Development of a model-based tool for the design of biotechnological processes under consideration of effects caused by heterogeneities

Vom Promotionsschuss der
Technischen Universität Hamburg

zur Erlangung des akademischen Grades

Doktor-Ingenieur (Dr.-Ing.)

genehmigte Dissertation (Monografie)

von
André Moser, M.Sc.

aus
Essen

2025

Gutachter:

1. Gutachter: Prof. Dr.-Ing. Ralf Pörtner
2. Gutachter: Prof. Dr. Andreas Liese
3. Gutachter: Prof. Dr.-Ing. Volker C. Hass

Tag der mündlichen Prüfung:

27. März 2025

DOI: https://doi.org/10.15480/882.15766

Abstract

Heterogeneous conditions caused by non-ideal mixing could cause worse process performances in large-scale reactors compared to lab-scale processes. To date, the industrial scale-up process often relies on empirical rules and costly experiments. Mathematical models and physical scale-down systems are rarely used for industrial scale-up, especially for bioeconomic processes, despite their potential to address scale-up challenges.

This thesis employs a physical scale-down system with two interconnected stirred tank reactors, and a novel mathematical model using a network of zones approach for the simulation of non-ideally mixed conditions. Model-based techniques are utilised to design the operational strategies in processes under controlled non-ideal conditions.

Two different research questions were formulated and answered in this thesis:

1. Is a structured, mechanistic mathematical model, calculating several interconnected ideally mixed zones, able to approximate the flow patterns in stirred tank reactors and describe the effects caused by non-ideal mixing in biotechnological processes with complex kinetics?
2. How can feeding strategies for biotechnological processes in non-ideally mixed reactors be systematically designed?

To answer the first question, a comprehensive experimental study was conducted with 19 cultivations in an ideally mixed lab-scale reactor and the non-ideally mixed scale-down system. After parameterising the mathematical model with data from 16 experiments, the model described the experimental data with high accuracy and was successfully validated with data of three experiments.

The model was also used in a modified model-based design of experiments to maximise biomass density under controlled heterogeneous conditions, predicting an experiment with high accuracy (R^2 = 0.94). In a second study, process differences between laboratory and pilot scale were investigated. Non-linear model-based predictive control was employed for the first time to design a process in the scale-down system, aiming to reduce differences to pilot-scale data. The mathematical model was then used to find potential explanations for these differences, identifying the different zones (and their volumes) in the large-scale system as the probable reason for the performance differences between scales.

In the future, this combination of physical and mathematical modelling techniques with model-based control methods may accelerate process development and scale-up, while increasing efficiency and reliability.

Kurzfassung

Heterogene Bedingungen, die durch nicht ideale Durchmischung verursacht werden, können dazu führen, dass die Prozessperformance in großtechnischen Reaktoren schlechter ist als im Labormaßstab. Zudem basiert der industrielle Scale-Up häufig auf empirischen Regeln und kostspieligen Experimenten. Mathematische Modelle und physikalische Scale-Down Systeme werden nur selten für das industrielle Scale-up verwendet, insbesondere für bioökonomische Prozesse, obwohl sie das Potenzial besitzen, die Problematik des Scale-up zu bewältigen.

In dieser Thesis wird ein physikalisches Scale-Down System eingesetzt, welches aus zwei miteinander verbundenen Rührkessel besteht und mit einem neuartigen mathematischen Modell kombiniert wird, welches ein „Network of Zones Modell“ verwendet, um nicht ideale Mischbedingungen zu simulieren. Modellgestützte Techniken werden eingesetzt, um Betriebsstrategien unter kontrollierten, nicht-ideal durchmischten Bedingungen zu designen.

In dieser Arbeit wurden zwei verschiedene Forschungsfragen formuliert und beantwortet:

1. Ist ein strukturiertes, mechanistisches mathematisches Modell mit mehreren verbundenen ideal durchmischten Zonen in der Lage, Strömungsmuster in Rührreaktoren zu approximieren und die Auswirkungen nicht idealer Durchmischung in biotechnischen Prozessen mit komplexer Kinetik zu beschreiben?
2. Wie können Fütterungsstrategien für biotechnologische Prozesse in nicht ideal durchmischten Reaktoren systematisch gestaltet werden?

Zur Beantwortung der ersten Frage wurde eine umfassende experimentelle Studie mit 19 Kultivierungen in ideal gemischten Reaktoren und dem nicht ideal gemischten Scale-down-System durchgeführt. Nach der Parametrierung des mathematischen Modells mit Daten von 16 Experimenten beschreibt das Modell die experimentellen Daten mit hoher Genauigkeit und wurde zudem erfolgreich mit Daten aus drei Experimenten validiert.

Das Modell wurde auch in einer modifizierten modellbasierten Versuchsplanung eines Experiments mit dem Ziel der Maximierung der Biomassedichte im Scale-Down System eingesetzt werden und konnte dies mit hoher Genauigkeit (R^2 = 0,94) voraussagen. In einer zweiten Studie wurden Prozessunterschiede zwischen Labor- und Pilotmaßstab untersucht. Zum ersten Mal wurde adaptive, modellbasierte Prozesssteuerung genutzt, um einen Prozess in einem Scale-Down System zu entwickeln, der die Unterschiede zwischen dem Scale-Down System und dem Pilotmaßstab reduzieren sollte. Das mathematische Modell wurde anschließend verwendet, um potenzielle Erklärungen für Unterschiede zu finden, wobei verschiedene Zonen (und deren Volumina) im Produktionsmaßstab als mögliche Ursache für unterschiedliche Performance identifiziert wurden.

In Zukunft könnte diese Kombination physikalischer und mathematischer Modellierungstechniken mit modellbasierten Strategien die Prozessentwicklung und den Scale-up beschleunigen und gleichzeitig deren Effizienz und Zuverlässigkeit erhöhen.

Table of contents

List of Figures

List of Tables

Nomenclature

Abbreviations

Ae	Aerobic
AI	Artificial Intelligence
AMPC	Adaptive Model Predictive Control
An	Anaerobic
ANN	Artificial Neural Network
Approx.	Approximately
ATP	Adenosine triphosphate
CCD	Central Composite Design
CFD	Computational Fluid Dynamics
Ch.	Chapter
CHO	Chinese Hamster Ovary
CMA	Compartment(alised) model(ling) approach
CoD	Coefficient of Determination (R^2)
DCU	Digital Control Unit
dMSCC	dynamic microfluid reactor for single cell cultivation
DCW	Dry Cell Weight
DNA	Deoxyribonucleic acid
DO	Dissolved Oxygen (also pO_2)
DoE	Design of Experiments
DSig	Double sigmoid function
DT	Digital Twin
e.g.	example given (*exempli gratia*)
EU	European Union
HPLC	High Performance Liquid Chromatography
IDE	Integrated Development Environment
Lab-scale	Laboratory-scale
LHS	Latin-Hypercube Sampling
mDoE	Model-assisted Design of Experiments
MFA	Metabolic Flux Analysis

MPC	Model Predictive Control
NMPC	Non-linear Model Predictive Control
NoZ	Network of Zones
ODE	Ordinary Differential Equation
OFAT	One factor at a time
OLFO	Open-Loop-Feedback-Optimal
OS	Operating System
PAT	Process Analytical Technology
PC	Personal computer
PFR	Plug Flow Reactor
PI	Proportional Integral
PID	Proportional Integral Derivative
RNA	Ribonucleic Acid
RQ	Respiratory Quotient
RSM	Response Surface Method
SDM	Scale-Down Model
SI	International System of Units (*Système international d'unités*)
STR	Stirred Tank Reactor
Suc	Sucrose
US	United States
WMSD	Weighted Mean Square Deviation

Capital Letter

Symbol	Unit	Meaning
CO_2	-	Carbon dioxide
D	m	Diameter
D	-	Total desirability (mDoE)
F	$m^3\ s^{-1}$	Volumetric flow rate
Fr	-	Froude number
H	-	Henry constant
K_i	$kg\ m^{-3}$	Saturation constant
K_{Sl}	-	Slope constant
$L(x)$	Unit of x	Lowest value (mDoE)
M	$g\ mol^{-1}$	Molar mass
MW	$g\ mol^{-1}$	Molar weight
N	s^{-1}	Stirrer speed
Ne	-	Newton number
O_2	-	Oxygen
P	-	Power number
P	-	Product
Q	-	Aeration number
R^2	-	Coefficient of determination R^2
S	-	Substrate
T	K	Temperature
$U(x)$	Unit of x	Highest value (mDoE)
V	m^3	Volume
Y	$kg\ m^{-3}$	Yield coefficient

Lower case letters

Symbol	Unit	Meaning
$a\ to\ j$	-	Parameter of quadratic function
c	g L^{-1}	Concentration / Partial density
d	m	Stirrer diameter
d	-	Part desirabilities (mDoE)
f	-	Function
g	m s^{-2}	Acceleration due to gravity
$k_L a$	s^{-1}	Volumetric oxygen transmission rate
$k_{weighting}$	-	Weighting constant
m	kg	Mass
n	-	Number
q_G	$m^3\ s^{-1}$	Gaseous volume flow
r	s^{-1}	Specific reaction rate
t	s	Time
$t_{M,90}$	s	Mixing time (90% mixing)
$t_{M,95}$	s	Mixing time (95% mixing)
$v(x)$	Unit of x	Variability (mDoE)
$\bar{x}$	Unit of x	Mean result (mDoE)
x, X	-	Value at x
x, y, z	-	Coordinates in 3D-room
y, Y	-	Value at x or t
y	mol mol^{-1}	Molar fraction
w_r	-	Weighting of result (mDoE)
z	-	Number of stirrers

Indices

Index	Meaning
DCW	Dry Cell Weight
DSig	Double sigmoid
EtOH	Ethanol
G	Gaseous phase
Glc	D-Glucose
h	high
i	Component i, compartment i
j	Component j, compartment j
L	Liquid phase
l	low
min	Minimum
max	Maximum
N	Nitrogen-Source
P	Product
S	Substrate
Suc	Sucrose
X	Biomass

Greek letters

Symbol	Unit	Meaning
$\bar{\varepsilon}_T$	$m^3\ s^{-1}$	Average energy dissipation
η	$kg\ m^{-1}\ s^{-1}$	Dynamic viscosity
ρ	$kg\ m^{-3}$	Density

1 Introduction

The process performance on the production scale is often different or even worse compared to the laboratory scale (George et al. 1998; Takors 2012). Nonetheless, the scale-up procedure of industrial biotechnological processes is often based on rather empirical approaches (e.g. constant volumetric oxygen mass transfer coefficient (k_La) or volumetric power input) (Yang 2007). According to Formenti (2014), industrial scale-up is rarely but increasingly based on mathematical models which consider the interaction between process scale and metabolism of cells (Formenti et al. 2014). In academic research, mathematical models are employed that integrate biokinetic (mechanistic and/or data-driven models) with hydrodynamic modelling (e.g. computational fluid dynamics). This approach enables the description of large-scale process behaviour using data obtained from laboratory experiments (Du et al. 2022; Monsalve-Bravo et al. 2015; Wang et al. 2020; Cappello et al. 2021). Nonetheless, industrial scale-up is still expensive, of high-risk, and usually more expensive than the development of the lab-scale process (Junker 2004; Crater and Lievense 2018). Experiments (e.g., optimisation experiments) in production scale reactors are economically not feasible and are performed on the pilot and laboratory scale e.g. in physical scale-down models. (Formenti et al. 2014; Neubauer and Junne 2016)

The performance loss is most likely attributed to non-ideal mixing conditions which cause heterogeneities and gradients (heterogenous conditions) within large scale bioreactors. The mixing times in production scale bioreactors can be longer than a minute, whereas small laboratory reactors are considered to be ideally mixed with mixing times of only a few seconds (< 5 s). For yeasts, mostly oxygen gradients and/or substrate concentration gradients might influence process performance (George et al. 1998; Larsson et al. 1996). Even for well-known microorganisms such as *Saccharomyces cerevisiae*, lower biomass densities are achieved on the production scale as compared to laboratory scale even if identical strains of *S. cerevisiae* and identical substrates are used (George et al. 1998; Bohrer 2020).

Physical laboratory and micro-scale reactors are used for the improvement of process understanding and the parameter identification of mathematical process models. Another special case of laboratory reactors are physical scale-down models (SDMs). SDMs are utilised to replicate large scale, heterogeneous conditions and often are parallelised millilitre-scale reactors or consist of two (or more) connected reactors, which are either combinations of stirred tank reactors (STR) and plug flow reactors (PFR) or two connected STRs. The conditions are controlled to resemble certain effects occurring in large-scale reactors such as the size ratio of zones, dissolved oxygen concentrations or

mixing time. (Oosterhuis and Kossen 1984; Neubauer and Junne 2016; Nickel et al. 2017; Heins and Weuster-Botz 2018; Delvigne et al. 2017)

Besides physical models, mathematical models are simulating bioprocesses and are able to accelerate research and process development by reducing costly and time-consuming experiments. Especially, the combination of biokinetics and hydrodynamics has potential for the application during scale-up as well as for the optimisation of large-scale production processes (Delvigne et al. 2017). The biokinetics form the core of the mathematical model and are usually described by systems of ordinary differential equations with a mechanistic background or by empirical equations and relationships. The hydrodynamics can be calculated using computational fluid dynamics (CFD) or may be accounted for by compartment modelling approaches (CMA) (e.g. network of zones) (Delafosse et al. 2014). CMA are considering networks of ideally mixed zones. The properties of these zones and the exchange flows between zones can be based on experimental observations or CFD simulations. CFD simulations are excellent for reactor design and process understanding, but their main drawbacks are their complexity, and the computing power required. So far, mostly simple biokinetic models have been coupled with hydrodynamic modelling. (Formenti et al. 2014; Delvigne et al. 2017; Fitschen et al. 2023)

The process control strategy must be modified during the scale transfer due to influences of the heterogeneous conditions on the process performance. Therefore, the mathematical model does not only need to be valid for ideally mixed and non-ideally mixed systems but as well for different process control strategies, such as different batch or fed-batch feeding strategies.

The long computation times make an application of biokinetics coupled with CFD and detailed CMA models difficult or impossible within algorithms for process control and optimisation. This so far untapped direct application potential could be realised through a CMA model with a limited number of zones. The application of a CMA model might enable parameter identification with data from ideally mixed laboratory processes and from scale-down models with controlled heterogeneous conditions. For this reason, a new validated mathematical model was developed which couples a structured biokinetic submodel with a network of zones model.

Model-based process design as well as the optimisation of the process performance require suitable methods (e.g. model-based design of experiments) and are required to be performed with experiments on the laboratory scale. The design and optimisation of the process is scale-dependent and must lead to a robust process in which small changes to the control do not cause large changes to the process performance (Formenti et al. 2014). In order to achieve comparable process behaviour and performance in different scales,

process operational or control strategies must therefore be adapted to each scale. The adaption of these strategies usually necessitates considerable experimental effort, which might be reduced through the implementation of adaptive feeding strategies. For this reason, the potential heterogeneous mixing conditions must be investigated and described with a mathematical model in order to be able to design and optimise the process.

Possible model-assisted design methods include the calculation and optimisation of fixed-feeding profiles or adaptive feeding profiles. Fixed feeding profiles are calculated before the start of the process and include methods like model-based design of experiments (Möller et al. 2015; Moser et al. 2021). In the case of adaptive feeding profiles, new process data is acquired and used to identify the model parameters. The adjusted model is then used for the calculation of adaptive feeding profiles with strategies like adaptive non-linear model predictive control (Witte 1996; Santos et al. 2010; Hass et al. 2013).

Cultivations in physical bioreactor systems including non-ideal mixing conditions are investigated in this work with the assistance of a mathematical process model and the utilisation of model-based design methods. This combination has the potential to lead to innovative applications for scale-up, scale-down and even optimisation of industrial processes based on lab-scale experiments and mathematical modelling.

2 State of the art

The process performance of industrial processes on the production scale, is often different compared to the laboratory scale (George et al. 1998; Takors 2012). This difference is probably caused by heterogenous conditions, which mainly occur in large-scale production systems for example in stirred tank reactors. In this work heterogenous substrate concentrations (incl. DO) are defined as non-uniform, locally different substrate concentrations in non-ideally mixed bioreactors or bioreactor systems. In non-ideally mixed bioreactors, the reactor medium exhibits heterogeneous concentrations which might impact the process performance (Enfors et al. 2001; Hewitt and Nienow 2010).

This is even the case for well-known and industrially relevant processes such as the production *Saccharomyces cerevisiae,* e.g. for the production of commercial and organic yeast (as defined in the EU regulation on organic production), including the growth on sustainably produced substrates (European Parliament 2018). These industrial *S. cerevisiae* processes are still producing less biomass per reactor volume on the production scale compared to the laboratory scale which was reassured by an industrial partner.

For example, George et al. (1998) reported that a *S. cerevisiae* cultivation in a large scale reactor with a volume of 120 m³ had a 7% lower biomass yield compared to a 10 L laboratory scale reactor, which was mainly attributed to a substrate distribution in the production scale (George et al. 1998).

Non-ideally mixed conditions might also cause a limitation of oxygen availability (Garcia-Ochoa and Gomez 2009), population heterogeneities (Heins and Weuster-Botz 2018; Lemoine et al. 2017) or the accumulation of byproducts (Junne et al. 2012; Bylund et al. 1998; Enfors et al. 2001) such as ethanol for yeast cultivations (Sweere and Mesters et al. 1988). These consequences of heterogeneities often negatively affect the process performance (George et al. 1998).

Therefore, an effective scale-up and scale-down procedure is necessary, ideally with similar performances on the large-scale production systems and laboratory scale. Usually heuristic approaches are selected and applied, utilising physical scale-up criteria such as geometric similarity (Najafpour 2007), volumetric oxygen mass transfer (k_La) (Yawalkar et al. 2002) or volumetric power input (Ju and Chase 1992).

Chapter 2.1 introduces the properties, cultivations, and applications of the selected organism *S. cerevisiae*. Different bioreactor scales are briefly introduced in chapter 2.2. In chapter 2.3, relevant scale-dependent effects and gradients are described. In chapter 2.4, several types of scale-down models are introduced. This is followed by chapter 2.5 which summarises mathematical model types and modelling approaches for cultivations and

applications under ideally mixed and non-ideally mixed conditions. The last chapter of the state of the art presents different model based control and optimisation methods (chapter 2.6).

2.1 *Saccharomyces cerevisiae* cultivations

Saccharomyces cerevisiae (or baker's yeast) is an industrially relevant microorganism with an estimated worldwide production of over 3 million tons (Lesaffre 2020). The industrial production of *S. cerevisiae* is performed mostly with fed-batch processes in large scale bioreactors ranging from tens up to over 100 m³. At that scale, non-ideally mixed conditions are expected and the resulting gradients might have a negative impact on the process performance (e.g. resulting biomass density) (George et al. 1998; Bohrer 2020).

The main applications of *S. cerevisiae* are the production of alcoholic beverages like wine and beer and the usage in the baking industry. Other applications are the production of bioethanol as fuel, as well as biobased alcohols for biosynthesis processes and drug production (e.g. recombinant proteins) (Parapouli et al. 2020; Nandy and Srivastava 2018). Furthermore, Hirschmann (2021) optimised a whole cell biocatalysis process with *S. cerevisiae* for the production of ethyl(3)hydroxybutyrate (E3HB) (Hirschmann 2021).

