to note that up until 60 hours acetate appear to be consumed however at 62.5 hours acetate is again being formed. From 50 hours and onwards there is only a modest production of ethanol in agreement with the proper operation of the control loop. It should be

noted that during closed loop operation the oxidative phosphorylation flux reaches 2 mole/mole/hr as seen in figure 9.23(b).

9.6.5 Insulin Precursor Production Rate

From figures 9.24 and 9.25 it could be seen how the trajectories of the offline measurements of the insulin precursor concentration were reasonably well described by the proposed model in equation 9.51.

It was mentioned in section 9.4.2 that the first order behavior of equation 9.49 could be interpreted as a slow buildup of capacity or filling up *storage* for producing and expressing the insulin precursor. The investigation of expression rates of the insulin precursor described by Kjeldsen *et al.* (2001) and reviewed in section 5.2.1, argued that secretion may reflect saturation of a sorting mechanism due to over-expression of the IP or that secretion occurs in competition with intracellular retention. No information has been reported on the size or capacity of this intracellular retention, however it does not seem likely that it is the filling of such an intracellular capacity that is the reason for the time it takes for the specific productivity of the insulin precursor to reach the high level observed in the cultivations without significant process disturbances shown in figure 9.24.

It is clear from Kjeldsen *et al.* (2001) that the proposed model does not capture the full dynamics of the expression system of the insulin precursor, however the model provides valuable information and a starting point for further studies into the improvement and optimization of the productivity. It would be highly interesting to investigate if it is possible to change the operation of the process to obtain a higher specific productivity earlier in the process. This would require investigation of the causes responsible for the observed slow increase in the specific productivity, where detailed studies of the response to fast changes in the process conditions *i.e.* pulse chase and other dynamic studies, will be relevant.

9.6.6 Online Analysis

One of the aims of this work has been to provide a framework and tools to facilitate online monitoring and control of cultivation processes. In the construction of the models it has been important that the measurements used for estimation of concentration profiles and flux distributions were similar to the information available in the online process control system of the production facility of an industrial company. This makes it possible to test and implement the methodologies at the production facility and help turning the focus onto a better utilization of the process information available from the process control system in production.

It was furthermore the aim of this work to implement and test the proposed frameworks and tools, however due to insufficient time for this part of the project, such an implementation was never carried out. Implementation of the proposed modeling framework in a pilot plant facility would facilitate further investigations of the physiology of the recombinant strain, where closed loop control of the substrate feeding rates based on the model outputs could provide new and valuable information about the metabolic control mechanism of the strain.

9.7 Conclusion

This work has proposed and presented 3 soft sensors that in combination provides a framework for dynamic online monitoring of cellular activities in terms of the distribution of intrinsic metabolic fluxes as well as of the balancing of the important cofactors ATP and NADH, combined with an estimation of the product concentration obtained by modeling the productivity of the desired product.

Firstly it was shown how unexpected fluctuations in the flow of ammonia to maintain a constant level of pH in the culture broth could be ascribed to the conversion of acetate. This observation led to a model that was able to estimate the conversion rate of acetate as well as the production rate of biomass, based on a combination of a proton balance (equation 9.40) and a carbon balance (equation 9.41), and using these rates in dynamic mass balances online estimations of acetate and biomass concentrations were provided. Applying the model on data from a number of cultivations, provided a surprising observation namely that acetate was being produced in large amounts 1-2 hours before formation of ethanol occurred. The reason for the onset of the acetate formation has not yet been determined. In order to closer investigate this phenomena, dedicated experiments combined with the use of the prosed soft sensors are needed.

A small metabolic flux model was proposed using calculated and estimated conversion rates of substrate, biomass and key metabolites combined with physiological parameters reported in the open literature on another strain of *Saccharomyces cerevisiae*. The model was used to illustrate and discuss observations from cultivation showing both normal and abnormal process behavior. The model illustrated how acetate was produced prior to ethanol formation. The model also showed how the activity of the oxidative phosphorylation changed extensively as ethanol formation started and as ethanol consumption ended, which was interpreted as effects from repression/derepression of the oxidative phosphorylation. It was not possible to explain what mechanism was responsible for this control of the oxidative phosphorylation, although it was discussed that it could not be a fixed limitation in the capacity of the oxidative phosphorylation, since an experiment using closed loop control of the ethanol concentration in the offgas showed higher activity of the oxidative phosphorylation than were seen in similar open loop experiments.