Yeasts are eukaryotic single cell organisms. *S. cerevisiae* is a well-known yeast and one of the most researched microorganisms. As an example, the chromosome III of *S. cerevisiae* became the first completely sequenced chromosome of any organism in 1992 (Oliver et al. 1992), and the genome was completely sequenced in 1996 (Goffeau et al. 1996).

The growth of *S. cerevisiae* (and other yeasts) can be divided into five distinct growth phases, which were described by Kolehmainen et al. (Kolehmainen et al. 2003):

1. Phase: Lag phase (acclimation to new environmental conditions)
2. Phase: Exponential growth (decreasing glucose and oxygen concentration while ethanol is produced)
3. Phase: Exponential growth (glucose present, with oxygen limitation, while ethanol is produced)
4. Phase: Respiratory growth using both glucose and ethanol carbon substrates.
5. Phase: Respiratory growth solely on ethanol

During the 4th and 5th phase growth is decelerated and eventually a stationary phase and afterwards a death phase without growth are reached (Herskowitz 1988).

S. cerevisiae is a facultative anaerobic microorganism and is able to produce ethanol under both aerobic and anaerobic conditions (van Dijken et al. 1993). Under anaerobic conditions, ethanol is produced through alcoholic fermentation and under aerobic conditions through an overflow metabolism when the glucose concentration is above 3.6 mg L^{-1} to 210 mg L^{-1}

(0.02 to 0.85 mM) as shown by multiple research groups (Woehrer and Roehr 1981; Rose 1993). The glucose uptake rate under anaerobic conditions might be higher compared to aerobic conditions (Hollander et al. 1986). However, more glucose is needed under anaerobic conditions for the same generation of energy since only 2 ATP are produced per glucose molecule (instead of 36 ATP for the full aerobic respiration of glucose). During fermentative growth on glucose, pyruvate is formed through glycolysis which is further fermented to ethanol and CO_2 (Hollander et al. 1986; Rodrigues et al. 2006).

S. cerevisiae also produces ethanol under aerobic conditions through an overflow metabolism of glucose which is commonly known as the Crabtree effect (Deken 1966; Crabtree 1929). At first, the Crabtree pathway may appear inefficient, as it produces only 2 ATP (equal to fermentation under anaerobic conditions) compared to the direct oxidation of glucose (Thomson et al. 2005). However, many yeasts (including *S. cerevisiae*) have an increased tolerance to ethanol of up to 60 g L^{-1}, in contrast to other microorganisms (Casey and Ingledew 1986; Thomson et al. 2005; Fleet 1993). Additionally, *S. cerevisiae* can metabolise ethanol for the supply of energy and for biomass growth.

Historically, this production of ethanol has been used by humankind for consumption and preservation and is still one of the main industrial application cases of *S. cerevisiae*. However, in controlled bioprocesses with the aim of maximising the biomass density, the Crabtree effect should be minimised in order to optimise biomass growth (Sonnleitner and Käppeli 1986; Fiechter et al. 1981; Deken 1966; Lemoigne et al. 1954; Crabtree 1929; Thomson et al. 2005).

The overall metabolic pathway of glucose to ethanol and the pathway of the total oxidisation of glucose are:

$$C_6H_{12}O_6 \rightarrow 2 \cdot C_2H_6O + 2 \cdot CO_2 \quad (2.1)$$

$$C_6H_{12}O_6 + 6 \cdot O_2 \rightarrow 6 \cdot CO_2 + 6 \cdot H_2O \quad (2.2)$$

If glucose is at low levels under aerobic conditions, *S. cerevisiae* is able to consume ethanol for growth and energy production. The oxidisation of ethanol has the following stoichiometric equation:

$$C_2H_6O + 3 \cdot O_2 \rightarrow 3 \cdot H_2O + 2 \cdot CO_2 \quad (2.3)$$

This two-phase growth on two different substrates is known as diauxic growth (Battley 1960; Lemoigne et al. 1954).

S. cerevisiae was chosen for the investigations of the influence of heterogeneous concentrations (non-ideally mixed conditions) on the process performance because even well-known *S. cerevisiae* cultivation processes often yield lower biomass densities on the industrial scale compared to processes on the laboratory scale (George et al. 1998; Bohrer 2020). This reduction of the process performance might be caused by gradients

such as dissolved oxygen and substrate concentration gradients within industrial scale reactors. One viable way to investigate these heterogeneous conditions is by monitoring the ethanol concentration, since ethanol is only produced at elevated glucose concentrations or under oxygen limitation. Furthermore, ethanol is consumed at low glucose levels under aerobic conditions (Sweere and Matla et al. 1988; Sweere and Mesters et al. 1988). In addition, the cultivation of yeast is relatively easy, robust, and short with cultivation times of 1 to 2 days. All these features are important for acquiring reliable data for the development and adaption of mathematical process models (Karathia et al. 2011).

2.2 Bioreactor scales

There are several scales of bioprocesses; the boundaries and transitions between the scales are not strictly defined and vary from author to author (Venus and Richter 2007; Crater and Lievense 2018).

The smallest scale are microfluidic devices with volumes close to the volume of single cells. This allows for a direct observation of single cell dynamics under accurately adjustable environmental conditions. Since single cells are used, the cellular state can directly be observed, which is not possible with other bioreactor systems (Dusny and Schmid 2015; Täuber et al. 2022).

Microreactors or millilitre-scale bioreactors range from about 50 µL to 10 mL (e.g. microtiter plates). They are often run in parallel and are used for high-throughput screening to test microorganisms, improve process efficiency during process development or to determine model parameters (Fink et al. 2021; Tajsoleiman et al. 2019). They also might be used to provide the inoculum for the next scale (laboratory scale) within seed-trains (Bareither and Pollard 2011; Formenti et al. 2014; Weuster-Botz 2005).

Laboratory scale (lab-scale) reactors are often shake flasks with maximum volumes of about 1 L or stirred tank reactors with maximum volumes of about 50 L. One important field of applications for lab-scale reactors are assessment of the process performance and the optimisation of the process control (Dusny and Schmid 2015).

Above this range the pilot scale starts. Pilot reactors are defined with varying maximum volumes of up to 10 m³ (Crater and Lievense 2018). They are used for process scale-up and as a pre-stage of industrial production reactors.

The production scale starts above 1 m³ depending on the process. Industrial production reactors can exceed volumes of above 200 m³ for selected biotechnological processes like the production of yeasts (Crater and Lievense 2018; George et al. 1998).

2.3 Scale-dependent effects

With increasing scale of bioreactors, heterogeneities and gradients are probable. These are caused by non-ideal mixing, since ideally mixed conditions would require unfeasible power input and stirrer speeds. The gradients might be substrate gradients (chapter 2.3.1), gradients of dissolved gas concentrations (chapter 2.3.1), and other gradients (e.g. gradients of pH value or shear forces) (Lemoine et al. 2017; Bylund et al. 1998; Neubauer and Junne 2016) (see Figure 1).

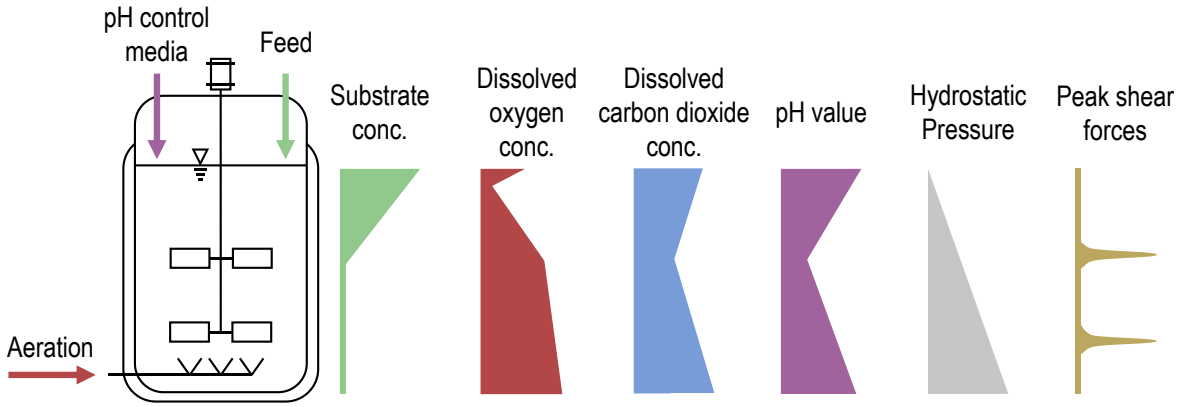


Figure 1: Selection of hypothetical axial gradients and heterogeneities and their exemplary & simplified distributions that might be present in large-scale stirred tank bioreactors for yeast cultivations. Substrate concentration: Highest at inlet, decreases gradually until nearly depletion. Dissolved oxygen concentration: Highest at aeration inlet, raised by hydrostatic pressure at bottom and lowered by high metabolic reaction rates at the top. Dissolved carbon dioxide concentration: Raised by hydrostatic pressure and metabolic activity. pH value: Influenced by controlling agents, mixing times and dissolved oxygen concentration. Hydrostatic pressure: Highest in the bottom area of the bioreactor, gradually decreases to the top. Shear forces: Force peaks in the immediate vicinity of the stirring levels (graphic based on Neubauer and Junne (2016) and further developed with data from Oosterhuis (1984)).

Heterogeneous conditions often occur near the inlet point of the feed medium, aeration gas and pH adjustment media (Neubauer and Junne 2016; Bylund et al. 1998). These heterogeneities are a major issue during scale-up and ultimately in large-scale production systems. They often cause a reduction of the process performance.

In the case of industrial cultivations of many microorganisms such as *S. cerevisiae* or *Escherichia coli*, heterogeneous substrate and oxygen distributions in industrial large-scale stirred tank reactors are often causing a reduction of the biomass or product yield (Larsson et al. 1996; George et al. 1993).

Heterogeneous distributions (gradients) of substrate concentrations might cause overflow metabolisms and/or inhibition in many different microorganisms. For example, *E. coli* produces acetate under fully aerobic conditions, generally at high glucose uptake rates and/or high growth rates (Akesson et al. 1999; Andersen and Meyenburg 1980). *S. cerevisiae* produces ethanol through overflow metabolism when the glucose concentration exceeds levels between 3.6 mg L^{-1} and 210 mg L^{-1} (Woehrer and Roehr 1981; Rose 1993). Both, acetate in *E. coli* cultivations and ethanol in *S. cerevisiae* cultivations are able to inhibit biomass growth and metabolic activity. Acetate inhibits growth and protein formation even at low concentrations of 0.5 g L^{-1} (Eiteman and Altman 2006; Nakano et al. 1997; Koh et al. 1992; Reiling et al. 1985). For *S. cerevisiae* cultivations growth inhibition might start already at ethanol concentrations above 20 g L^{-1} (Casey and Ingledew 1986; Bajaj and Sharma 2010).

Oxygen limitation affects microbial metabolism and causes anaerobic reactions and stress responses that might influence process performance. Some microorganisms are capable of surviving and growing under both aerobic and anaerobic conditions. *Lactococcus lactis* and *S. cerevisiae* are well-known facultative anaerobic microorganisms (Sijpesteijn 1970). *S. cerevisiae* consumes sugar and produces ethanol and carbon dioxide when oxygen is at low levels (Battley 1960). *L. lactis* generates lactic acid from sugar under anaerobic conditions. The respirative metabolism of *L. lactis* produces acetic acid and acetoin at the expense of lactic acid production (Cesselin et al. 2018; Kandler 1983). Both microorganisms are widely employed for production of fermented food and drinks as well as for food preservation (Neves et al. 2005).

These examples illustrate the relevance of experimental investigations of these gradients with regard of the influence on process performance (e.g. biomass growth or product formation) (Neubauer and Junne 2016; Bylund et al. 1998).

The gradients of shear forces also change during scale-up. The power input per volume during scale-up usually decreases and the diameter of the impellers increases, resulting in lower rotational speeds but increased tip speed of the impeller. This results in high peak shear forces but overall lower average shear forces in the large-scale (Humphrey 1998; Junker 2004). Simulations of radial and axial shear forces around the stirrer were performed for example for Rushton turbines (Torrez and André 1999) and pitched blade turbine (Márquez-Baños et al. 2019).

Kar et al. (2008) investigated the influence of multiple gradients (pH and DO fluctuations, as well as methyl oleate dispersion) on the lipase production of *Yarrowia lipolytica*. pH gradients arise because pH-media is added at the top of the reactor. These pH gradients

(and methyl oleate gradients) led to a decline in cell growth and enzymatic activity (Kar et al. 2008).
Heterogeneous conditions in reactors regarding limiting substrates, pH or DO are expected to increase with scale. (George et al. 1998)

2.3.1 Heterogeneous substrate concentration

Heterogeneous conditions (gradients) of substrates are mainly caused by feeding. Bylund et al. (1998) determined up to 400 times higher glucose concentrations when the feed was located in a stagnant mixing zone around the feeding spot compared to the mean concentration for an *E. coli* cultivation. This substrate gradient resulted in an elevated acetate production due to increased overflow metabolism. The increased acetate production led to a biomass reduction of up to 20% in a 12 m³ (initial volume of 8 m³) large-scale cultivation compared to a 3 L (initial volume of 2.5 L) lab-scale process. (Bylund et al. 1998)
Larsson et al. (1996) determined a twice as high glucose concentration (40 mg L^{-1}) at the top of a 30 m³ bioreactor (liquid volume: 19.8 – 22.3 m³) compared to the bottom sampling level (20 mg L^{-1}) for a *S. cerevisiae* cultivation. These substrate gradients were more pronounced when the feed was located in a stagnant zone at the top, compared to gradients of a process where the feed was located in a well-mixed zone around the bottom impeller (Larsson et al. 1996).
George et al. observed glucose and fructose gradients in an industrial 215 m³ production-scale bubble column reactor with 120 m³ medium volume for a fed-batch production process of *S. cerevisiae*. Sampling was performed at two levels in the production-scale reactor. At the bottom sampling level (at 2.7 m) a fructose concentration of up to 150 mg L^{-1} was determined, whereas at the same time at the top sampling (at 6.3 m) level a lower concentration of 110 mg L^{-1} was determined. When compared to a homogeneous 10 L reactor using identical strain, medium and process control, the biomass yield in the production scale was 6 - 7% lower whereas the maximum ethanol concentration during the process was up to 80% higher.

2.3.2 Heterogeneous dissolved oxygen concentration

Oxygen gradients may arise in large-scale bioreactors as a consequence of interactions between high mixing times, oxygen transfer and kinetics. Oxygen limitation resulting from gradients is predominantly occurring when the time constant for the oxygen uptake is smaller than the circulation time of the fluid.
The circulation time is usually determined utilising regime analysis (Amanullah et al. 2004; Sweere and Mesters et al. 1988; Sweere et al. 1987; Oosterhuis 1984; Palomares et al.

2010). As a general guide, the circulation time t_c can be determined as one-quarter of the mixing time t_M until the medium is perfectly mixed and no gradients are within the reactor medium (Sweere et al. 1987):

$$t_c = t_M/4 \tag{2.4}$$

In a study by Oosterhuis and Kossen, dissolved oxygen concentrations (DO or pO_2) were measured at multiple vertical and radial locations in a 25 m³ bioreactor equipped with two Rushton turbines. The research revealed that DO ranged from 13 - 22% in direct proximity of the impeller down to 0% saturation at the reactor wall. Surface aeration at the liquid-gas phase boundary increased the DO again to values up to 8% (Oosterhuis and Kossen 1984; Oosterhuis 1984). As a result, cells in bioreactors circulate through and are exposed to regions with different dissolved oxygen concentrations.

Sweere et al. simulated oxygen gradients within a 2 L bioreactor by periodically oscillating the oxygen concentration for a *S. cerevisiae* cultivation. These oscillations caused a reduction of biomass growth as well as an increased production of metabolites (acetic acid & glycerol) and ethanol. (Sweere and Mesters et al. 1988)

Heterogenous dissolved oxygen concentrations were also physically measured in cultivations with *Streptomyces niveus* and *Streptomyces aureofaciens* in bioreactors with different sizes, including 112 m³ (Manfredini et al. 1983) as well as 20 L, 250 L and 15 m³ bioreactors (Steel and Maxon 1966). However, the axial oxygen gradients observed during these fermentations were amplified because the resulting suspensions were highly non-Newtonian Bingham fluids. The radial gradients in these cultivations were negligible. (Steel and Maxon 1966; Manfredini et al. 1983)

2.3.3 Population heterogeneity

The concept of population heterogeneity is only briefly introduced in this work, as the simulation times of models considering population heterogeneity are usually too time-consuming for model based process design methods. Segregated, population models were therefore not considered in the selection of the mathematical model, and instead an unsegregated modelling approach was chosen. Instead, only extrinsic factors such as substrate concentration distributions are considered in the model (chapters 2.5.3 and 4.5).

Population heterogeneity is defined as the unequal behaviour of individuals of an otherwise isogenic population and is present in all biotechnological processes, even in ideally mixed, homogeneous environments (Schlüter et al. 2015). Nonetheless, heterogeneous conditions and the resulting fluctuations experienced by cells can favour the formation of cell populations (Heins and Weuster-Botz 2018). These fluctuations might be caused by long mixing times above 100 s. These mixing times are multitudes longer than metabolic

reactions or gene expression caused by stress reactions generated by the environmental fluctuations (Wang et al. 2015; Lara and Galindo et al. 2006). The heterogeneous populations might express different yields, productivity, biomass growth or an increased by-product formation (Lidstrom and Konopka 2010).

The description of population heterogeneity is a complex process, requiring the comparison of multiple individual cells during a cultivation over an extended time period. The heterogeneity of a population is influenced by a number of different factors, including the growth phases, the age of individual cells, as well as variable external conditions (Martins and Locke 2015; Heins and Weuster-Botz 2018). This may result in gaussian or even complex multimodal distributions of the cell population with regard to physiological parameters like growth, metabolic activity or respiration (Lemoine et al. 2017).

In *S. cerevisiae* cultivations the formation of two distinct cellular populations was observed during ethanol production of 140 g L^{-1} in two days under fully aerobic conditions. The populations were isolated by centrifugation. The lower fraction was predominantly composed of uniform yeasts exhibiting a high respiratory activity whereas the upper fraction was heterogeneous in size with a lower respiratory but relatively high glycolytic activity (Benbadis et al. 2009).

The consequences of population heterogeneities of *S. cerevisiae* may include an increased ethanol production and reduced biomass growth (Kortmann et al. 2009). These effects could also be direct consequences of non-ideally mixed reactor media. The quantitative description of population heterogeneity is complex, and the resulting mathematical models usually have long computation times. Therefore, the use of a structured model that does not consider population heterogeneity as the initial modelling approach is more suitable for the description and model-based control of *S. cerevisiae* processes under non-ideal mixing conditions.

2.4 Physical scale-down models

The chemical and physical gradients in large-scale stirred bioreactors may be replicated on laboratory scale using physical lab-scale reactors and configurations of multiple connected reactors also called (physical) scale-down models (SDMs) (Lara et al. 2016).

The idea of considering non-ideal mixed conditions with a surrogate reactor model has already been proposed in the 1970s by Levenspiel (1972) for chemical reactors. The reactor model considers several compartments in order to represent the real non-ideal behaviour. For example, a physical mixed stirred tank reactor with a stagnant (dead) zone is mathematically represented by an ideally mixed compartment, which is connected to a second compartment representing the dead zone. (Levenspiel 1972)

This idea was transferred by Oosterhuis (1984) to biotechnological processes in order to investigate the influence of gradients which result in environmental fluctuations with scale-down models (Oosterhuis 1984; Oosterhuis et al. 1985).

Scale-down models can be grouped into mathematical and physical models. The physical scale-down models are described in this chapter; the mathematical models, which include CFD simulations and compartmentalised process models, are described in chapter 2.5.4.

Physical scale-down models can be applied to various bioprocess development stages, from upstream cell line selection to growth medium optimisation (Olughu et al. 2019). Table **1** provides an overview of SDMs which are explained in the following sections. Single compartment systems are described in chapter 2.4.1, two-compartment SDMs in chapter 2.4.2, and chapter 2.4.3 describes a three-compartment system.

Table 1: Examples for physical scale-down models consisting of single compartments, two compartments and multiple compartments often consisting of stirred tank reactors (STR) in combination with plug flow reactors (PFR).

	Description	**Gradient**	**Organism**	**Reference**
Single compartment scale-down models (ch. 2.4.1)	Reactor with controlled oscillations	Glucose feeding in intervals	*E. coli*	(Lin and Neubauer 2000)
	Reactor with controlled oscillations	Oxygen concentrations in oscillations	*S. cerevisiae*	(Sweere and Mesters et al. 1988)
	Reactor with installed cylindrical discs	Substrate gradient via longer mixing times	*Coryne-bacterium glutamicum*	(Schilling et al. 1999)
	Reactor with oscillating control and cell retention	Oscillating pH conditions	*C. glutamicum*	(Täuber et al. 2022)
	Loop reactor with forced circulation	Substrate gradients	*Aspergillus niger*	(Papagianni et al. 2003)
Two compartment scale-down models (ch. 2.4.2)	STR-PFR configuration	Glucose gradients	*S. cerevisiae*	(George et al. 1993)
	STR-PFR configuration	Seven different settings with pH, glucose, and dissolved oxygen gradients	*E. coli*	(Onyeaka et al. 2003)
	STR-PFR configuration	pH gradients	*Bacillus subtilis*	(Amanullah et al. 2001)
	STR-STR configuration	Oxygen gradients	*Gluconobacter oxydans*	(Oosterhuis et al. 1985)
	STR-STR configuration	Glucose gradients	*S. cerevisiae*	(Sweere and Matla et al. 1988)
	STR-STR configuration	Oxygen gradients	*E. coli*	(Sandoval-Basurto et al. 2005)
	STR-STR configuration	Oxygen gradients	*C. glutamicum*	(Limberg et al. 2017)
	STR-STR configuration	Oxygen gradients	*E. coli*	(Lara and Leal et al. 2006)
Multiple compartment systems (ch. 2.4.3)	PFR-PFR-STR configuration	Oxygen and substrate gradients	*C. glutamicum*	(Lemoine et al. 2015)

2.4.1 Single compartment physical scale-down models

Figure 2 depicts schematic graphics of representations of physical single compartment scale-down models (see also Table 1 for examples).

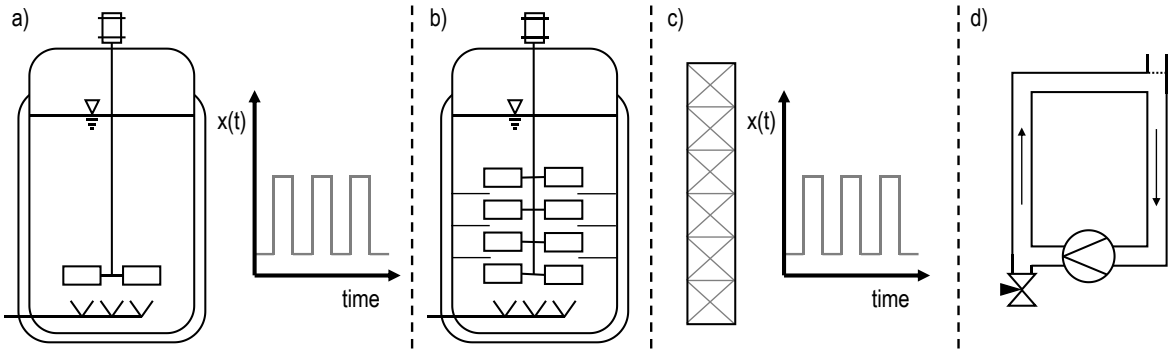


Figure 2: Examples of different scale-down models with a single reactor: a) Stirred tank reactor (STR) with oscillating control (e.g. feed) (Sweere and Matla et al. 1988); b) STR with installed cylindrical discs (Schilling et al. 1999); c) Plug-flow reactor with oscillating control and cell retention (Täuber et al. 2022); d) Loop reactor with forced circulation, valve for feeding and sampling (Papagianni et al. 2003).

One of the most common single reactor scale-down models is a stirred tank reactor (STR) with controlled oscillations, which might be oscillations of the substrate or oxygen concentration (Figure 2a) (Lin and Neubauer 2000; Sweere and Matla et al. 1988). Other examples of scale-down models with one compartment include single STRs with installed cylindrical discs to prolong mixing time (Figure 2b) (Schilling et al. 1999), plug flow reactors (PFR) with cell retention and oscillating control (Figure 2c) (Täuber et al. 2022) as well as loop reactors where the cultivation broth circulates (Figure 2d) (Papagianni et al. 2003).

Single reactor with controlled oscillations:

Lin and Neubauer (2000) compared the results of a recombinant *E. coli* culture in an STR with oscillations through glucose feeding in fast (1 min) and slow (4 min) intervals against continuous glucose feeding. The results indicated a reduction in product (α-glucosidase) stability. (Lin and Neubauer 2000)

Sweere et al. (1988) changed the oxygen concentration periodically for a *S. cerevisiae* cultivation by switching between air and nitrogen in the inlet gas flow. Oscillations of 1 min and 2 min led to a reduced biomass growth as well as an increased production of metabolites (acetic acid & glycerol) and ethanol. (Sweere and Mesters et al. 1988)

Reactor with installed cylindrical discs:

A 42 L STR with six stirrers and five cylindrical discs was designed by Schilling et al. (1999) for the scale-down of mixing time dependent cultivations. The installation of these discs resulted in an increase of the mixing time from 10 s to 130 s while the power input per liquid

volume remained constant. A fed-batch experiment with *C. glutamicum* was performed where the limiting substrate was fed onto the liquid surface. The 13 times longer mixing time, led to a reduced sugar and ammonium consumption rate, along with a decreased biomass growth and product formation ($_L$-lysine). (Schilling et al. 1999)

Reactor with oscillating control and cell retention:

Recently Täuber et al. (2022) published a special scale-down model and compared it to an STR-STR configuration (see section 2.4.2) for *C. glutamicum* cultivations. The scale-down model is a single compartment “dynamic Microfluid reactor for Single Cell Cultivation” (dMSCC). The reactor is consisting of three zones, one zone with oscillating pH conditions and two control zones, with continuously high and low pH values. Oscillations of pH conditions (8 min at pH 7 & 2 min at pH 6) resulted in a reduced growth rate of 21% and the STR-STR configuration of 27% compared to homogeneous conditions at pH 7. (Täuber et al. 2022)

Loop reactor with forced circulation:

Papagianni et al. (2003) designed a tubular loop reactor that employs a peristaltic pump for circulation of the reactor medium. The system was utilised for the aerobic cultivation of *Aspergillus niger*. The reactor produces comparable biomass and citric acid curves, when compared to a 10 L stirred tank reactor with a similar mixing time. The cultivation results indicate a reduced citric acid production with increasing circulation times. When the circulation time was increased from 11 s to 40 s, the citric acid production was halved. (Papagianni et al. 2003)

2.4.2 Two compartment physical scale-down models

Two compartment scale-down models were originally proposed by Oosterhuis (1984) to study large-scale production bioreactors on laboratory scale (Oosterhuis 1984). These compartments are representing different zones of the production scale reactor e.g. the well mixed zone around the stirrer and the more stagnant zone at the top of the large-scale reactor (Lara et al. 2016). Figure 3 displays two common scale-down models consisting of two compartments.

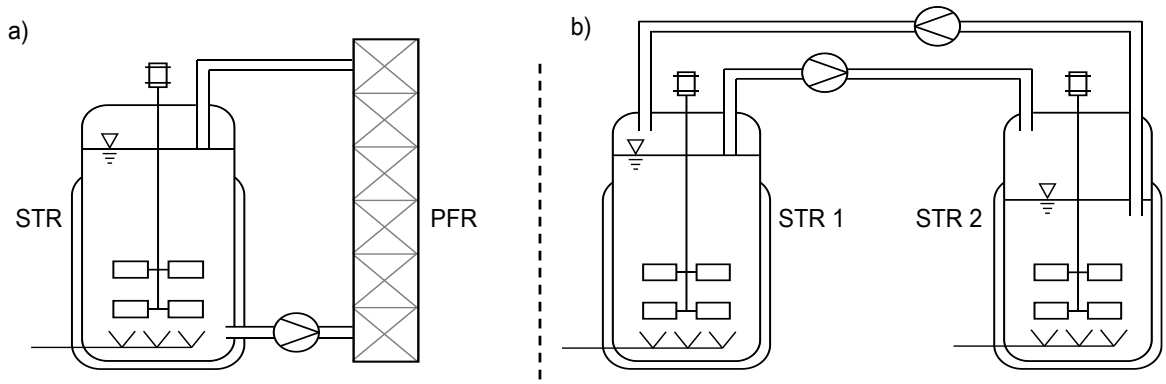


Figure 3: Examples of scale-down models with two compartments: a) STR with a PFR (George et al. 1998); b) two connected STRs (Oosterhuis 1984)

In both the STR-PFR and the STR-STR combination the cultivation broth is circulated between the two compartments. The circulation times and resulting mixing times are crucial settings of these systems. The plug flow reactor might be aerated and might include static mixers (Onyeaka et al. 2003). The inlet of the substrate feed could be positioned before the PFR or in the STR (George et al. 1998).

2.4.2.1 Scale-down model with a STR-PFR configuration:

The STR-PFR system is the most common scale-down model combining a plug flow reactor with time and space dependent concentration profiles with a stirred tank reactor. The PFR represents the zone where the feeding or air inlet is located in large-scale production system; the STR the well-mixed homogeneous zone around the stirrer (Papagianni 2015; Heins and Weuster-Botz 2018).

An STR-PFR configuration was utilised by George et al. (1993) to analyse the aerobic ethanol production by *S. cerevisiae*. The medium was circulating between the PFR and STR; the feeding was located at the inlet of the PFR. This process was compared to a cultivation in the STR under good mixing conditions. The heterogenous glucose distribution in the PFR caused a decrease in biomass growth of 6%, a reduction in the sugar uptake rate and an increase in ethanol production. (George et al. 1993)

Onyeaka et al. (2003) combined a well-mixed 5 L stirred tank reactor with a plug flow reactor for the cultivation of *E. coli*. The system was operated at seven different settings regarding pH, glucose concentration, and dissolved oxygen concentration. The obtained results were compared to those of a 5 L well-mixed stirred tank reactor and a 20 m³ large-scale reactor. The results indicate that the highest similarity occurred when the PFR experienced a high glucose concentration coupled with a low DO with a residence time in the PFR of 50 s and pH 7 in the whole scale-down system. (Onyeaka et al. 2003)

Amanullah et al. (2001) investigated the effects of pH gradients on a *B. subtilis* culture, which is sensitive to pH levels. The STR was kept at pH 6.5; with NaOH addition between the STR and PFR. While the maximum biomass was unaffected by the pH gradients, there were significant changes to other factors such as product formation or acetic acid concentration. (Amanullah et al. 2001)

2.4.2.2 Scale-down model with a STR-STR configuration

Oosterhuis et al. (1985) developed an STR-STR system with heterogenous dissolved oxygen concentrations and compared it to a single bioreactor system with fluctuating DO concentrations. The results highlight that not only the residence time is important, but as well the flow rate between the reactors and how much oxygen is transferred from the aerated to the unaerated reactor. Oosterhuis et al. also carried out optimisation experiments via gassing one reactor with pure oxygen, which resulted in irreversible damage to the cells. (Oosterhuis et al. 1985)

Sweere et al. (1988) employed an STR-STR system to investigate the impact of fluctuating glucose concentrations for different circulation rates and reactor volume ratios on cultivations of *S. cerevisiae*. The heterogeneous substrate concentrations in the system are caused by feeding into only one of the reactors. Experiments with a circulation time of 6 min resulted in a decreased biomass concentration of more than 20% and an increased ethanol production compared to a process in a single bioreactor. (Sweere and Matla et al. 1988)

Sandoval-Basurto et al. (2005) simulated the effects of oxygen gradients regarding *E. coli* cultivations by keeping one STR under aerobic and the other under anaerobic conditions for different circulation times. The growth rate of *E. coli* was reduced by 30% at a circulation time of 180 s. The negative effect on biomass yield and product (pre-proinsulin) yield was even higher. An accumulation of acetic acid, lactic acid, formic acid and succinic acid indicated that the anaerobic metabolism was responsible for the reduction of the growth rate and product yield. (Sandoval-Basurto et al. 2005)

Limberg et al. (2017) cultivated a strain of *C. glutamicum* in a STR-STR two-compartment scale-down model, with one STR featuring oxygen limitation. The residence time was set to approximately three minutes to simulate heterogenous dissolved oxygen concentrations on the industrial scale. No impact on biomass growth and target product (1,5-diaminopentane) formation was observed but the heterogeneous DO demonstrated a clear influence on the build-up of glycolytic intermediates and side products. Overall, the experimental setup provided insights into the cells' behaviour under a large scale environment. (Limberg et al. 2017)

Lara et al. (2006) cultivated recombinant *E. coli* in an STR-STR scale-down model under heterogeneous DO concentrations as expected in large-scale reactors. Molecular

phenomena (transcription) and macroscopic data e.g. regarding kinetic rates or stoichiometric coefficients are presented in this study. The results indicate that all products are produced during oscillating DO (dissolve oxygen tension) conditions of mixed-acid fermentation. Furthermore, the study reveals that certain genes are affected by oscillating DO environment, whereas other relevant genes remain unaffected. (Lara and Leal et al. 2006)

2.4.3 Three compartment physical scale-down models

Lemoine et al. (2015) developed a new three compartment scale-down (Figure 4) model consisting of two PFRs and one STR. This scale-down model was developed with the aim to replicate three distinct regions of a large-scale production reactor with a substrate feed situated at the bottom.

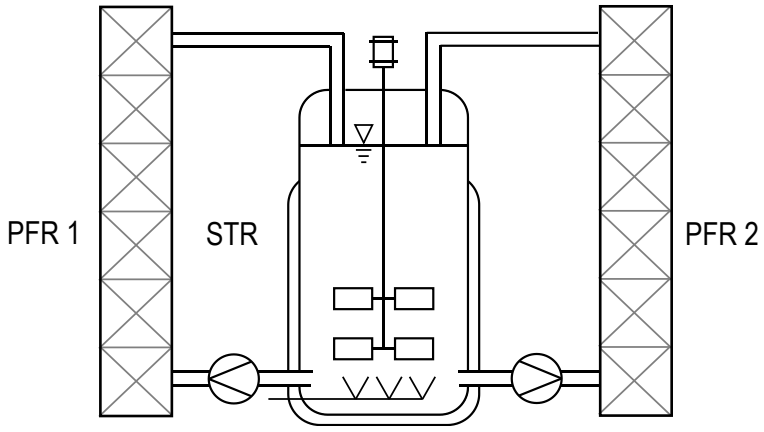


Figure 4: Scale-down model consisting of two PFRs and a STR (Lemoine et al. 2015)

This setup is designed to mimic three zones of industrial bioreactors, a relatively stagnant zone at the top, a stagnant zone at the bottom with high substrate concentration as well as the well-mixed zone around the stirrer. The representation of those three zones is not possible with two-compartment scale-down systems.
Lemoine et al. (2015) compared the results of a fed-batch cultivation of *C. glutamicum* with cultivations performed in an STR-STR scale-down model. The substrate was solely fed into one PFR, and both PFRs were not aerated. The process in the three compartment system produced more lactate and succinate as well as other metabolites. However, there was no significant impact on the overall bacterial growth. (Lemoine et al. 2015)

2.4.4 Comparison of physical scale-down systems

All presented bioreactor systems are able of simulating certain conditions of large reactors on a laboratory scale with different focuses and characteristics.

Single bioreactors (PFR and STR) with oscillating control and looped reactors offer the possibility of precisely timed and controlled environments. However, in reality, the residence times in the respective conditions are less precise and harmonised which can be achieved with a single bioreactor with prolonged mixing times or with combinations of two reactors. Single bioreactors might be installed with discs or other installations to prolong the mixing time (Schilling et al. 1999). However, they are difficult to describe with simple mathematical models and complex CFD simulations may be needed to obtain information about the fluid dynamics and conditions within this bioreactor type.

The advantage of the STR-STR configuration in comparison to the STR-PFR configuration is the possibility of featuring two well-defined and controlled environmental zones. The reactors in the STR-STR configuration are able to exhibit different liquid volumes. Other options are, that only one reactor might be aerated, or the inlet gas composition might be adapted to simulate oxygen gradients of large-scale systems. The microorganisms circulate through the systems throughout the cultivations experiencing heterogeneous conditions (Sandoval-Basurto et al. 2005; Lara et al. 2016; Limberg et al. 2017). In other cases the substrate is fed into only one compartment to simulate substrate gradients (Sweere and Matla et al. 1988). Main advantage of the STR-PFR system is the possibility of setting a defined residence time and defined substrate/oxygen gradients in the plug flow reactor which is not feasible in the STR-STR system (George et al. 1993).

The STR-STR system facilitates the experimental investigation of controlled heterogenous dissolved oxygen and substrate concentrations which could also simultaneously occur. These possible heterogeneities are likely to be critical for the performance of industrial *S. cerevisiae* processes. In addition, for a simulation of two connected bioreactors both individual bioreactors could be approximated as ideally mixed. The resulting model would be simpler and faster than potential models of the STR-PFR system and consequently of the three-compartment STR-PFR-PFR system. For these reasons, an STR-STR combination was chosen for the experimental investigations during this work (see chapter 4.3.2 for a detailed explanation and specifications).

One aim of this work is to answer the question how cultivations in this STR-STR system can contribute to finding explanations for the effects of heterogeneities on the production scale. This investigation will be supported by a mechanistic, mathematical model that simulates the cultivations in the STR-STR scale-down model (chapter 2.5). It should also be researched, if cultivations in this physical scale-down system can be designed using model-based bioprocess design strategies (chapter 2.6).