Finally a simple model was proposed to describe the specific productivity of the product, an insulin precursor. The description was based on a first order model expression for the dependence of production rate and biomass synthesis rate, with a time constant proportional to the specific glucose uptake rate provided by the metabolic flux model presented above. The model gave a reasonable description of the observed trajectories of product concentration in a normal cultivation and by a small extension of the model, it was also able to provide a reasonable estimation of the product concentration profile during process upset in the form of acetate formation.

9.7.1 Further Work

It is highly relevant to conduct further experiments to elucidate the mechanisms behind some of the observations reported in this work. It would be interesting using

the proposed model for online evaluation to conduct additional experiments with the aim of elucidating the mechanisms controlling distribution of metabolic fluxes. It is interesting to determine how and why acetate is being formed. Investigations using the proposed model framework should be combined with more frequent sampling and offline analysis of the culture broth, where analysis of the activity of key enzymes during different process conditions could provide valuable insight in the control of physiological mechanisms related to the reductive catabolism.

It is believed that studies using closed loop control based on physiological responses, as was briefly discussed in relation to cultivation MTF02, will be able to provide valuable information as to the mechanisms leading to changes in the metabolic flux distribution. Closed loop control based on physiological responses is especially interesting in relation to the investigations and screening of genetically modified organisms in order to understand and evaluate the effects of the modifications to the genome and their impact on performance and productivity of a given recombinant strain.

10

Conclusions

10.1 Modeling Production Data

The investigations in chapter 6 illustrated how model-based soft sensors can be used as tools for analysis and monitoring of industrial cultivation processes and enhance the understanding of the processes. Despite using noisy, low quality process data it was possible to provide both online estimates and predictions that would enable an extended supervision and evaluation of cultivations. The online evaluation of biomass and product concentrations are not only valuable for the personnel at the cultivation plant, but also for the subsequent downstream processing, providing information of the expected quantity and quality of the culture broth that is arriving. In continuous cultivations such information becomes even more valuable, since the culture broth is processed while the cultivation is still running and results from offline analysis are not available yet.

10.2 Elemental Composition of Biomass of an Industrial Recombinant Strain of *Saccharomyces cerevisiae*

During the work with process data from a production facility it was found that the current quality of data would render it impossible to construct and evaluate more sophisticated process models. In stead experiments were carried out in pilot plant scale using another production strain of *Saccharomyces cerevisiae*.

In order to use component mass balances for online monitoring of unknown components it was found necessary to estimate the element composition of the biomass during oxidative growth conditions. This was described in chapter 7. Two different approaches were attempted: elemental analysis and macroscopic mass balancing. Inspired by the work of Lange and Heijnen (2001) it was assumed that 3.6 w% of the biomass concentration measured as dry weight was due to water, and an ash content of 3.8 w% accounting for the contributions of metals, sulfur and phosphor to the biomass composition. The following elemental composition of the recombinant strain of *S. cerevisiae* is suggested:

$$CH_{1.82}O_{0.576}N_{0.146} \quad \kappa_x = 4.23; \quad M_{DW} = 27.1 \text{ g/C-mole } (f_r\text{: } 7.3 \text{ w\%}) \qquad (10.1)$$

The results obtained from this investigation also shows that the mass balances close under stationary conditions during continuous operation. This would not have been the case if the biomass composition estimated using macroscopic mass balances

was significantly different from the elemental analysis and the compositions cited in the literature.

10.3 Estimating Ammonia Flow Rates During Oxidative Growth

While conducting experiments in pilot plant surprising variations in the process variables were observed, most noticeable in the ammonia feed flow rate to maintain a constant pH. In chapter 8 a model was proposed to provide a reasonable estimate of the ammonia feed flow rate during balanced oxidative growth, the normal operating conditions. Comparing the model-based estimated flow rate of ammonia to the measured values highlighted a number of discrepancies especially during the switch over from fed-batch to continuous operation. It is believed that these discrepancies can be ascribed to undetermined metabolic activities *i.e.* changes in metabolic flux distribution, which influence the amount of protons transported from the cell to the surrounding broth and subsequently the addition of ammonia to neutralize this effect. Furthermore the use of the proposed model for monitoring of oxidative growth condition was also discussed. A detection limit of $\pm$ 1.5 mmole NH_3/mole biomass/hr for violation of the oxidative growth condition was suggested.

The findings illustrate how model-based soft sensor can be used to provide information on the current state of a cultivation process by comparing expected process behavior, formulated as a model, to the observed process behavior. Discrepancies between the two signals can then be analyzed in order to extend the understanding of the system.