2.5 Mathematical process models in biotechnology

In this thesis, a novel mathematical process model shall be presented that can be employed to simulate biotechnological cultivation processes under non-ideally mixed conditions. The mathematical model must be able to describe all relevant determinable effects of *S. cerevisiae* processes under homogenous and heterogenous conditions with a satisfying precision. Ideally, it should also be usable for model-based control methods of processes performed under non-ideal mixing conditions.

The development and validation procedure of biotechnological process models is described in chapter 2.5.1. In most cases, process models can be categorised into a model structure and submodel framework, which is presented in chapter 2.5.2. Various different model types (and models) exists (chapter 2.5.3), which are capable of simulating bioprocesses and have different performance characteristics (chapter 2.5.3.4).

Computational-driven (in silico) design and experimentation are also crucial factors that leverage recent advantages in both scale-up and scale-down (Formenti et al. 2014; Delvigne and Noorman 2017). For this purpose, mathematical models are employed that are able to simulate non-ideally mixed (heterogeneous) conditions. Different model types are presented: approaches utilising CFD simulations (chapter 2.5.4.1) and compartmentalised models (chapter 2.5.4.2), as well as combinations of both.

From the presented model types and modelling methods, one was selected that meets the requirement of being able to simulate processes with non-ideally mixed conditions and be used for model-based control and optimisation methods. Detailed explanations for the selection of the mathematical model are outlined in the chapters 2.5.3.4, 4.5 and 5.2.

These model-based methods play a key role in modern bioprocess control and optimisation (chapter 2.6). For this reason, two strategies for the model-assisted design of processes are presented, which were modified and applied during this thesis: Nonlinear model predictive control (NMPC) (chapter 2.6.1) and model-assisted Design of Experiments (mDoE) (chapter 2.6.2).

2.5.1 Model development and validation process

The model development process must begin with the clear definition and establishment of a goal and purpose of the model. The model development can either start from nothing or an existing model can be extended, modified, and/or adapted. The complexity of the resulting mathematical model is a consequence of the definition of the goal and purpose. In general, mathematical models should always be as simple as possible, but as complex and detailed as necessary (Moran and Ostrom 2005; Bailer-Jones 2017; Robinson 2018).

Workflows were developed to standardise and facilitate the procedure of model development for biotechnological processes (Moser 1988; Kroll et al. 2017; Almquist et al. 2014). Figure 5 illustrates a simplified model development workflow based on Sixt et al. (Sixt et al. 2018) which was slightly modified during this work (Moser et al. 2020).

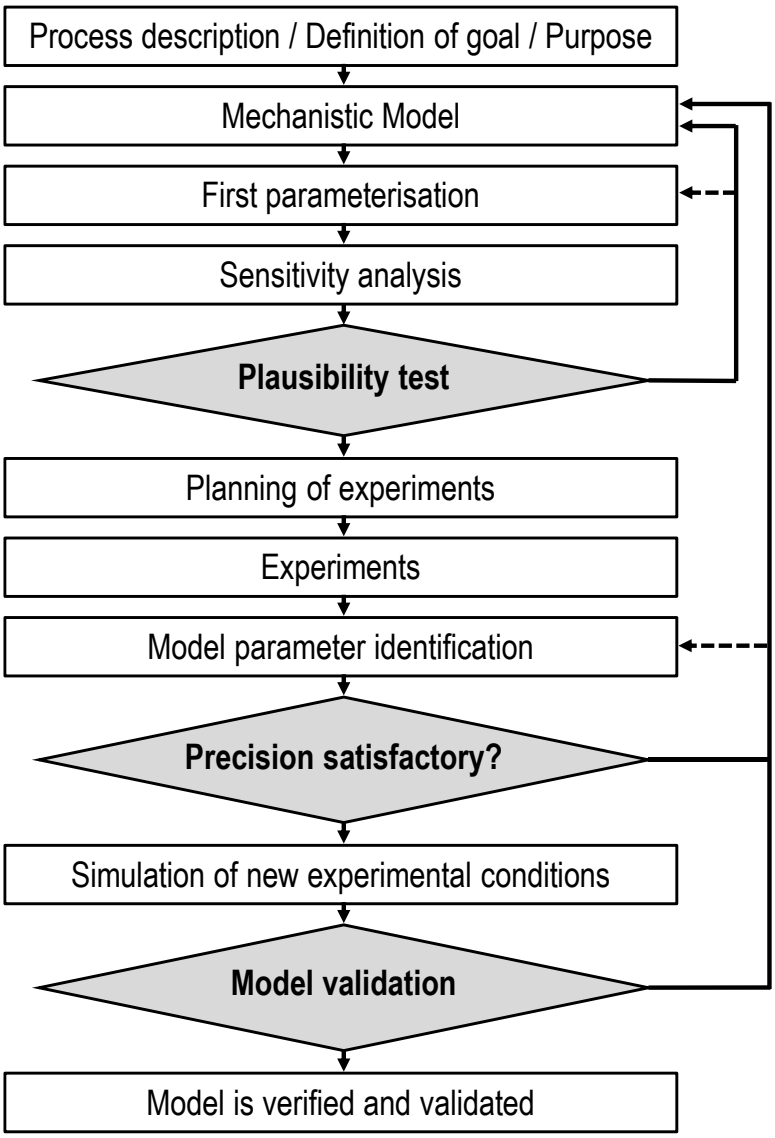


Figure 5: Simplified modelling workflow for mechanistic process models used as a basis in this work (dashed arrows are alternative paths).

The goal encompasses a comprehensive and detailed description of the process that is to be modelled and a definition of the system boundaries (Daume et al. 2020). Possible purposes might be the utilisation of the mathematical model in bioprocess control and optimisation or as a soft sensor. In the context of this work, the model shall be usable in processes under heterogeneous conditions and for cultivations within two interconnected stirred tank reactors. The defined goal and the system boundaries derived from the process

description can also be modified and adapted during model development (and process development) if required.

The process description includes all relevant cause-effect relationships and control variables. This requires the identification and analysis of typical, relevant atypical and special cultivations, which shall be described by the model. The necessary process variables, parameters, input/outputs of the model and boundary conditions are defined and derived from the description. Based on this information the kinetics, metabolic pathways, and reaction schemes are selected, formulated, and implemented into the mathematical model. The result is the first running version of the model. This model is tested for the model integrity, which checks if the model is producing reasonable simulation results.

A rudimentary parameterisation of the model is performed based on literature or historical data. Afterwards, a sensitivity analysis is frequently conducted, where individual parameters and values might be modified one at a time. The results of the sensitivity analysis are used to test the model simulations for the general plausibility. If the simulation results do not meet expectations and are not plausible, it may be necessary to update the initial parameter set, introduce and implement new process variables and/or modify or expand the model equations. (Moser 1985; Schaich et al. 2001; Leifheit et al. 2007; Sixt et al. 2018; Daume et al. 2020)

If the plausibility test is successful, experimental conditions and settings are planned. These experiments are often designed with empirical methods such as one factor at a time (OFAT) or design of experiments (DoE) approaches (Lee and Gilmore 2006; Franceschini and Macchietto 2008). The planning phase might also already be assisted by the mathematical process model (e.g. through determining or approximating reasonable boundaries of the experimental design space). The model development, parameter identification and sensitivity analysis could also be supported by model-based design of experiments methods (Cenci et al. 2023; Franceschini and Macchietto 2008).

Afterwards, the experiments are carried out and offline as well as online data are determined. The newly acquired experimental data are utilised to identify the model parameters. Optionally, initial values of process variables could be identified to improve the quality of the identification. The parameter identification needs a quality function which should be minimised or maximised during identification. For example, the simplest case would be the minimisation of the sum of the absolute differences between experimental and simulated data or more complex the maximisation of the coefficient of determination R^2 (chapter 4.6.2). In this this work the weighted mean squared deviation is minimised (WMSD; explained in chapter 4.6.1), which calculates the squared difference between experimental and simulated data points and applies a weighting value for each comparison The WMSD

is minimised using a downhill simplex method (based on the Nelder-Mead algorithm) (Nelder and Mead 1965).

After identification, the model precision/accuracy is evaluated. This is most of the time qualitatively done by calculating the coefficient of determination R^2. A purely quantitative evaluation of the "fit" may not be sufficient to assess the quality or precision of a model. For this reason, the quality of the identification should always additionally be checked by visualising data and comparing the experimental and simulated data graphically. If the model does not meet the predetermined precision (e.g. determined by calculating the coefficient of determination R^2), it needs to undergo a reparameterisation or even more likely, modification and/or extension. These steps are crucial for achieving a well-adjusted and calibrated model. (Schaich et al. 2001; Sixt et al. 2018)

The last step involves the model validation, where the model is utilised to simulate entirely new experimental conditions that are not covered by the data used for the parameter identification. The model is used to predict state variables and measurements. The experimental conditions might be determined with the application of model-based design and optimisation methods (chapter 2.6). The outcomes of these experiments are compared with the already acquired simulated data. Achieving the required level of precision indicates that the model accurately reproduces experimental data and possesses the ability to make reasonable predictions. If the required level of precision is achieved, the model is deemed validated. (Moser 1985; Sixt et al. 2018)

This workflow was followed during the development and validation of the novel mathematical model to simulate *S. cerevisiae* cultivations in homogeneous and controlled heterogeneous (non-ideally mixed) environments (chapters 5 to 8).

2.5.2 Model structure and submodel framework

One approach to facilitate model development is by organising the model structure. This can be done in the form of a shell structure as proposed by Blesgen, Hass and Kuntzsch (Blesgen and Hass 2010; Kuntzsch 2014). The shell structure is consisting of six shells, the biokinetic submodel, physico-chemical submodel and reactor submodel, as well as the plant and periphery model, control and automation model and the real-time connection to the physical bioreactor (Figure 6).

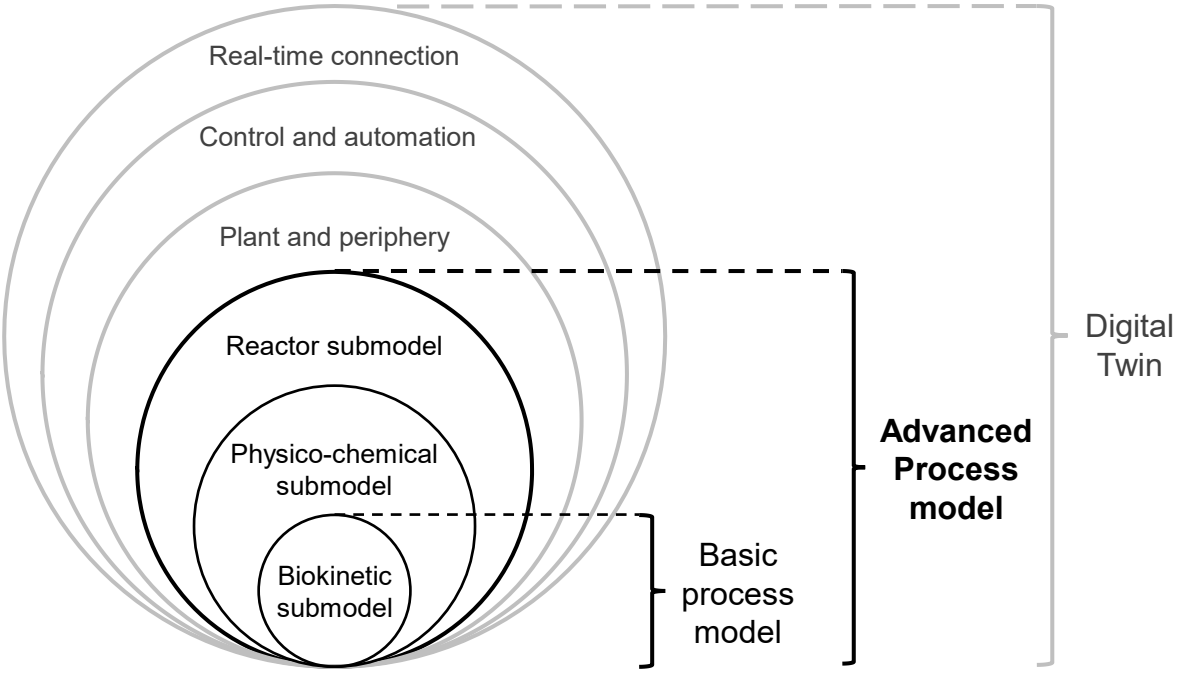


Figure 6: Shell structure of basic process models, advanced process models and full Digital Twins containing the biokinetic, physico-chemical and reactor submodel, as well as the plant and periphery model, control and automation model and the (real-time) connection to a physical plant. Based on Blesgen and Kuntzsch (Blesgen and Hass 2010; Kuntzsch 2014)

The inner core within the shell structure is the biological submodel which calculates the biomass growth rates, as well as uptake and production rates of metabolites based on kinetics, involving inhibition and limitation terms. These kinetic equations are frequently integrated directly into the mass and/or concentration balances of the reactor submodel. Consequently, there is not an essential requirement for a definite distinction between reactor and biological submodels. If the reactor and biological submodels are separated, the resulting rates derived from the biological part are transmitted to the reactor submodel. Types of biokinetic submodels (e.g. mechanistic and empirical) are presented and explained in chapter 2.5.3) (Himmelblau and Riggs 2012; Chmiel et al. 2018).

The second layer of the shell structure contains the physico-chemical submodel. This submodel calculates the rates of change of physico-chemical variables such as medium temperature, pH value, dissolved oxygen concentration, gas-phase concentrations, foam level etc. Physico-chemical variables might be simulated since they contain valuable process information which are frequently available in the form of continuously recorded and swiftly available online-data. Applying physico-chemical submodels allows for the usage of these data and process knowledge to identify parameters. This might improve parameter identification, if the data contains useful information, but might also impede

parameterisation, as the number of identifiable parameters and the models' simulation time increase. (Blesgen and Hass 2010; Hass 2016; Brüning 2016)

The reactor submodel contains differential equations that compute the values of all physico-chemical and biokinetic variables (e.g. biomass and metabolites) from the rates calculated in the biokinetic and physico-chemical submodel. These three submodels are forming the basis of advanced process models (Moser et al. 2024). These advanced process models are well suited for the application during process development and for model-based control and optimisation methods. The further development to full Digital Twins briefly outlined below was not necessary for the fulfilment of the objectives defined in the context of this work. (Blesgen and Hass 2010; Moser et al. 2020)

Various authors and research groups have published definitions for Digital Twins (Glaessgen and Stargel 2012; Grieves 2015, 2016; El Saddik 2018; Zobel-Roos et al. 2019). Consensus between authors is that Digital Twins are detailed digital replications of physical entities with data transfer between the physical and digital entities (El Saddik 2018). For this reason, besides the advanced process model, a plant and periphery model, a control and automation model and the (real-time) connection to the plant might be needed to obtain a Digital Twin. The plant and periphery model describes the periphery connected to the bioreactor such as pumps, tanks, valves, or pipes. The control and automation methods of the physical twin are realised in the control and automation model. Here, control strategies can be tested and transferred to the physical bioreactor. (Blesgen 2009; Appl et al. 2021; Moser et al. 2020; Appl 2023)

2.5.3 Mathematical (biokinetic) model types

Mathematical process models can be categorised as mechanistic, empirical and hybrid models (see Figure 7) (González-Figueredo et al. 2019; Craven et al. 2013). This applies to advanced process models but also to the respective submodels (e.g. biokinetic, physico-chemical submodels and reactor submodels). The model types presented in this chapter are mainly based on the example of biokinetic models.

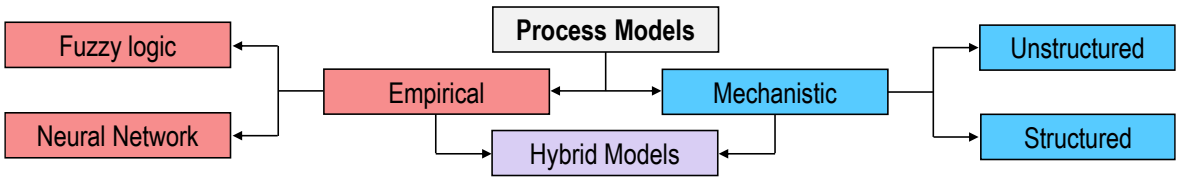


Figure 7: General classification of process models into mechanistic and empirical models.

Mechanistic models incorporate process knowledge and are derived from mechanical, physical, chemical, and biochemical relations and equations (Schubert et al. 1994). They are classified as structured or unstructured models, which are further subdivided into

segregated (e.g. population models) and unsegregated models (Bailey 1998). Empirical models are purely data driven process models or derived from process knowledge that rely solely on the processing of information, such as artificial neural networks (ANN) or statistical models (Doherty et al. 1997; Ronen et al. 2002a; Schubert et al. 1994; Stosch et al. 2014). Heuristic process knowledge can be included into process models in the form of fuzzy sets including thresholds for inhibition, limitation and/or optimal temperature, pH and DO ranges. (Zadeh 1965; Konstantinov and Yoshida 1990; Kosko 1992). Hybrid models combine the mechanistic and empirical approaches into one process model (see Figure 7) (Stosch et al. 2014).

2.5.3.1 Mechanistic models

Mechanistic models used for biotechnological processes are based on biochemical and physical principles. This facilitates the description of cause-and-effect relationships and enables the possibility of reasonable extrapolation beyond the data used for the identification of the model. Non-mechanistic models, on the other hand, strictly speaking, only allow for interpolation within the data used for the identification of the model.
Mechanistic models can be categorised into structured and unstructured models as well as non-segregated and segregated models (see Figure 8).

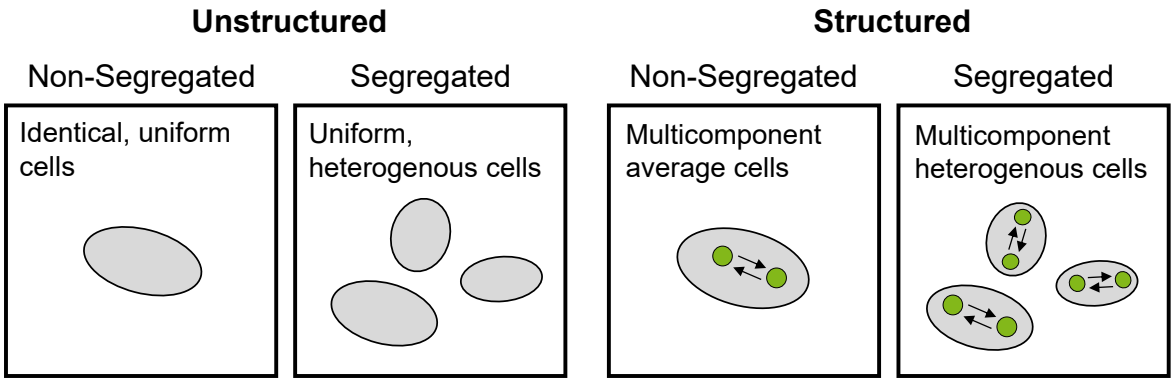


Figure 8: Classification of mathematical, biotechnological process models into unstructured and structured as well as non-segregated and segregated models. (Bailey 1998)

Structured models divide the cellular structure into different compartments; each composed of different cellular components with specific functions. In contrast, unstructured models view cells as homogeneous. Both structured and unstructured models can be subdivided into segregated and non-segregated models. Non-segregated models consider all cells to be identical and in the same state (e. g. in the same metabolic state), whereas segregated models considers that cells are different from each other and thereby establish cell populations (Tsuchiya et al. 1966; Fredrickson et al. 1970; Ramkrishna 1979; Bailey 1998).

Unstructured models

Unstructured, non-segregated models assume that all cells are in the same physiological state and model them as black boxes. Only extracellular factors have an influence on the growth and metabolism of cells in unstructured models. The most well-known and prevalent model is the Monod-Model (Monod 1949), where the growth limiting factor is the concentration of a single substrate. There are several other unstructured kinetics primarily influenced by substrate limitation (Blackman 1905; Teissier 1936; Moser 1958; Ming et al. 1988). The Monod-Model model has been extended multiple times to incorporate effects such as substrate or biomass inhibition (see Figure 9) (Contois 1959; Andrews 1968).

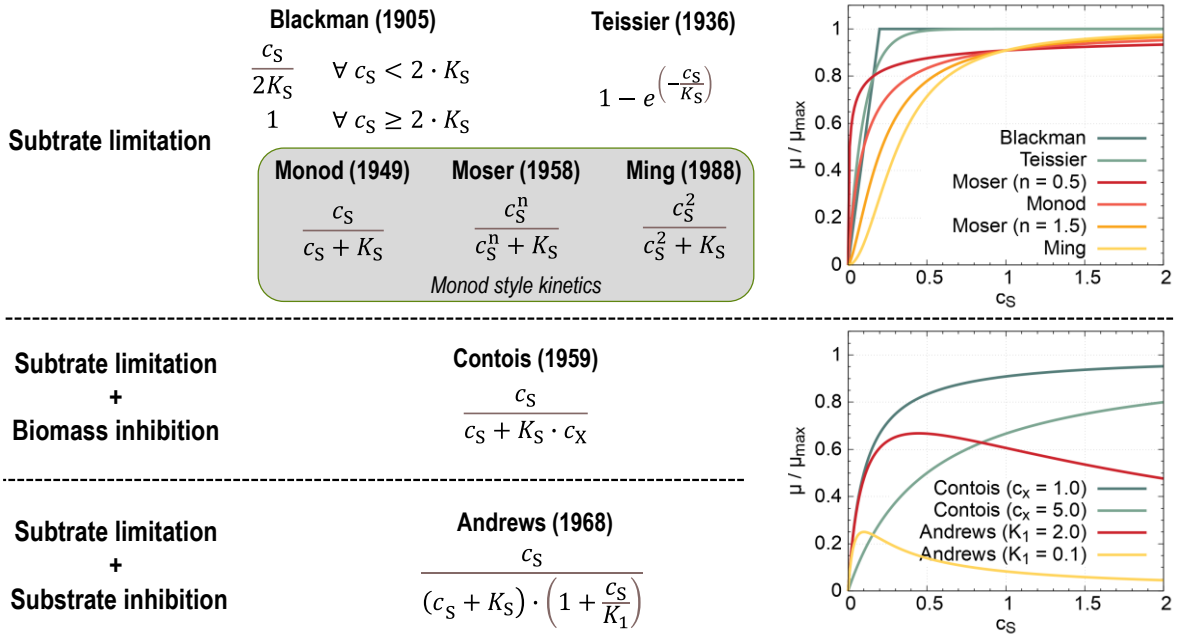


Figure 9: Selection of kinetic expressions used in unstructured models. All displayed graphs are for half saturation constants K_s with numerical values of 0.1 and their corresponding units (Blackman 1905; Teissier 1936; Monod 1942, 1949; Moser 1958; Contois 1959; Andrews 1968; Ming et al. 1988).

Few model parameters and the generally low complexity of unstructured models accelerate and facilitate model development. In comparison to other model types, unstructured models can be parameterised with a limited amount of process data. As a result, it is possible to employ them during early stages of process development. Simple unstructured models have limitations in describing complex growth and metabolic behaviour, such as transitions between growth phases, different growth patterns, and their influence on biocatalytic and metabolic rates. They are primarily suited for the quantitative reproduction of cultivations. Unstructured models are mainly utilised to describe individual effects of microbiological cultivations, such as substrate limitation or product inhibition. Several studies have

extended unstructured models to encompass more effects (Sonnleitner and Käppeli 1986; O'Neil and Lyberatos 1990). Despite yielding favourable results, these extensions elevate the complexity of the models and thereby diminishing and compromising the core advantage of unstructured models.

One of the most known unstructured models for *S. cerevisiae* was developed by Sonnleitner et. al (1986), which incorporates stoichiometric equations for the growth of *S. cerevisiae* on glucose and ethanol. The model includes different oxidative metabolic pathways, with and without ethanol production and for reductive metabolism for growth on ethanol. Glucose and ethanol uptake rates are modelled by Monod kinetics while the oxygen uptake is linearly correlated with the glucose uptake rate during oxidative metabolism. Ethanol is only consumed when glucose levels are low. Overall, this unstructured model can successfully describe yeast cultivations with good agreement between determined and simulated data. However, the model is only capable of modelling the growth phase of yeast. A lag-phase, stationary phase or death phase are not considered. Furthermore, the inhibition by ethanol and other influences on relevant rates were also not included in this model at this stage. (Sonnleitner and Käppeli 1986)

Structured models:

In structured models, the cells are not viewed as homogeneous. Instead, the biomass is divided into compartments with distinct functions. This allows for the description of changes in cell properties, while all cells are identical and change as one in time. Structured models are more complex than their unstructured counterparts and have an increased number of model parameters.

Williams (1967) proposed a structured model for the universal features of cell growth from the lag-phase to the stationary phase. The model divides the biomass into two portions, a synthetic portion and a structural/genetic one. The synthetic portion is fed by an available substrate, and the structural portion is fed from the synthetic portion. Cell division is only possible when the structural portion has doubled, independent of the state of the synthetic biomass portion. (Williams 1967)

Zeigler and Weinberg (1970) developed a technique for aggregating multiple cell compounds into single compartments for bioprocesses with *E. coli*. The model considers mass balances of these biomass compartments and of extracellular variables (Zeigler and Weinberg 1970). Shuler et al. (1979) employed this approach to construct a growth and cell division model of a single *E. coli* cell comprising 12 compartments (Shuler et al. 1979).

Nielsen et al. (1991) and Nielsen (1993) proposed a structured model of filamentous microorganisms in which the growth is described by dividing the hyphal elements into three completely different compartments (apical, subapical, and hyphal compartment). Formation

of new biomass only occurs in the apical and subapical compartments whereas hyphal cells are inactive. Cells are initially apical cells formed from subapical cells and transform into subapical cells. They eventually inactivate into hyphal cells (Nielsen 1993; Nielsen et al. 1991). The model was experimentally verified for cultivations of *Aspergillus oryzae* (Agger et al. 1998).

Witte (1996) describes a dynamic model with four biomass compartments for the description of *Cyathus striatus* cultivations. This model structure allows for the description of multiple growth phases. (Witte 1996)

These four biomass compartments are:

- Primary metabolism biomass including reactions linked to cell growth (production of primary and secondary biomass)
- Secondary metabolism biomass including reactions (e.g. for product formation)
- Structural biomass representing cell components like cell wall, membranes, and complex polysaccharides (not responsible for catalysing growth or product formation)
- Inactive biomass representing decaying and inactive primary and secondary biomass e.g. defect enzymes or RNA/DNA

This structured modelling approach was extended to six compartments by Brüning (2006) to achieve a greater generality than most segregated models at the same or lower levels of complexity, enabling the description of multiple different microorganisms and mammalian cells (Figure 10). In both models the mass balance is closed by implementing the overall stoichiometric equations of metabolic reactions into the model (Brüning 2016; Witte 1996).

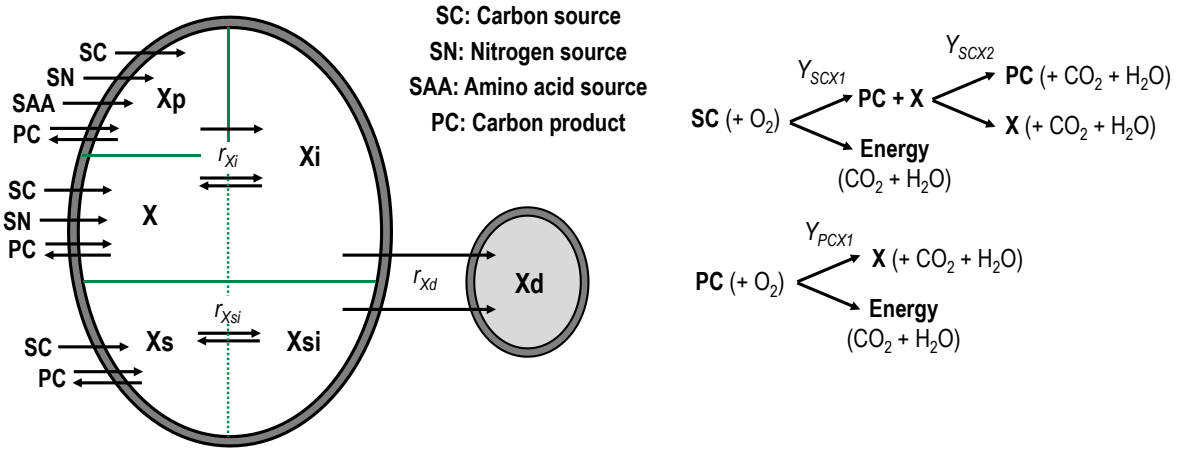


Figure 10: Schematic structure of the biokinetic submodel (6-compartment model). Illustrated are the six different biomass compartments, selected rates between compartments and the substrates. Furthermore, two chosen metabolic processes are presented. The aerobic metabolism of the carbon substrate (SC) and O_2 to energy via respiration, carbon product (PC) via fermentation and biomass (X), the respective yield coefficients and side products (CO_2 and H_2O) as well as the diauxic aerobic metabolism of the carbon product to biomass and energy via respiration (based on (Brüning 2016; Brüning et al. 2017)).

The model by Brüning (2006) utilises a special nomenclature of the biomass compartments (marked by X, with g L^{-1} as the unit), which is utilised in the following section to explain the model.

In the model, the biomass is divided into active and inactive compartments including a dead biomass compartment. Compartment X comprises of proteins and RNA/DNA and necessitates at least a carbon source and nitrogen source for formation. The production of non-growth related products is accomplished by Xp and its formation from X is requiring a carbon and nitrogen source. Lipids, like the cell membrane, cell wall, and polysaccharides are represented by Xs. Xs is not auto-catalytically active and is also generated from X. X and Xp can be inactivated to compartment Xi and Xs can be inactivated to form Xsi. These inactivated compartments can be reactivated; the reactivation requires maintenance energy through e.g. respiration. Finally, a mortality rate determines the generation of “dead” biomass Xd formed from the inactivated compartments. The mortality, inactivation and activation rates are influenced by cultivation conditions (Brüning 2016).

This structured model utilises double-sigmoid functions to extend kinetics and to model variable, state-dependent yield coefficients (Brüning 2016; Gerlach et al. 2013). The extension to six biomass compartments enabled the simulation of desired growth phases

(lag, exponential, linear, stationary, and death phase) and the description of energy metabolism.

The structure allows simulations of various organisms, such as yeast (*S. cerevisiae*), bacteria (*E. coli* & *L. delbrueckii*), basidiomycetes in suspension culture (*C. striatus*), as well as mammalian cell lines (hybridoma and CHO cells), without changing the model structure (Brüning 2016). For these reasons, the model was utilised as the basis for the mechanistic model developed for this work (described in chapter 4.5.3).

Metabolic Flux Models

Metabolic flux models are a subtype of structured models which are the result from a metabolic flux analysis (MFA). They provide a quantitative description of the microorganisms' metabolic pathways, reactions, and their responses by different conditions, actions or genetic manipulations (Varma and Palsson 1994; Falco et al. 2022). They take changes of extra- and intracellular concentrations of substrates and products into account, and consider mass balances through stoichiometric equations and constraints (Vallino and Stephanopoulos 1993; Iwatani et al. 2008). The intracellular fluxes are often quantified via carbon labelling e.g. by using 13C labelled substrates (Wiechert 2001). MFA has been instrumental in metabolic engineering by determining and characterising the metabolic state of cells and has been extensively utilised in various biotechnological processes (Iwatani et al. 2008; Chassagnole et al. 2002; Rizzi et al. 1997).

MFA is based on stoichiometric equations of the biochemical reactions within biological systems. Main assumption for MFA is that the system is in a stationary state. Result of the MFA is a stoichiometric matrix representing intra- and extracellular metabolites (rows) as well as the metabolic fluxes (columns) (Antoniewicz 2015; Wiechert 2001). Stoichiometric MFA is not able to calculate certain situations such as parallel reaction pathways or pathway cycles where none of the fluxes in the parallel pathways or the cycle are measurable (Wiechert 2001). MFA often considers only seconds of a stationary state of cells or specific metabolic pathways, such as the glycolysis in *S. cerevisiae* (Rizzi et al. 1997).

Reuss et al. developed metabolic models for *S. cerevisiae* (Rizzi et al. 1997; Vaseghi et al. 1999; Visser et al. 2000; Müller et al. 2003; Zakhartsev et al. 2015) and even combined those with hydrodynamics calculated with CFD simulations (Lapin et al. 2010; Lapin et al. 2006).

Dynamic MFA models are dividing the process into short quasi-stationary time periods of a few seconds to minutes. Average external rates and fluxes are calculated for each time interval. The results of these models are evaluated at various times and combined to achieve time profiles (Antoniewicz 2015).

Another option is to develop detailed dynamic models which are derived from MFA. These models describe the dynamics of enzyme kinetics with coupled differential equations. This necessitates the measurement of many numerous kinetic parameters, which can be extremely difficult to obtain for complex reaction systems and microorganisms (Rizzi et al. 1997; Theobald et al. 1997; Colijn et al. 2009).

For these reasons stationary MFA and derived dynamic metabolic flux models cannot be generalised and are strictly specific to the process and organism involved. Moreover, they require a large amount of experimental data for the model development, calibration, and validation, as well as computational resources (Hameri et al. 2021).

Segregated Models and Population Models

Segregated models and population models differentiate between biomass states, such as stages of the cell cycle and cell age, or morphological differences among cells. The simulations feature multiple segregated cell populations with different growth and production behaviours which can change independently over time. Their potential is extremely high, but due to the high complexity of these models, their usage is challenging (Heins and Weuster-Botz 2018; Charlebois and Balázsi 2019).

The application is usually restricted to a single process as the conditions necessary for the change of states are highly specific and usually determined by experts (knowledge-based system). Moreover, identifying parameters and utilising optimisation algorithms can be both time consuming and complex. These aspects preclude the usage of population models for this work as the simulation time is usually too long for an effective application in model-based control methods.

An example for a population models is an age models developed by Shu (1961) as well as Blanch and Rogers (1971) (Blanch and Rogers 1971; Shu 1961). Blanch and Rogers (1971) categorised cells into two age groups, immature and a mature cell stage for the production of gramicidin S in *Bacillus brevis*. The antibiotic gramicidin S is only produced during the mature phase of a cell.

Megee et al. (1970) developed a mathematical model with five different states for the growth and production metabolisms of *Aspergillus awamori.* The first state is responsible for cell growth, the four other states for the formation of various products such as DNA, RNA, proteins and enzymes (Megee et al. 1970).

Functional state models

Functional state models represent a mechanistic modelling approach that links multiple models. The process is divided into phases (functional states). Each of these states is usually described by an individual mechanistic model whose parameters are exclusively

valid for this particular state, dominated by a particular functionality (e.g. metabolic pathway) (Zhang et al. 1994; Hristozov et al. 2001; Roeva et al. 2007).

The individual model equations of each functional state can be solved independently, provided that the initial states of every stage are known (Roeva et al. 2007). The process of separating the model into individual functional states has the advantage that the models of each stage are easier or even analytically solvable. Each local model comprises only a restricted number of parameters, and these can be identified independently (Vilums et al. 2012).

It is of importance to define the individual states, boundaries, and transitions between states accurately in order to describe a process. The boundaries and transitions between functional states are delineated by a meta-model. This meta model may be represented by a set of rules, for instance, the cultivation state (e.g., lag, exponential, stationary or death phase), the respiratory quotient (RQ), the oxygen concentration, or the concentration of critical metabolites that may inhibit or limit cell growth, such as ethanol or acetic acid (Roeva et al. 2007; Hristozov et al. 2001; Roeva and Pencheva 2014). For example, the transition between states of functional state models may be determined by fuzzy logic (see chapter 2.5.3.2).

The characterisation of these states is of importance in the utilisation of functional state models given the potential for overlap of biological effects associated with individual states. This presents a challenge when attempting to derive rules that adequately and precisely represent each state.

2.5.3.2 Empirical, non-mechanistic models

Non-mechanistic models are built solely from mathematical correlations. This model type does not include mechanical, physical, chemical, or biochemical relations. Instead they describe processes in a more abstract manner (Konstantinov and Yoshida 1990). Due to their generic structure, non-mechanistic models are typically less time-consuming in their development compared to mechanistic models (Stosch et al. 2014).

Non-mechanistic models are limited to reproduce user knowledge and preexisting experimentally acquired data. Additionally, non-mechanistic models solely consider consequences and perturbations that surpass a user-defined significance level in the available data. It is critical and crucial that the signal-to-noise ratio (significance level) is accurately identified. Simple examples of empirical models might be decision trees or graphical characteristics derived from experimental data.

The simulation outcomes from non-mechanistic models are constrained to reproducing only the responses reflected in the data of previously conducted experiments or the process knowledge of the model developer (Glaessgen and Stargel 2012; Zobel-Roos et al. 2019).

Therefore, non-mechanistic models, such as artificial neural networks (ANN) or statistical models, have a limited capability for prediction. The validation of these models poses a challenge. Non-mechanistic models are more specific to the real process than mechanistic models because they are purely data driven. Extrapolation beyond the data used for model development and parameter identification is difficult, if not impossible, which greatly restricts the transferability to other processes and their applicability in process development (Schubert et al. 1994).

Modelling with fuzzy sets

Fuzzy logic also known as fuzzy set theory was introduced by Zadeh (1965) as a form of multiple-valued logic. It is characterised by the fact, that an assumption has more than two true values (Zadeh 1965). In contrast, (binary) Boolean logic permits only two outcomes ("True" or "False") and all propositions take exactly one of those truth values (Boole 1847; Barnes 1995). The Łukasiewicz–Tarski logic proposed by Łukasiewicz–Tarski (1930) allowed three outcomes: "False", "True" and "Unknown", thereby employing a three-valued logic (Lukasiewicz 1930).

Fuzzy logic extends this to an infinite-valued logic to enable the numerical capture of a variable or characteristic between 0 and 1. Fuzzy set theory is therefore not constrained by pure definition if whether the considered propositions or objects are (or are not) elements of the set, it also defines to which degree they belong to the set.