10.4 Soft Sensors for Estimating Biomass and Acetate Concentrations

The construction of 3 models for dynamic online monitoring of cellular activities were described in chapter 9.

The unexpected fluctuations in the flow of ammonia to maintain a constant level of pH in the culture broth, as described in chapter 8, could be ascribed to the conversion of acetate. This observation led to the development of a model in section 9.2 that was able to estimate the conversion rate of acetate as well as the production rate of biomass, based on a combination of a proton balance and a carbon balance, and using these rates in dynamic mass balances online estimations of acetate and biomass concentrations were provided. The large production of acetate was quite surprising and applying the model on data from a number of cultivations, provided a surprising observation namely that acetate was being produced in large amounts 1-2 hours before formation of ethanol occurred. It is not yet known why the acetate formation occur, however with the model-based soft sensor at hand it is now possible to monitor the onset of the acetate formation as well as the acetate and biomass concentration. This information is valuable in order to avoid the acetate formation or to provide a robust and efficient procedure to get the system back into the desired operation regime, balanced oxidative growth.

The availability of online estimates of the biomass concentration will make it possible to apply more aggressive, yet robust, feeding profiles during fed-batch operation shortening the time before the production phase, the continuous operation, is initiated.

10.5 Soft Sensors for Estimating Intrinsic Metabolic Fluxes

In process development, the information of the biomass concentration can also be used for more detailed analysis of the physiology of the recombinant production strain. This was attempted in section 9.3 where a small metabolic flux model was proposed using calculated and estimated conversion rates of substrate, biomass and key metabolites combined with physiological parameters reported in the open literature on another strain of *Saccharomyces cerevisiae*.

The model was used to illustrate and discuss observations from cultivation showing both normal and abnormal process behavior. As a soft sensor the model was able to provide estimates for online monitoring of the intrinsic metabolic flux distributions, highlighting that acetate was produced prior to ethanol formation and how the activity of the oxidative phosphorylation changed extensively as ethanol formation started and as ethanol consumption ended.

Having such soft sensors available during process development would support the investigations of the process boundaries in order to construct control strategies for obtaining more robust operating procedures. Applying and/or improving the models of the soft sensors in the production environment could provide a unique possibility to monitor, control and optimize process performance.

10.6 Soft Sensors for Estimating Product Concentration

Finally a simple model was proposed in section 9.4 to describe the specific productivity of the product, an insulin precursor. The analysis of the specific productivity showed surprising features, namely a slow adaptation of the specific production rate and impact on this from process upsets such as the formation of acetate. The model gave a reasonable description of the observed trajectories of product concentration in a normal cultivation and through a small extension of the model, it was also able to provide a reasonable estimation of the product concentration profile during process upset in the form of acetate formation.

10.7 Soft Sensors for Improved Process Understanding

In this thesis it has been demonstrated that the construction and application of soft sensors provide new and valuable information on the dynamics of cultivation

processes. The modelbased analysis of cultivation process data has led to the surprising discovery that large amounts of acetate were sometimes being formed, especially prior to the onset of ethanol formation. Constructing models with increasing complexity provided the means for estimating the variations in the biomass and acetate concentrations, and furthermore to estimate how changes to the intrinsic metabolic flux distribution affected the performance of the cultivation.

The approach to modeling of cultivation data that has been used in this thesis, is in line with the main idea behind PAT (see chapter 2), namely that combining data and knowledge to improve process understanding is the correct way to ensure product quality. Using the terminology of the information hierarchy presented in figure 2.2 on page 11, the soft sensors condensate the *data* provided by the process measurements into *information* and *knowledge* of the physiology and productivity of the production strain.

The applicability of the information and knowledge now available is extensive. Using the new information to formulate (better) monitoring algorithms and control strategies is obvious, as is the use of new information and knowledge for process optimization purposes. Other interesting applications are the use of the information in other plants, most likely in downstream processing, but also in the research and development departments where an explanation for the observed production of acetate as the initial metabolic product formed during onset of oxidoreductive growth can be investigated.

10.8 Outlook and Further Work

The investigations carried out in this thesis have demonstrated the potential of model-based monitoring and control in the cultivation processes of biopharmaceutical company, Novo Nordisk. Already some of the ideas developed have been taken up and are being pursued, however there are still many possible model applications that have not yet been attempted or addressed.