Hence, fuzzy modelling offers an alternative approach to deterministic mathematical models (Honda and Kobayashi 2000). Modelling with fuzzy sets does not require a detailed mathematical description of the process. Instead, this approach relies solely on expert and process knowledge, which must be available and organised in the form of cause and effect chains (Karakuzu et al. 2006; Horiuchi 2002; Zadeh 1965). A main application of fuzzy logic is in process control, where promising results were achieved (Konstantinov and Yoshida 1990; Ying 1995; Shi and Shimizu 1992; Horiuchi 2002; Filev et al. 1985; Ronen et al. 2002a). Fuzzy sets modelling can also be used in combination with mechanistic or ANN models (Ronen et al. 2002b).

Artificial neural networks

Artificial neural networks (ANN) imitate the operational principles of the brain. They consist of an input layer and an output layer, connected by one or more hidden layers (Montague and Morris 1994; Ławryńczuk 2008; Grahovac et al. 2016). Figure 11 illustrates a basic neural network with three process inputs and two outputs.

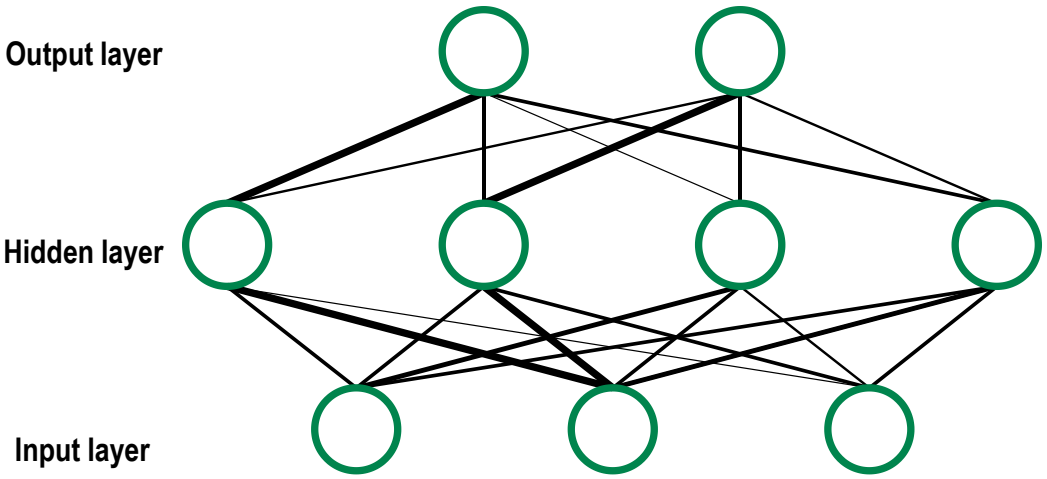


Figure 11: Simple artificial neural network. Circles represent simplified neurons. Thickness of lines represent the exemplary strengths of the connection between the layers.

The nodes in the hidden layer sum up the weighted inputs and calculate the outcome through sigmoid functions in most cases. The weightings of the inputs and parameters of the sigmoid functions are parameterised through the "training" or "learning" process. To obtain reliable results using ANN, large quantities of data are required for the learning process. If insufficient experimental data are available, further data points are produced occasionally by mechanistic models (Grahovac et al. 2016; Karakuzu et al. 2006). The model training requires time but after that ANNs exhibit extremely fast computing speeds. ANNs are purely data-driven models and are not dependent and/or based on process knowledge/insight. Recently, ANNs have been used for bioprocess control and optimisation (Chen et al. 2004; Franco-Lara and Weuster-Botz 2005; Karakuzu et al. 2006; Nagy 2007; Ławryńczuk 2008). Since ANNs are purely mathematical data-driven correlations, they cannot offer mechanistic insights into the process (Karim et al. 1997). Additionally, ANNs have limitations since only effects and relationships are modelled that are significant within the data sets used for the training and learning process. This means, strictly speaking, ANNs only allow for a description, but not for prediction (Zobel-Roos et al. 2019).

Neural networks were already employed for the dynamic temperature modelling of continuous *S. cerevisiae* cultivations by Nagy (2007). The resulting ANN successfully modelled cultivations considering the heat of the metabolic reaction and was integrated into a control algorithm to regulate cultivation temperature (Nagy 2007).

However, since ANNs are data-driven and therefore need a lot of experimental data for the training process, they are not applicable in the early stage of process development. The required data is not available for *S. cerevisiae* processes under controlled heterogeneous conditions in scale-down models. Furthermore, the application in process

control and optimisation requires a model which is capable of extrapolating beyond the data used for the parameterisation. This cannot be fully satisfied by ANNs.

2.5.3.3 Hybrid models

Hybrid models may be considered as a combination of mechanistic models and data-driven non-mechanistic models in order to merge different types of process knowledge (Stosch et al. 2014). This includes prior knowledge about the process in the form of material and energy balances, thermodynamic, and kinetic laws. Historical data can be implemented into hybrid models in the form of ANNs (Schubert et al. 1994; Doherty et al. 1997; Kosko 1992). Hybrid models strive to balance the benefits and drawbacks of both mechanistic and non-mechanistic models. While mechanistic models are capable of producing predictive results, their development is often time-consuming and requires in-depth process knowledge. In contrast, non-mechanistic models, can be created quickly and have widespread applicability, but solely offer satisfactory descriptive properties (Stosch et al. 2014). Hybrid models offer a distinct advantage over purely data-driven modelling approaches as they offer a higher accuracy, more efficient model development, and better extrapolation properties compared to purely data-driven modelling approaches.

An evaluation by Stosch et al. indicates that the implementation of hybrid models and their parameterisation algorithms is error- and labour-intensive, making the implementation challenging (Stosch et al. 2014). However, hybrid models have been successfully and effectively utilised in process control and optimisation (Madni et al. 2019; Galvanauskas et al. 2004; Schubert et al. 1994).

Hybrid models also need experimental data or process knowledge for the development of the empiric model part, neither are available at the current state for processes under controlled heterogeneous conditions.

2.5.3.4 Classification of model types

The introduced model types can be categorised into six distinct categories as illustrated in Figure 12. This categorisation was used to aid the selection of an adequate model type for the mathematical process model of this work. The selected categories comprise data requirement for parameterisation and model development, model complexity, potential of utilisation, process knowledge needed for model development, time, and effort necessary for model development and lastly the computational effort or simulation time required for simulations.

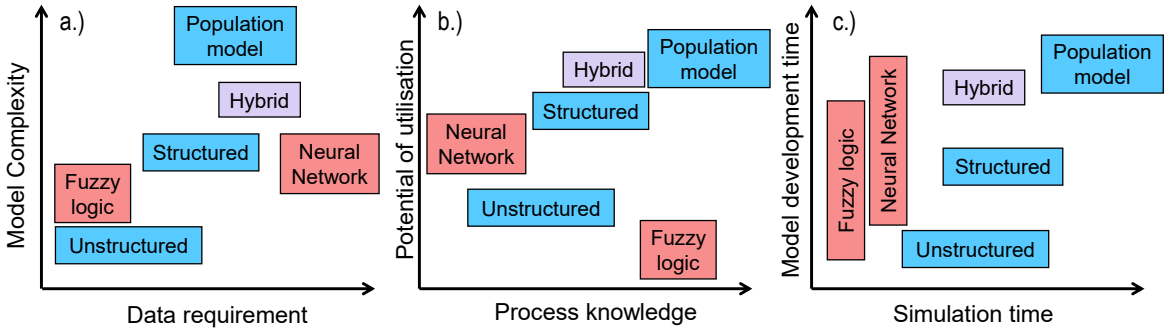


Figure 12: Simplified allocation of different process model types into six categories.

Unstructured models are simple in structure, have low model development time and need low amount of data for the creation. The disadvantages are derived directly from these advantages, as unstructured models only describe a few effects and can therefore only describe processes to a limited extent which constrains the application possibilities in model-based control and optimisation.

Structured models on the other hand allow for a sufficient description of the dynamics of cultivations (Brüning 2016), while providing a simplistic structure. Additionally, structured models offer the foundation for an adequate description of key effects for process optimisation based on biological understanding (Witte 1996). Furthermore, the simulation time and model development time are low enough for the effective application in model-based control and optimisation.

Population models show significant potential especially for the simulation of cultivations under heterogeneous conditions since gradients in the reactor medium influence the population. The complexity and high number of parameters of population models make the computation time a limiting factor (Figure 12 c.), particularly for parameter identification and model-based optimisation methods.

Non-mechanistic models offer fast simulation times and a potentially fast model development (Figure 12 c.). Neural networks in particular need a lot of data for model development. Given the limited amount of data and process knowledge available during early stages of process development for heterogeneous cultivation, non-mechanistic models are not an option.

In this work, it shall be tested if a structured, mechanistic modelling approach is sufficient and capable of accounting the influence of non-ideal mixing conditions and hydrodynamics on cultivations of *S. cerevisiae*. Furthermore, the chosen model should be usable in model-based design and optimisation methods. Possible approaches for the consideration of non-ideally mixed conditions in mathematical models are presented in the following chapter 2.5.4.

The selected process model and the structure are further described and elaborated in greater detail in the material and methods section (chapter 4.5) and modifications to the model are described in chapter 5.2.

2.5.4 Mathematical (reactor) models with non-ideally mixed conditions

Heterogeneous conditions might have an influence on the process performance of large-scale cultivation processes. Different approaches exist to consider and simulate the hydrodynamics of the reactor medium and approximate the influence on the cultivation.
One option is the simulation of the fluid dynamics in large-scale (and small-scale) bioreactors. This is usually performed with computational fluid dynamics (CFD). The results of the CFD simulations can be used to develop hypotheses with respect to the cultivation behaviour. In addition, measures can be taken to improve the process on a large scale e.g. internal installations or stirrers can be designed to improve the mixing behaviour.
A different option is the compartmentalisation of the reactor submodel into compartments (compartment model approach) or zones which are interconnected (network of zones). Each compartment (zone) represents an ideally mixed system. Fluid and fluid properties are exchanged through the connections between zones (hydrodynamics). The hydrodynamics are usually based on experimental observations or CFD simulations.
Both CFD simulations and network of zones models have already been coupled with biokinetic models. Main advantage and potential of these combinations is, that the model parameters identified on the laboratory scale in ideally mixed systems are usable within the coupled biokinetic models without compensating for the hydrodynamics. This is possible since the hydrodynamics of the system are already described by the CFD or compartment model.
CFD models are introduced in chapter 2.5.4.1; compartment model approaches and network of zones models in chapter 2.5.4.2

2.5.4.1 CFD simulations

Computational fluid dynamics (CFD) simulations are able to simulate the flows and flow patterns within reactors. CFD simulations divide the volume of the medium into smaller volumes (mesh cells). The numerical solution of CFDs is usually based on solving the Navier-Stokes equation for simulating fluid velocities. Variables that may be relevant for the scale-up, such as the k_La value, shear forces or the mixing time of the system might be estimated. (Fitschen et al. 2023)
CFD simulations can be two-dimensional (2D) and three-dimensional (3D). In addition to the liquid phase (reactor medium), the gas phase (aeration) and, more rarely, the solid phase (biological phase, cells) might also be considered in the simulations. This increases

the complexity of CFD simulations, as the phases might influence each other. With the exponential increase of computational power, it is also now possible to incorporate biokinetic reactions into CFD simulations (Bezzo et al. 2003; Schmalzriedt et al. 2003; Moilanen et al. 2005; Morchain et al. 2014; Delafosse et al. 2014).

Euler-Lagrange (Lagrangian) or **agent-based modelling** is one approach to utilise the results of CFD using the discrete element method. The positions of the cells are described as they move through the reactor over time. The positions of each simulated cell are stored as a function of time. This enables the computation of the environment each cell experiences at each time in their lifetime. (Enfors et al. 2001; Kuschel et al. 2017; Delvigne et al. 2017)

Increasing computational power allowed not only better insight into metabolic dynamics but also increased the possible resolution of the reactor contents. This enables the possibility of simulating millions of cells and their journeys through bioreactors. The influence of concentrations gradients (e.g. substrates, products, oxygen) on the cells can be considered using their individual positions in the reactor with biokinetic equations. These equations are calculating rates, such as the substrate uptake rate, growth rate or byproduct formation rate based on the surroundings of each simulated cell (Enfors et al. 2001; Lapin et al. 2004; Delvigne et al. 2017).

Haringa et al. simulated the positions of 175000 particles and their experienced conditions. They analysed the positions over time (lifelines), residence times and transition patterns of the particles in a 54 m³ stirred tank reactor. The results were transferred to the simulation of a fermentation of *Penicillium chrysogenum*. The simulated circulation times were comparable to experimentally determined times although the gas phase was not considered in the simulations. Furthermore, it was simulated that the substrate was completely consumed in more than half (57%) of the reactor. This resulted in the cells experiencing very different regimes (excess substrate and depleted substrate) every second to several tens of seconds. (Haringa et al. 2016)

Kuschel et. al used this Euler-Lagrange approach with a structured cell cycle model to investigate large-scale heterogeneity in *Pseudomonas putida* cultivations. The model employs CFD simulations to calculate the process conditions along Lagrangian trajectories through the bioreactor. The results demonstrate that the population undergoes strong heterogeneity, indicated by an up to 1.5 higher adenine triphosphate maintenance determined demand. (Kuschel et al. 2017)

In **Euler-Euler (Eulerian) models**, the simulated phases (liquid and gas phase for bioreactors) are considered as continuous, and the flow fields are viewed as a whole. The fluid properties are functions of time and space. Some examples for these fluid

properties are the velocity, pressure, and temperature. The direction and momentum of the fluid are calculated for each volume element by solving the Navier-Stokes equations.
Sarkar et al. used a multiphase (liquid and gas) Euler-Euler approach for the description of an aerated stirred tank reactor for the production of monoclonal antibodies with mammalian cell cultures. The mass transfer coefficient k_La was estimated based on process parameters such as the stirrer speed and aeration gas flow rate. Furthermore, the CFD simulation was combined with a population balance model to successfully estimate the size distribution of bubbles considering effects like coalescence and breakage of bubbles. The presented approach saved time and resources for the optimisation of mixing conditions compared to experimental investigations. (Sarkar et al. 2016)
Elqotbi et al. incorporated a simple Monod-type model into a three-dimensional CFD approach to simulate of *A. niger* cultivations. The model simulations provided a satisfactory agreement between experimental and simulated data, including the glucose substrate and bioproduct concentrations. (Elqotbi et al. 2013)
Morchain et al. (2013) coupled a population balance model with an Euler-Euler approach for a two-phase flow calculation. This model was utilised to investigate the biological adaptation to concentration gradients. The Eulerian balance was solved for each phase (gas and liquid) with CFD simulations. The group simulated oxygen and substrate gradients in a 70 m³ production reactor. The simulation results showed that the relatively slow growth rate of the microorganism was sensitive to the mean concentration in the medium, whereas the fast substrate assimilation was more sensitive to the substrate distribution. (Morchain 2017)
The many simultaneously occurring effects in biotechnological processes are causing complex biokinetics that can only be mapped with a corresponding biokinetic model. This impedes the integration of the biokinetic differential equations into CFD simulations and possibly makes effective simulation impossible due to e.g. excessive computation times and/or computer power requirements. Furthermore, the need for three-phase CFD simulations would have to be clarified and viscosity changes over the cultivation period might have to be simulated.
The mathematical model for this work must also be usable for model-based optimisation methods. A model consisting of a CFD simulation or calculating the lifelines of hundreds of thousands to millions of cells with high temporal resolution would certainly not be usable in at-line process control (e.g. in NMPCs) due to high simulation times of CFD simulations.

2.5.4.2 Compartment model approaches

Network of zones models (NoZ) or compartment model approaches (CMA) are different methods of simulating the influence of heterogeneous conditions on cultivations. These

models are used when concentration gradients are expected in either the liquid or the gas phase (Oosterhuis and Kossen 1984). This has been traditionally modelled through interconnected networks of zones (compartments), which exchange fluids and their properties (van Barneveld et al. 1987a; Delafosse et al. 2014).

Levenspiel (1972) developed a compartment reactor model to mathematically describe the non-ideal behaviour of chemical reactors. The model considers two connected compartments. A physical mixed stirred tank reactor with a stagnant (dead) zone is mathematically represented by an ideally mixed compartment, which is connected to a second compartment representing the dead zone. (Levenspiel 1972)

The determination of fluid exchange between compartments was initially based on visual observations of the flow patterns in reactors, measurements of fluid velocity or based on experimental measurements of circulation times / mixing times (van Barneveld et al. 1987a, 1987b). With increasing computational power, more realistic simulations were enabled, and the number of compartments increased (Vrábel et al. 1999; Zahradník et al. 2001).

CMAs only require the fluid velocities between compartments (cells) to describe mixing on the reactor scale (Mann et al. 1995; Delafosse et al. 2014). Interdisciplinary expertise is required for the combination and application of CMA and CFD, which remains a challenge (Morchain 2017). Moreover, the combination still requires a huge amount of computational power, limiting their application to the representation of fluid and reaction dynamics. Parameter identification of the biokinetic model equations and model-based optimisation methods are hardly possible and are performed on ideally mixed systems (Wang et al. 2015).

Delafosse et. al (2014) developed a CFD-based compartment model by combining CFD simulations and compartment modelling to create a 3D network of zones models. Main advantage of the network of zones modelling approach is the lower computation time compared to CFD simulations. The models were validated by recreating an experimentally determined mixing time. The mixing time was experimentally determined at about 72 s using a conductivity tracer approach for a reactor equipped with two Ruston turbines stirring at 300 rpm and a working volume of 16.5 L. The network of zones model is dividing the bioreactor into a radial, axial and tangential direction, with two closed loops around each stirrer. The model is considering the main circulation flows caused by the stirrers and turbulent flows. Even without the incorporation of information from a CFD simulation, the network of zones model was able to simulate the mixing time of the bioreactor by adjusting the ratio of the turbulent flows. (Delafosse et al. 2014)

In the next step, Delafosse et. al (2014) calculated the flow rates between the compartments from the velocity fields of a CFD simulation. The CFD-based compartment model was able

to reproduce the concentration distribution during the mixing process within the bioreactor with good accuracy. However, only the liquid and the solid phase were simulated without taking a gas phase into account. The model is therefore restricted to processes which are anaerobic or aerated solely through surface aeration. Furthermore, even though the model is magnitudes faster than a CFD simulation, the model still required significant computational power since above 13000 zones were necessary to successfully recreate the mixing times, with less compartments the mixing time was underestimated (Delafosse et al. 2014; Delafosse et al. 2015)

Pigou et al. (2015) investigated how the bioreactor performance may be impacted by reactor heterogeneities through a compartmentalised modelling approach that divides the bioreactor into volume compartments (reactor cells). The work combines this compartmentalised model describing substrate and biomass distribution, a population balance model modifying growth rate, and a metabolic model computing the reaction rates of individual cells. The article compares the simulation results to experimental data from *E. coli* cultivations. Under specific large-scale conditions, the model can simulate the simultaneous production and consumption of acetate in different regions of the reactor. (Pigou and Morchain 2015)

These approaches show the potential of combining the simulation of heterogeneous conditions caused by non-ideal mixing of large-scale reactor with biokinetic equations. They focussed on recreating and understanding the development of heterogeneous conditions in bioreactors. A next step and alternative approach, which can be derived from this chapter, should be the combination of mathematical models considering heterogeneities and model-based control and optimisation methods. This combination could show that process control of processes under heterogeneous conditions can be systematically designed by model-based strategies.

2.6 Model-based bioprocess design

Mathematical models can be used effectively in bioprocess design and development of process operational strategies (Appl et al. 2021; Madni et al. 2019; Grieves 2015). So far, these model-based operational and design methods have rarely been used for processes in non-ideally mixed systems (e.g. scale-down models) and with models considering non-ideally mixed conditions. One main reason for this was the long computing time of the employed models. There are many possible applications for these model-based operational and design methods: optimising process control and conditions in scale-down reactors, model development, improvement and validation as well as assisting scale-down.

Models used for process control are often heuristic and based on purely mathematical relationships and correlations (e.g. fuzzy logic control or control with ANNs) and not based on mechanistic principles. They do not provide insight into the process, have limited predictive capabilities, and are limited to reproduce responses that are within the data of previously performed experiments (Glaessgen and Stargel 2012; Zobel-Roos et al. 2019).
The second group of mathematical process models utilised for the development of process operational strategies and process design are mechanistic models. The employed mechanistic process models must be able to quantitatively describe cause and effect relationships within the selected operating range including the optimisation horizon. Furthermore, the model must consider actuating control variables (e.g. feeding rates, aeration rates) as well as relevant process variables (e.g. biomass density, metabolite concentrations, product concentration). After identifying the model parameters, the model must be capable of replicating experimental data of all relevant process variables with a high quantitative and qualitative accuracy. Additionally, it is essential and required that the process models used for the application in model predictive control (MPC) are robust. Mainly unstructured models are utilised in MPCs for bioprocesses due to their simpler structure, better usability and faster computation times when compared to structured models (Craven et al. 2014; Dewasme et al. 2013; Hodge and Karim 2002; Frahm et al. 2002).
The possibility of multivariable operation provided by MPCs allows potentially all media components, feed streams, pH media, aeration rates and further actuating control quantities to be used for control, design and operation of the bioprocess (Galvanauskas et al. 2013; Birol et al. 2002; Gustafsson et al. 1995). Feeding rates, in particular, have shown a great potential for the process design through mathematical process models using model-based strategies (Craven et al. 2014; Witte 1996; Dewasme et al. 2013; Mears et al. 2017; Birol et al. 2002). The feeding rate has an impact on the entire process, in addition to directly influencing the concentration of the substrates, products and metabolites. Modifying feeding rates leads to changes in reactor volume, causes dilution effects and influences uptake rates, biomass growth, product formation as well as indirectly influences the exhaust gas composition (O_2 uptake and CO_2 production) (Villadsen et al. 2011).
A precise and well-defined design criterion is necessary for the successful implementation of model predictive operation and design methods. This criterion usually involves maximising or minimising specific process quantities, such as the maximisation of the biomass density (see chapter 8.3.2.3) (Kovárová-Kovar et al. 2000; Chang et al. 2016; Santos et al. 2012; Galvanauskas et al. 2013). Occasionally, the criterion considers the reduction of the difference between a predefined or determined course of available

experimental data and the process that is controlled, if similar cultivation courses to a preceding cultivation are desired. (Zhang and Lennox 2004). Multiple process objectives may also be mathematically combined into one single design criterion.

2.6.1 (Nonlinear) model predictive control

Process development and the scale-down procedure might be supported by (nonlinear) model-based control methods. These MPCs are using a mathematical process model for the optimisation and control. The control strategies are suitable for single- and multivariable control and can be utilised for the design and optimisation of complex bioprocesses affected by several different influencing factors. They are mainly employed for the development of operational process strategies in lab-scale bioreactors. In comparison, classical strategies without mathematical models such as PI-controlled temperature, pH, pressure, flow rate or concentration control are single loop controls and only allow for the modification of a limited number of control variables. MPCs with implemented linear or nonlinear mathematical models (NMPC) are more suitable for the application of predictive control and optimisation. (Laurí et al. 2014; Mears et al. 2017; Frahm et al. 2002)

In general, model predictive control and operational strategy development takes one of the three common forms:

- Open-Loop Control
- Feedback Control
- Adaptive Model Predictive Control (AMPC)

The three control methods listed above are applicable to both MPCs and NMPCs. Figure 13 presents block diagrams for Open-Loop control, feedback, and adaptive model predictive control.

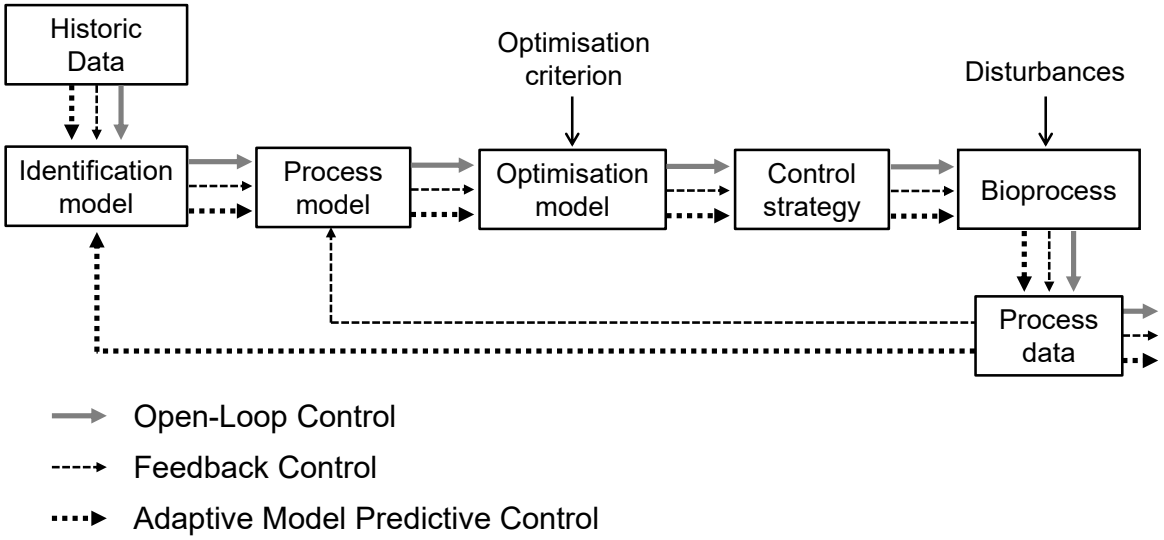


Figure 13: Process model and data utilisation within open loop control, feedback control and adaptive model predictive control.

The variants of model predictive control are described for the main application case, the control of volumetric feeding rate profiles of fed-batch processes (or in general: profiles of actuating variables including possible set-point trajectories). During open-loop control of fed-batch processes, the feeding rates profiles are calculated and optimised before the start of the process. These control profiles rely exclusively on available (historic or literature) data, a model derived from this data, the predetermined initial conditions, and the chosen optimisation criterion. The application of open-loop control is advantageous if the future process behaviour can be predicted with a high degree of accuracy and if major (process altering) disturbances are not expected during the bioprocess (Jenzsch et al. 2006).
Contrary to feedback control and AMPC, no data acquisition is required for process control. Low computation times of the mathematical model are not as critical as in other model-based control methods as all simulations are performed before the actual experiment. Therefore, it is easier and at first glance more appealing to introduce open-loop control and optimisation during early stages of process development (Wechselberger et al. 2012; Baeza 2017).
(Model predictive) Feedback control is industrially accepted and extensively applied in the (petro-)chemical industry (Forbes et al. 2015). Nevertheless, implementing MPCs in biotechnological processes is more challenging due to exceedingly complex and non-linear relationships present in biological systems. Feedback process control necessitates the use of process analytical techniques (PAT) for determining the current process state, which is then used as the new starting point for the subsequent optimisation computation. Feedback control can utilise the process data of the currently running process for recalculation and readjustment of the calculated operational strategies to achieve optimal results.

Adaptive MPCs (AMPCs) are feedback controllers that are specialised in utilising an identification and optimisation model during an ongoing process (Frahm et al. 2002; Witte 1996; Pörtner et al. 2017a). Prior to each optimisation of the process control function, the model parameters are identified based on current experimental data. The new identified parameter set is then transferred to the optimisation model, which may be identical or different from the parameter identification model. As a result, the optimised feeding strategy is always adjusted to the process based on new process data (Witte 1996). During cultivations, AMPCs and feedback control are repeated several times (Witte 1996; Frahm et al. 2002; Caramihai and Severi 2013; Baeza 2017).

Cultivation reproducibility, which ensured equal growth or production rate as well as similar durations of cultivation phases between two or more processes is vital for the application of open-loop control, whereas in processes controlled by feedback loop, the reproducibility can be increased by the feedback loop.

The Open-Loop Feedback Optimal strategy (OLFO-strategy) is a special form of an AMPC that utilises initial parameters and settings that are then transferred to the process model (Witte 1996). An optimised control function is calculated while the process is in progress and the model parameters are identified multiple times using available process data (offline and online data). The primary advantage of the OLFO-strategy is that it is not necessary to optimise or perfectly parameterise the process model beforehand, as the model parameters are identified during the process, and the control trajectories are updated frequently. Therefore, the OLFO-strategy (along with other AMPCs) is particularly appealing for early stages of process development (Witte 1996). During the process control (with OLFO strategy, AMPCs, or feedback control) the model either simulates the process in advance up to a predetermined time (moving horizon) (Frahm and Hass et al. 2003; Frahm and Lane et al. 2003; Levermann et al. 2020; Kim et al. 2023) or until the end of the process (finite horizon) (Witte 1996; Hedengren et al. 2014; Li 2015; Kim et al. 2023). In this work, the OLFO strategy was utilised to optimise a feeding strategy for an experiment in a scale-down model consisting of two connected stirred reactors with multiple goals: (1) acquisition of new process data, (2) improvement of the model parameters and (3) replication of process data of a pilot-scale experiment on the laboratory scale (chapter 9.3).

The identification of model parameters and determination of the process design may potentially utilise classical algorithms such as the Nelder-Mead algorithm. The parameter identification could theoretically even be performed manually (Nelder and Mead 1965; Kovárová-Kovar et al. 2000). The initial process design model could also be based on model-based design of experiments which is explained in more detail in chapter 2.6.2 (Nelder and Mead 1965; Kovárová-Kovar et al. 2000; Mandenius and Brundin 2008; Möller

et al. 2015; Möller and Pörtner 2017; Abt et al. 2018; Moser et al. 2021; Gassenmeier et al. 2022).

One possible case for the application of nonlinear model-based control is supporting the scale-down procedure. For an effective scale-down, the process on the small scale should produce similar results compared to the process on the large scale. The objective criterion could be formulated with the aim to reduce the difference between the existing data of the large scale and the controlled process. This method might have exciting potential for an effective scale-down procedure and was tried for the first time in this work. The strategy was performed as part of a study with industrial media and process conditions for a process with controlled heterogeneous conditions (chapter 9.3).

2.6.2 Model-based Design of Experiments

Model-based design of experiments (mDoE) is utilised in this work to develop open-loop operational strategies for bioprocess operation and design. The process development and design of bioprocesses is experimentally complex, time-consuming, and costly.

Process development with traditional design of experiments (DoE) requires an extensive number of experiments (Mandenius and Brundin 2008; Pörtner et al. 2017b). DoEs can be utilised to simultaneously optimise multiple process parameters and factors under consideration of combinatorial effects. However, selecting the right design space is a huge challenge and crucial. The choice of the design space (factor boundaries) can greatly impact the outcome of the DoE and narrowing the mDoE design space, in particular, is complex (Stosch and Willis 2017).

mDoE combines design of experiment strategies with process models to design, test and evaluate initial conditions and operation strategies before conducting experiments (Sercinoglu et al. 2011). The objective might involve maximising the amount of information gained for each experiment to improve and parameterise the model (Deflorian and Zaglauer 2011; Kreutz and Timmer 2009), and/or for optimising process operation (Möller and Pörtner 2017; Möller et al. 2019; Kuchemüller et al. 2020; Bayer et al. 2020; Moser et al. 2021).

The design of the process operation should also lead to robust processes in which minor changes to the control do not cause substantial changes to the process result. This can be achieved by incorporating uncertainty analyses into mDoEs. The uncertainty of model parameters, process factors and variables or control functions can be included in the calculation of the predicted process robustness by considering the predicted variability. (Candioti et al. 2014; Möller et al. 2019; Moser et al. 2021)

One advantage of mDoE is that the application leads to a reduction of experimental effort since less experiments are needed for the optimisation of the chosen factors. Furthermore, mDoE supports narrowing down the factor boundaries of the experimental design space (Moser et al. 2021; Gassenmeier et al. 2022; Möller et al. 2019).

A selected workflow of an mDoE (part of the mDoE-Toolbox) may consist of the following five steps (Möller et al. 2019; Moser et al. 2021):

1. Mathematical model formulation
2. Parameter identification (with e.g., historic or literature data)
3. Selection of an experimental design and boundaries
4. Simulation of each point in the design space
5. Evaluation of design (with e.g., response surface method)

The mDoE version of Möller et al. (2019) utilises classical design of experiments such as full factorial design or Box-Behnken design. These designs are applied since they are also used for experimental investigations. Higher number of experiments than full factorial design (n^3 with n = factors) are often not feasible to be performed experimentally. (Möller and Pörtner 2017; Möller et al. 2019; Moser et al. 2021)

However, model simulations are faster and can be automatically performed. For this reason, it might be advantageous to simulate highly detailed designs with thousands of experimental settings to achieve a higher resolution of the response surface.

In this work, the presented mDoE version of Möller et al. (2019) was modified to simulate a user-defined number of points within the selected boundaries of the design space. This modified version was already provided to and utilised by Appl (2023) for the design and optimisation of an enzymatic starch hydrolysis batch experiment. The factors were the initial concentrations of two enzymes, and the process objective was the maximisation of a profit function. The method and the utilised starch hydrolysis model was validated with four experiments. The experiment in the predicted optimum also showed the highest value of the profit function. (Appl 2023)

The modified mDoE was employed in this thesis to support simultaneous model development and process design (chapter 7.2) in order to minimise experimental effort. The resulting experimental data were used for model parameter identification in the area of optimal biomass growth. The modified mDoE was also applied for the design of experimental settings for a cultivation used for simultaneous model validation and process design. The objective of this experiment was the maximisation of the biomass density in a two-compartment scale-down model (chapter 8.3.2.3).

3 Research aims, strategies and thesis scope

Heterogeneous substrate and dissolved oxygen concentrations that might occur on the production scale probably cause performance differences between *S. cerevisiae* cultivations in different scales. These performance differences are still not fully understood (George et al. 1998; Larsson et al. 1996; Takors 2012). This work aims to contribute to finding new explanations for the different performances using a mechanistic, mathematical model capable of providing new insights into processes affected by heterogenous conditions combined with experiments in a physical scale-down system consisting of two connected stirred tank reactors.

Physical scale-down systems (or scale-down models, SDM) can be used to experimentally investigate the influence of heterogeneous conditions on the cultivation process by mimicking the conditions of large-scale processes (Oosterhuis and Kossen 1984; Neubauer and Junne 2016).

In the chosen physical scale-down system, the influence of heterogeneous DO and substrate concentrations for fed-batch *S. cerevisiae* cultivations will be investigated. These experimental investigations will be supported for the first time by model-based methods. Various studies have already demonstrated that model-based experimental design strategies are very powerful for developing informative experiments (Deflorian and Zaglauer 2011; Möller et al. 2019; Moser et al. 2021).

It can be derived from the state of the art that there is a need for mathematical models that are based on mechanistic principles and represent non-ideally mixed bioreactors and ideally mixed bioreactors simultaneously. In particular, there is a need for models that can be used for the model-based design of processes. To date, only simple, unstructured or population models have been incorporated into models capable of representing hydrodynamics and/or non-ideal mixing conditions. However, these models are limited in their ability to describe cultivations and/or have limited use for the application in model-based control methods often due to long computation times.

For this reason, a physical, experimental scale-down model (system) combined with a mathematical model-based approach is intended to be developed that can be used to improve the experimental design, while considering effects and their influence on cultivations caused by heterogeneous conditions. For the mathematical modelling of this work, a structured biokinetic model was extended by a reactor model consisting of a large number of connected zones (“Network of zones” model). The reactor model in combination with the chosen structured, biokinetic model should be able to consider the influence of the heterogeneous conditions within the SDM on cultivations with complex kinetics.

3.1 Research questions

Although there are already mathematical models and process optimisations regarding cultivations of *S. cerevisiae*, the performance in small reactors is often still better compared to reactors on the production scale. To investigate this issue, two research questions have been formulated, based on the state of the art, which will be answered in the thesis:

1. Is a structured, mechanistic mathematical model, calculating several interconnected ideally mixed zones, able to approximate the flow patterns in stirred tank reactors and describe the effects caused by non-ideal mixing in biotechnological processes with complex kinetics?
2. How can feeding strategies for biotechnological processes in non-ideally mixed reactors be systematically designed?

The process control and the heterogeneous conditions within bioreactors have an influence on the performance of cultivations (Neubauer and Junne 2016; Heins and Weuster-Botz 2018; Sweere and Matla et al. 1988). Therefore, a targeted design of process control is required. In this work, yeast cultivations are designed while heterogeneous glucose (or sucrose) and DO concentrations as well as the interaction with ethanol production/uptake are influencing the cultivation. The designed operational strategies could be used for process development and optimisation as well as for model development and the assistance of the scale-down procedure.

3.2 Thesis scope and workflow

To be able to answer the research questions, two further research objectives are defined:

1. A laboratory bioreactor system consisting of two connected stirred tank reactors shall be developed. This system shall be used to investigate cultivations of *S. cerevisiae* that are influenced by heterogeneities.
2. A modelling and software system needs to be developed and tested to investigate the research questions.

The workflow of this thesis and the employed model-based strategies and software tools are presented in Figure 14.