In this thesis the model construction, evaluation and validation has been carried out offline. It remains to be tested how the soft sensors perform in online environments both in pilot plant and in production. It can be expected that such testing will lead to the uncovering of shortcomings in the soft sensors, since process conditions and equipment can have changed since the model development and unforeseen process disturbances might trigger new and unexpected metabolic responses. During the online testing it is important to involve plant personnel in the evaluation of the tools in order to assure that the tools are accepted and understood by the people who are expected to use them daily. Furthermore the involvement of quality assurance and quality control staff is also important in order to extend the anticipation of model-based monitoring and control and to address possible conflicts with quality assurance guidelines as early as possible.

If an online evaluation of the soft sensors proves to be successful, it would be attractive to investigate the possibility of applying control loops based on the signals provided by the soft sensors. Experiments in pilot plant could be used to illustrate the potential and providing experience for extended physiological studies of production strains and formulation of enhanced, yet robust, production process with

the aim of optimizing the productivity. In the development of new products and processes it would be desirable to address the process modeling for monitoring and control as early as possible in the development project. Remember that the models probably have to be changed, recalibrated or reconstructed as new aspects come into play during scale up of processes. However the analysis of the required changes to a model can be used to shed further light on the mechanisms that are responsible for process variations.

Another area that can be addressed in future work is the use of the information provided by the soft sensors in the downstream processes in order to estimate the quality and quantity of inputs to these processes. Taking a more holistic approach to optimization of the production pipeline can lead to improved overall process economics by sharing of information and experience in order to limit the effects of bottlenecks and optimize equipment utilization.

A big question is where the responsibility of model construction, validation and maintenance should be placed in the company organization. With the process analytical technology (PAT) initiative by U.S. Food and Drug Administration it can be expected that eventually model-based process monitoring and control will be given much more attention in the biopharmaceutical industry. It is desirable that the personnel using models in the shape of soft sensors on a daily basis have a sense of responsibility for the tools, however it is unlikely that plant floor personnel can be expected to have high expertise in both cultivation processes and modeling at the same time. A better approach might be to set up a central modeling team that can support the plant personnel in the maintenance and supervision of the models and at the same time head project teams in the development of new models and soft sensors. The best solution for such organization remains to be investigated and tested as does the aspects for iterative model revalidation and calibration.

Appendices

Abbreviations

A-data	additional batches used in chapter 6
ADP	adenosine diphosphate
AMP	adenosine monophosphate
API	active pharmaceutical ingredient
ARMAX	autoregressive moving average with exogenous input
ATP	adenosine triphosphate
BAC	software sensor for estimating biomass and acetate concentrations
cdos	primary substrate: glucose syrup
cGMP	current good manufacturing practice
CoA	coenzyme A
CUSUM	cumulated sum
CVA	canonical variate analysis
DNA	deoxyribonucleic acids
DOT	dissolved oxygen tension
DW	dry weight biomass
EA	elemental analysis
EMP	Embden-Meyerhof-Parnas
ER	endoplasmic reticulum
FAD	flavin adenine dinucleotide
FPEM	first principles engineering model
GDP	guanosine diphosphate
GMO	genetically modified organism
GRAS	generally regarded as safe
GTP	guanosine triphosphate
HPLC	high pressure liquid chromatography
IMF	software sensor for estimating intrinsic metabolic fluxes
IP	insulin precursor
IPP	software sensor for estimating the insulin precursor production rate and concentration
M-data	modeling batches used in chapter 6
MES	manufacturing execution system
MIMO	multiple inputs/multiple outputs
MIS	manufacturing information system
MMB	macroscopic mass balance
MPC	model predictive control
MPLS	multiway projection to latent structures
MTE	Cultivation code, series 2
MTF	Cultivation code, series 3
MTS	Cultivation code, series 1
MTV	Cultivation code, series 4

NADH	nicotinamide adenine dinucleotide
NADPH	nicotinamide adenine dinucleotide phosphate
ndos	secondary substrate: nutrients
P_i	inorganic phosphate
P/O	the ratio of phosphate to oxygen in the oxidative phosphorylation
PAT	process analytical technologies
PCA	principle component analysis
PCR	principle component regression
PCS	process control system
PLC	programmable logic controller
PLS	projection to latent structures
POT	tpi from *Schizosaccharomyces pompe*
PP	pentose phosphate
RMSE(C/P/V)	root mean square error of calibration/prediction/validation
RNA	ribonucleic acids
RQ	respiratory quotient
SDE	stochastic differential equation
SISO	single input/single output
TCA	tricarboxylic acid
TPI	triose phosphate isomerase
UDP-NAG	uradine diphosphate N-acetylglucosamine
UDP-NAM	uradine diphosphate N-acetylmuramic acid
USFDA	U.S. Federal Drug Administration
V-data	validation batches used in chapter 6
wt	wild type

List of Symbols

$\boldsymbol{A}$	Kalman filter model parameter matrix
ax	mole content of hydrogen in biomass per C-mole biomass
bx	mole content of oxygen in biomass per C-mole biomass
$\boldsymbol{B}$	Kalman filter model parameter matrix
B	regression matrix
c	concentration [mole/L, g/L, g/kg]
C	number of principle components
$\boldsymbol{C}$	Kalman filter model parameter matrix
CER	CO_2 evolution rate [mole/hr]
cx	mole content of nitrogen in biomass per C-mole biomass
CXC	difference between off-gas outlet and air inlet concentration of CO_2 [%]
D	dilution rate [L/L/hr, kg/kg/hr]
DOT	dissolved oxygen tension [%]
e	residual
F	flow rate [L/s, kg/min]
f	function; fraction
I	number of batches
J	number of variables; objective function
K	number of data/sample points
$\boldsymbol{K}$	Kalman filter gain matrix
K_m	saturation constant [mM]
M	molar [g/mole]
OUR	oxygen uptake rate [mole/hr]
OXC	difference between air inlet and off-gas outlet concentration of oxygen [%]
p	loading vector
$\boldsymbol{P}$	Kalman filter state variance matrix
q	specific conversion rate [mole/(g dry weight biomass)/hr, mole/mole/hr]
q	loading vector
$\boldsymbol{Q}$	covariance matrix of state noise
q_{s,H^+}	proton equivalents [mmole H^+/L]
r	volumetric conversion rate [mole/L/hr, mole/kg/hr]
$\boldsymbol{R}$	Kalman filter output variance matrix
res	residual
$\boldsymbol{S}$	covariance matrix of measurement noise
t	time [s, min, hr]
t	score vector
$t_{1/2}$	halftime [s, min, hr]
t_{fb}	time for start of fed-batch operation
u	score vector
V	uptake rate [nmole/s/(mg dry weight)], volume [L]
v	specific intracellular flux
w	weighting vector

W	weight [kg]
$w\%$	weight percent [g/g] $\cdot$ 100
x	biomass concentration [g DW/L, gDW/kg]; state in Kalman filter
y	mole fraction in gas phase; measurement
Y_{sx}	yield coefficient of biomass on substrate [C-mole/C-mole]
Y_{xATP}	yield coefficient of ATP per biomass [mole ATP /C-mole]
Y_{xH}	yield coefficient of protons per biomass [mole H^+ /C-mole]
Y_{xNADH}	yield coefficient of NADH per biomass [mole NADH /C-mole]
α	model parameter
β	model parameter
γ	model parameter
κ	degree of reduction
μ	specific growth rate [L/L/hr, kg/kg/hr]
ω	frequency [s^{-1}, min^{-1}, hr^{-1}]
ψ	model parameter
ρ	density [kg/L]
τ	characteristic time for adaptation to specific productivity
θ	model parameter; temperature [$^\circ C$]
ζ	filter model parameter

Superscipts and subscripts

$\hat{\cdot}$	estimate
a	ash
a	acetate
ana	anabolic reaction
b	broth
c	carbon; cdos
cat, ox	oxidative catabolism
cat, red	reductive catabolism
$crit$	critical
DW	dry weight biomass
e	effluent; ethanol
$etoh$	formation of ethanol
f	feed
g	gas phase; gauge
g	glycerol
glc	glucose uptake
gly	glycolysis
hac	formation of acetate
in	inlet
m	mass
max	maximum
n	ammonia; ndos
o	oxygen
out	outlet
oxp	oxidative phosphorylation
p	product; physiological contribution
r	residual
s	substrate; glucose; sampling
t	top
t_{fb}	time for start of fed-batch operation
v	vessel property contribution
w	water
x	biomass

References

Åström, K. J. and Wittenmark, B. (1995). *Adaptive Control*. Addison-Wesley Publishing Company, second edition.

Barford, J. and Hall, R. (1979). An Examination of the Crabtree Effect in *Saccharomyces cerevisiae*; the Role of Respiration Adaptation. *Journal of General Microbiology*, **114**, 267–275.

Casal, M.; Cardoso, H. and Leão, C. (1996). Mechanisms regulating the transport of acetic acid in *Saccharomyces cerevisiae*. *Microbiology*, **142**, 1385–1390.

Casal, M.; Cardoso, H. and Leão, C. (1998). Effects of Ethanol and Other Alkanols on Transport of Acetic acid in *Saccharomyces cerevisiae*. *Applied and Environmental Microbiology*, **64**(2), 665–668.

Castrillo, J.; de Miguel, I. and Ugalde, U. (1995). Proton production and consumption pathways in yeast metabolism. A chemostat culture. *Yeast*, **11**, 1353–1365.

Compagno, C.; Bonschi, F. and Ranzi, B. M. (1996). Glycerol Production in a Triose Phophate Isomerase Deficient Mutant of *Saccharomyces cerevisiae*. *Biotechnology Progress*, **12**, 591–595.

Diers, I.; Rasmussen, E.; Larsen, P. and Kjærsig, I.-L. (1991). *Yeast Fermentation Processes for Insulin Production*, pages 166–176. Marcel Dekker, New York, NY.

Duboc, P. and von Stockar, U. (1998). Systematic Errors in Data Evaluation Due to Ethanol Stripping and Water Vaporization. *Biotechnology and Bioengineering*, **58**(4), 428–439.

Duboc, P.; Stockar, U. v. and Villadsen, J. (1998). Simple Generic Model for Dynamic Experiments with *Saccharomyces cerevisiae* in Continuous Culture: Decoupling between the Anabolism and the Catabolism. *Biotechnology and Bioengineering*, **60**, 180–189.

Egel-Mitani, M.; Hansen, M.; Norris, K.; Snel, L. and Fiil, N. (1988). Competitive expression of two heterologous genes inserted into one plasmid in *Saccharomyces cerevisiae*. *Gene*, **73**, 113–120.

Gregersen, L. and Jørgensen, S. (1999). Supervision of fed-batch fermentations. *Chemical Engineering Journal*, **75**, 69–76.

Hastie, T.; Tibshirani, R. and Friedman, J. (2001). *The Elements of Statistical Learning. Data Mining, Inference and Prediction.*. Springer Verlag, New York.

Herwig, C. (2001). *On-line Exploitation Tools for the Quantitative Analysis of Metabolic Regulations in Microbial Cultures*. Ph.D. thesis, The Department of Chemical Engineering, Swiss Federal Institute of Technology, Lausanne.

Herwig, C. and von Stockar, U. (2002). A Small Metabolic Flux Model to Identify Transient Metabolic Regulations in *Saccharomyces cerevisiae*. *Bioprocess and Biosystems Engineering*, **24**, 395–403.

Herwig, C.; Marison, I. and von Stockar, U. (2001). On-line stoichiometry and identification of metabolic state under dynamic process conditions. *Biotechnology and Bioengineering*, **75**(3), 345–354.

Jochumsen, K. (1995). *Production of Proteinase A by Saccharomyces cerevisiae*. Ph.D. thesis, Department of Biotechnology, The Technical University of Denmark.

Kawasaki, G. and Bell, L. (1999). Stable DNA Construct. United States Patent.

Kispal, G.; Cseko, J.; Alkonyi, I. and Sandor, A. (1991). Isolation and Characterization of Carnitine Acetyltransferase from *S. cerevisiae*. *Biochimica et Biophysica Acta*, **1085**, 217–222.

Kjeldsen, T. (2000). *Yeast Secretory Expression of Insulin Precursors*. Doctoral thesis, Technical University of Denmark.

Kjeldsen, T.; Pettersson, A. and Hach, M. (1999). The role of leaders in intracellular transport and secretion of the insulin precursor in the yeast *Saccharomyces cerevisiae*. *Journal of Biotechnology*, **75**, 195–208.

Kjeldsen, T.; Balschmidt, P.; Diers, I.; Hach, M.; Kaarsholm, N. C. and Ludvigsen, S. (2001). Expression of Insulin in Yeast: The Importance of Molecular Adaptation for Secretion and Conversion. *Biotechnology and Genetic Engineering Review*, **18**, 89–121.

Kristensen, N. R. (2003). *Fed-Batch Process Modelling for State Estimation and Optimal Control*. Ph.D. thesis, Department of Chemical Engineering, Technical University of Denmark.

Lange, H. and Heijnen, J. (2001). Statistical Reconciliation of the Elemental and Molecular Biomass Composition of *Saccharomyces cerevisiae*. *Biotechnology and Bioengineering*, **75**(3), 334–344.

Larsson, C.; Påhlman, I. and Gustafson, L. (2000). The Importance of ATP as a regulator of glycolytic flux in *Saccharomyces cerevisiae*. *Yeast*, **16**, 797–809.

Lei, F. (2001). *Dynamics and non-linear phenomena in continuous cultivations of* Saccharomyces cerevisiae.. Ph.D. thesis, Department of Chemical Engineering, Technical University of Denmark.

Lei, F.; Olsson, L. and Bay Jørgensen, S. (2003). Experimental Investigations of Multiple Steady-states in Aerobic Continuous Cultivations of *Saccharomyces cerevisiae*. *Biotechnology and Biochemical Engineering*, **82**, 766–777.

Lei, F.; Olsson, L. and Bay Jørgensen, S. (2004). Dynamic Effects Related to Steady-state Multiplicity in Continuous *Saccharomyces cerevisiae* Cultivations. *Biotechnology and Biochemical Engineering*, **88**, 838–848.

Louverse, D.; Tates, A.; Smilde, A.; Koot, G. and Berndt, H. (1999). PLS discriminant analysis with contribution plots to determine differences between parallel batch reactors in the process industry. *Chemometrics and Intelligent Laboratory Systems*, **46**, 197–206.

Madsen, H. and Holst, J. (2000). *Modelling Non-Linear and Non-Stationary Time Series*. Informatics and Mathematical Modelling, DTU, Denmark.

Makuc, J.; Paiva, S.; Schauen, M.; Krämer, R.; Andr, B.; Casal, M.; Leãao, C. and Boles, E. (2001). The putative monocarboxylase permeases of the yeast *Saccharomyces cerevisiae* do not transport monocarboxylic acids across the plasma membrane. *Yeast*, **18**, 1131–1143.

Münch, T. (1992). *Zellzyklusdynamik von* Saccharomyces cerevisiae *in Bioprozessen*. Ph.D. thesis, Eidgenössischen Technichen Hochschule Zürich.

Nielsen, J. and Villadsen, J. (1994). *Bioreaction Engineering Principles*. Plenum Press, New York, first edition.

Nomikos, P. and MacGregor, J. F. (1995). Multivariate SPC Charts for Monitoring Batch Processes. *Technometrics*, **37**(1), 41–59.

Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. *Organization Science*, **5**(1), 14–37.

Olsson, L. and Nielsen, J. (1997). On-line and *in-situ* monitoring of biomass in submerged cultivations. *Tibtech*, **15**, 517–522.

Pampulha, M. E. and Loureiro, V. (1989). Interaction of the effects of acetic acid and ethanol on inhibition of fermentation in *Saccharomyces cerevisiae*. *Biotechnology Letters*, **11**(4), 269–274.

Pampulha, M. E. and Loureiro-Dias, M. C. (2000). Energetics of the effect of acetic acid on growth of *Saccharomyces cerevisiae*. *FEMS Microbiology Letters*, **184**, 69–72.

Pham, H.; Larsson, G. and Enfors, S.-O. (1998). Growth and Energy Metabolism in Aerobic Fed-batch Cultures of *Saccharomyces cerevisiae*: Simulation and Model Verification. *Biotechnology and Bioengineering*, **60**, 474–482.

Pham, H.; Larsson, G. and Enfors, S.-O. (1999). Modelling of aerobic growth of *Saccharomyces cerevisiae*. *Bioprocess Engineering*, **20**, 537–544.

Pons, M.-N.; Rajab, A. and Engasser, J.-M. (1986). Influcence of acetate on growth kinetics and production control of *Saccharomyces cerevisiae* on glucose and ethanol. *Applied Microbiology and Biotechnology*, **24**, 193–198.

Postma, E.; Scheffers, W. A. and van Dijken, J. P. (1989a). Kinetic of Growth and Glucose Transport in Glucose-limited Cultures of *Saccharomyces cerevisiae* CBS 8066. *Yeast*, **5**, 159–165.

Postma, E.; Verduyn, C.; Scheffers, W. A. and Van Dijken, J. P. (1989b). Enzymatic Analysis of the Craptree Effect in Glucose-Limited Chemostat Cultures of *Saccharomyces cerevisiae*. *Applied and Environmental Microbiology*, **55**(2), 468 – 477.

Pronk, J.; Steensma, H. and van Dijken, J. (1996). Pyruvate Metabolism in *Saccharomyces cerevisiae*. *Yeast*, **12**, 1607–1633.

Remize, F.; Andrieu, E. and Dequin, S. (2000). Engineering of the pyruvate dehydrogenase bypass in Saccharomyces cerevisiae: role of the cytosolic Mg^{2+} and mitochondrial K^{+} acetaldehyde dehydrogenases Ald6p and Ald4p in acetate formation during alcoholic fermentation. *Applied and Environmental Microbiology*, **66**(8), 3151–3159.

Roels, J. A. (1983). *Energetics and Kinetics in Biotechnology*. Elsevier Biomedical Press.

Schulze, U. (1995). *Anaerobic Physiology of* Saccharomyces cerevisiae.. Ph.D. thesis, Department of Biotechnology, Technical University of Denmark.

Siano, S. A. (1995). On the Use of the pH Control Reagent Addition Rate for Fermentation Monitoring. *Biotechnology and Bioengineering*, **47**, 651–665.

Sonnleitner, B. and Hahnemann, U. (1994). Dynamics of the respiratory bottleneck of *Saccharomyces cerevisiae*. *Journal of Biotechnology*, **38**, 63–79.

Sonnleitner, B. and Käppeli, O. (1986). Growth of *Saccharomyces cerevisiae* is controlled by its limited respiratory capacity; formulation and verification of a hypothesis. *Biotechnology and Bioengineering*, **28**, 927–937.

Stephanopoulos, G. N.; Aristidou, A. A. and Nielsen, J. (1998). *Metabolic Engineering. Principles and Methodologies*. Academic Press,.

Stryer, L. (1995). *Biochemistry*. W. H. Freeman and Company, New York, fourth edition edition.

Stückrath, I.; Lange, H. C.; Kötter, P.; van Gulik, W. M.; Entian, K.-D. and Heijnen, J. (2002). Characterization of Null Mutants of the Glyoxylate Cycle and Gluconeogenic Enzymes in *Saccharomyces cerevisiae* Through Metabolic Network Modeling Verified by Chemostat Cultivation. *Biotechnology and Bioengineering*, **77**, 61–72.

Swiegers, J.; Dippenaar, N.; Pretorius, I. S. and Bauer, F. F. (2001). Carnitine-dependent Metabolic Activities in *Saccharomyces cerevisiae*: three carnitine acetyltransferases are essential in a carnitine-dependent strain. *Yeast*, **18**, 585–595.

USFDA (2004a). Pharmaceutical CGMPs for the 21st Century - A Risk-Based Approach. Final Report. Technical report, US. Department for Health and Human Services, US. Food & Drug Administration.

USFDA (2004b). Guidance for Industry. PAT - A Framework for Innovative Pharmaceutical Development, Manufacuring and Quality Assurance. Technical report, US. Department for Health and Human Services, US. Food & Drug Administration.

van Gulik, W. M. and Heijnen, J. J. (1995). A Metabolic Network Stoichiometry Analysis of Microbial Growth and Product Formation. *Biotechnology and Bioengineering*, **48**, 681–698.

van Hoek, P.; van Dijken, J. P. and Pronk, J. T. (2000). Regulation of fermentative capacity and levels of glycolytic enzymes in chemostat cultures of *Saccharomyces cerevisiae*. *Enzyme and Microbial Technology*, **26**, 724–736.

Vanrolleghem, P. A.; de Jong-Gubbels, P.; van Gulik, W. M.; Pronk, J. T.; van Dijken, J. P. and Heijnen, S. (1996). Validation of a Metabolic Network for *Saccharomyces cerevisiae* Using Mixed Substrate Studies. *Biotechnology Progress*, **12**, 434–448.

Verduyn, C. (1991). Physiology of yeasts in relation to biomass yields. *Antonie van Leeuwenhoek*, **60**, 325–353.

Verduyn, C.; Postma, E.; A., S. W. and van Dijken J. P. (1990). Physiology of *Saccharomyces cerevisiae* in Anaerobic Glucose Limited Chemostat Cultures. *Journal of Genetical Microbiology*, **136**, 395–403.

Verduyn, C.; Postma, E.; A., S. W. and van Dijken J. P. (1992). Effect of Benzoic Acid on Metabolic Fluxes in Yeasts: A Continuous-Culture Study on the Regulation of Respiration and Alcoholic Fermentation. *Yeast*, **8**, 501–517.

Walker, G. M. (1998). *Yeast physiology and Biotechnology*. John Wiley & Sons.

Walsh, M. C.; Smits, H. P.; Scholte, M. and van Dam, K. (1994). Affinity of Glucose Transport in *Saccharomyces cerevisiae* is Modulated during Growth on Glucose. *Journal of Bacteriology*, **176**(4), 953–958.

Working Party on Control of Medicines and Inspections (2001). Final Version of Annex 17 to the EU Guide to Good Manufacturing Practice: Parametric Release. Technical report, European Commission, Enterprise Directorate-General.