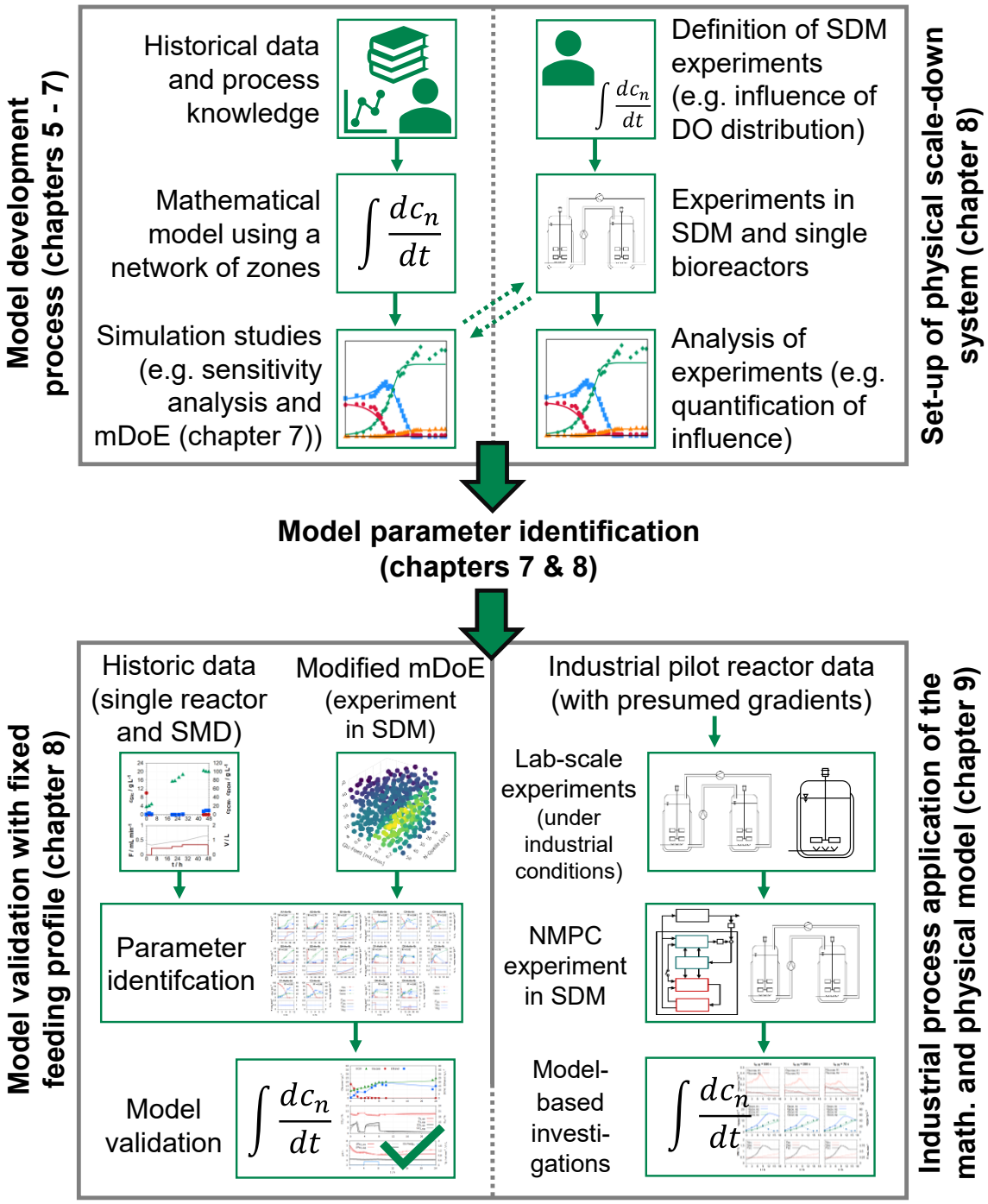


Figure 14: Simplified workflow of this thesis and employed strategies.

For the experimental investigation of the effects caused by non-ideal mixing, a physical scale-down model comprising of two reactors is developed and established. The requirement for this scale-down model is that it should be able to simulate the basic scale effects of production reactors. The heterogenous conditions of large-scale production systems can be simulated by adapting the mixing time, the division of the reactor system into (two) distinct zones with varying concentration ratios (e.g. oxygen or substrate distribution) and the positioning of the feed in different areas within the system.

A mathematical model is developed which utilises an existing structured biokinetic model. The biokinetic model is integrated into a reactor model using a network of zones model (chapter 5). The model is used in a simulation study (chapter 6) and is employed within a modified mDoE to design feeding profiles for model development and process optimisation in an ideally mixed bioreactor (chapter 7).

For the first time, an experimental investigation of the influence of heterogeneous dissolved oxygen and glucose concentrations supported by a mechanistic, mathematical models on *S. cerevisiae* cultivations is performed (chapter 8). The modified mDoE is also used for the determination of experimental settings used for model validation and process development (chapter 8.3).

The physical scale-down model and the mathematical model are utilised in an industrially relevant use case with conditions and media of an industrial process. The media and feeding strategy of the industrial process are transferred to and investigated on the physical scale-down model and ideally mixed bioreactors. An NMPC is used as a method to reproduce data of the pilot scale process and is simultaneously utilised for data acquisition. Finally, model-assisted investigations are conducted to identify possible explanations for the cause of the process differences between the scales (chapter 9).

4 Material and methods

This chapter presents the material and methods used for the cultivations of yeast in single bioreactors and the developed scale-down model (SDM) consisting of two connected bioreactors. Additionally, it outlines the analytical and in-silico methods employed for the experimental data acquisition, the mechanistic modelling, and the development of model-based strategies for process control and optimisation.

The materials utilised are presented in chapter 4.1, the analytical methods are described in chapter 4.2, the experimental methods (applied reactors and the SDM, the determination of mixing times and process performance indicators) are outlined in chapter 4.3, the software methods utilised are described in chapter 4.4, the mathematical model used as a basis for the model structure is explained in chapter 4.5 and the computer assisted methods (in-silico methods) for the evaluation of the simulations are detailed in chapter 4.6.

4.1 Materials

The materials needed for the cultivations of *S. cerevisiae* are carbon substrates (glucose-monohydrate and sucrose) and nitrogen substrates (peptone and yeast extract) as well as pH adjustment media and ethanol for analytical purposes (Table 2).

Throughout all experiments a genetically unmodified commercial organic dry yeast was used (Agrano GmbH & Co. KG, Germany). Glucose (cultivations in chapter 7 and 8) and sucrose (chapter 9) were utilised as the carbon substrates, the nitrogen source is a combination of yeast extract and peptone (all Carl Roth GmbH + Co. KG, Germany). In the experiments of the industrial application process (chapter 9) an industrial supplier provided an industrially utilised yeast extract. Potassium dihydrogen phosphate (Merck KGaA, Germany) and potassium hydroxide (Carl Roth) were used in selected cultivations (chapter 7) for adjustments of the pH.

Table 2: Materials used for experimental investigations.

Material	Specification	Manufacturer
S. cerevisiae	Product number: AG31000086 42100263	Agrano GmbH & Co. KG, Riegel am Kaiserstuhl, Germany
Glucose-Monohydrate	CAS No.: 77938-63-7	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Peptone	CAS No. 91079-46-8	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Yeast extract	CAS No. 8013-01-2	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Sucrose	Bio-Quality	Provided by an industrial supplier
Ethanol	CAS-No.: 64-17-5	Merck KGaA, Darmstadt, Germany
Potassium dihydrogen phosphate	CAS No. 7778-77-0	Merck KGaA, Darmstadt, Germany
Potassium hydroxide	CAS No. 1310-58-3	Carl Roth GmbH + Co. KG, Karlsruhe, Germany

4.2 Analytical methods

Glucose, ethanol, and sucrose concentrations were determined with enzymatic UV test-kits “Enzytec” (R-Biopharm AG, Germany) (see Table 3). The tests were performed according to the enclosed instructions of the test-kits and performed with the spectrophotometer UVmini-1240 (Shimadzu Corp., Japan) at wavelengths of 340 nm and 365 nm using disposable semi-micro cuvettes. Samples that exceeded 1 Abs (absorption unit of the spectrophotometer) were diluted appropriately to achieve concentrations in the validation range. The methods were regularly validated by determinations of standard samples containing 0.5 g L^{-1} glucose or 0.06 g L^{-1} ethanol (employed in all experiments except D1 – D5).

Table 3: Materials and methods used for the analyses in the performed experiments.

Analytical Method	Specification	Manufacturer
Enzymatic assays		
Enzytec Liquid D-Glucose	Art. No. E8140; Cat. No 10 716 251 035	R-Biopharm AG, Darmstadt, Germany
Enzytec Liquid Maltose/Sucrose/D-Glucose	Art. No. E8170	R-Biopharm AG, Darmstadt, Germany
Enzytec Liquid Ethanol	Art. No. E8340	R-Biopharm AG, Darmstadt, Germany
Dry cell weight		
Moisture Analyzer	MA45 (Infrared moisture analyse)	Sartorius AG, Göttingen, Germany
Cellulose Acetate Filter	0.45 mm (Art. No. 512-4316)	VWR, Radnor, Pennsylvania, United States
UV measurements		
UV-VIS Spectrophotometer	UVmini-1240	Shimadzu Corp., Kyōto, Japan
Disposable Cuvettes	10 mm, 300 - 900nm, 1.5 mL (Cat. No. 634-0678)	VWR, Radnor, Pennsylvania, United States
HPLC		
HPLC system	Nexera XR	Shimadzu Corp., Kyōto, Japan
HPLC column	Rezex ROA-Organic Acid H+ column (300 × 7.8 mm)	Phenomenex, Washington D.C., United States
Mobile Phase	0.005 N sulfuric acid	VWR, Radnor, Pennsylvania, United States
Further equipment		
pH/mV meter	FiveEasy	Mettler Toledo, Columbus, Ohio, United States
Thermometer	testo 110	Testo SE & Co. KGaA, Titisee-Neustadt, Germany
Universal oven	30 - 1060	Memmert GmbH + Co. KG, Schwabach, Germany
Magnetic stirrer with heating plate	RH basic 2	IKA-Werke GmbH & CO. KG, Staufen, Germany

In experiments D1 - D5 the glucose and ethanol concentrations were quantified and determined using high pressure liquid chromatography (HPLC), employing a Rezex ROA column (300 × 7.8 mm, Phenomenex, USA) with 0.005 N sulfuric acid as the aqueous phase, following the manufacturer's protocol. The sample volume was 10 µL. The calibration curve was generated with six different solutions of known concentrations within the expected concentration ranges. The measured retention times for ethanol were 24.3 min ± 1.7 min and for glucose 12.1 min ± 0.1 min. The enzymatic assay and HPLC results were expressed as partial densities / mass concentrations in g L^{-1} (equal to the SI-Unit kg m^{-3}). The HPLC method encounters difficulties in detecting low concentrations (e.g., < 0.5 g L^{-1} for glucose) for cultivation samples. These difficulties arise due to an unsmooth baseline and consequently imprecise peak detection. Therefore, cultivation samples with a glucose concentration of 0 g L^{-1} (determined by UV test-kits) usually produced readings above 0.2 g L^{-1}, when determined with the HPLC method. Only the ethanol and glucose concentrations of experiments D1 - D5 were determined by the HPLC method; the concentrations of the other experiments were determined by using the enzymatic test-kits.
To determine the dry cell weight (DCW) of a fermentation sample, a sample of known volume was filtered through a cellulose acetate filter (0.45 mm, VWR, USA). The resulting retentate was washed with at least double the sample volume of demineralised water. The weight of the retentate was determined after drying in a moisture analyser (MA45, Sartorius, Germany). The cell weight density was calculated using equation (4.1).

$$c_{DCW} = \frac{m_{dried\ retentate}}{V_{sample}} \tag{4.1}$$

This method was validated by weighing a sample of 100 mL after drying in a drying oven (M30 – 1060, Memmert GmbH + Co. KG, Germany).

4.3 Experimental methods

Three different bioreactor systems were utilised for the yeast cultivation experiments (see Table 4 and Table 5), including the scale-down model (SDM), which consists of two connected MDX bioreactors (MDX Biotechnik International GmbH, Germany) (see chapter 4.3.2). The other two bioreactors are a 1 L MDX bioreactor and a 2 L Biostat B (chapter 4.3.1). Additional experimental methods employed in this thesis include the experimental and mathematical determination of mixing times (chapter 4.3.3) and the calculation of process performance indicators.

Table 4: Bioreactor equipment of the physical scale-down model and the MDX system

Equipment Name	Specification	Manufacturer
MDX bioreactor	STR, V_{max} = 1 L	MDX Biotechnik International GmbH, Nörten-Hardenberg, Germany
DO probe	VisiFerm DO ECS 225 H0	Hamilton, Reno, Nevada, United States
pH probe	EasyFerm Plus PHI S8 225	Hamilton, Reno, Nevada, United States
Thermostat	Huber CC-104 A Pilot One	Peter Huber Kältemaschinenbau SE, Offenburg, Germany
Syringe feed pump	LSP 01-1A	Longer Precision Pump Co., Ltd., Baoding, Hebei, China
Scale	Entris 5201 ACX	Sartorius AG, Göttingen, Germany
Circulation Pump	MDX TC/200	MDX Biotechnik International GmbH, Nörten-Hardenberg, Germany

Table 5: Bioreactor equipment of the Biostat B bioreactor system

Equipment Name	Specification	Manufacturer
Biostat B	STR, V_{max} = 2 L	B. Braun Biotech International, Melsungen, Germany
DO probe	InPro 6800	Mettler Toledo, Columbus, Ohio, United States
pH probe	405-DPAS-SC-K8S	Mettler Toledo, Columbus, Ohio, United States

Each bioreactor system used the identical exhaust gas measurement device, the Sick Sidor SN 762860 (Sick, Germany). This exhaust gas analyser measures the oxygen and carbon

dioxide fractions in the exhaust gas stream from which the respiratory quotient (RQ) is calculated as the quotient of the produced carbon dioxide divided by the consumed oxygen.

$$RQ = \frac{n_{CO_2,produced}[mol]}{n_{O_2,consumed}[mol]} \quad (4.1)$$

4.3.1 Single reactor systems

Two different bioreactor types were employed for the experiments in the single reactor systems. Two identical Biostat B systems with maximum volumes of 2 L were used as well as two identical MDX biotech reactors (MDX Biotechnik International GmbH, Germany) with maximum volumes of 1 L (see Figure 15).

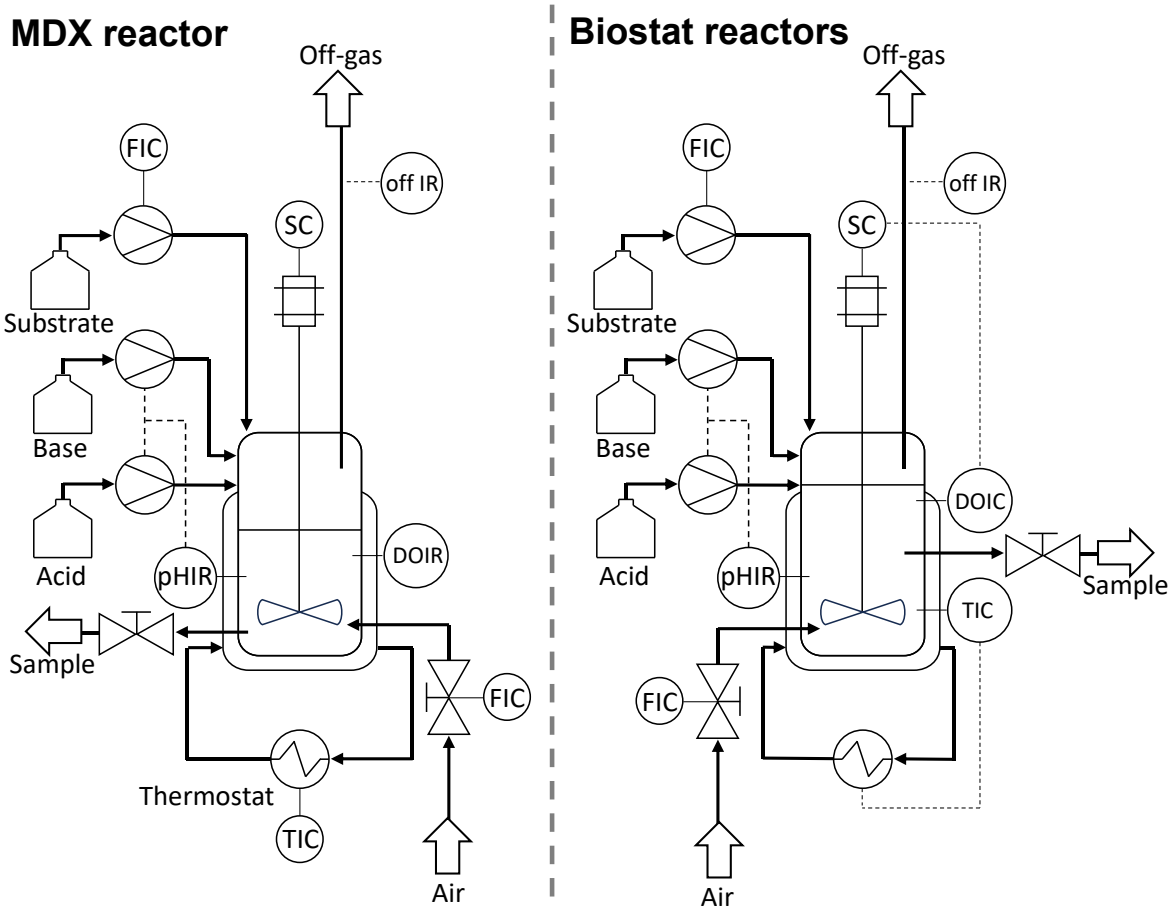


Figure 15: Single bioreactors and periphery. Left side MDX bioreactors (V_{max} = 1 L), right side Biostat B (V_{max} = 2 L)

The MDX reactor with a maximum volume of 1 L served as the smallest stirred tank bioreactor used in this work. This reactor type was used for the physical scale-down model constructed from two connected MDX bioreactors (for further information, see section 4.3.2). The reactors featured a hollow cylindrical built-in glass installation with the stirrer inside to improve mixing. Incorporated in the built-in cylinder are circular holes for liquid

passage, spaced regularly and 1 cm in diameter. The propeller stirrer with three blades is located at the lower end of the installation approximately 1 cm above the reactor bottom.
An in situ amperometric pH probe (EasyFerm Plus PHI S8 225, Hamilton, US in MDX reactors) was used for pH measurements. The pH could be controlled automatically. The dissolved oxygen (DO) was measured with an optical DO probe (VisiFerm DO ECS 225 H0, Hamilton, US) and was manually controlled by adjusting the stirrer speed and/or aeration rate. The medium temperature in the MDX bioreactor was controlled via an external water bath (Huber CC-104 A Pilot One) pumping cooling/heating water through the jacket of the bioreactor vessel.
The Biostat B has a maximum working volume of 2 L and features a glass vessel with a Rushton turbine stirrer on two levels. The pH value of the medium was measured with amperometric pH-probes (405-DPAS-SC-K8S, Mettler Toledo, US). The DO was also measured amperometrically using the InPro 6800 (Mettler Toledo, US). Additionally, the medium temperature was measured utilising a PT100 connected to the Biostat digital control units (DCUs). The temperature was controlled automatically by adjusting the water temperature and the volume flow in the cooling circuit of the reactors.

4.3.2 Physical scale-down model

The state of the art presented various physical scale-down models that could be considered for the experimental part of this thesis. The experimental work (chapters 8 and 9), that will be presented, has been performed in a system consisting of two coupled stirred tank reactors.
In the first section (chapter 4.3.2.1) the choice of the STR-STR configuration is explained and in the second section (chapter 4.3.2.2) the structure and specifications of the system are described.

4.3.2.1 Selection of the physical model

The selection of the STR-STR system was based on the two research questions posed in chapter 3.1. The experimental system must represent heterogeneous conditions (regarding substrates and dissolved oxygen concentration), and it must be possible to simulate the system with a structured mathematical model, which will be used for model-based process design of cultivations in the selected physical scale-down system.
The STR-STR system offers the advantage over other scale-down systems (STR-PFR or single STRs with oscillations etc.) that the aerobic and anaerobic conditions can be set accurately in the two individual reactors. The oxygen gradients (e.g. in the tubular reactor section) in the other scale-down systems with plug flow reactors, single reactors with oscillating process conditions or single stirred tank reactors with built-in installations that

prolong the mixing time are more difficult to measure and characterise. This also applies to heterogeneous substrate concentrations. Furthermore, strictly speaking, lab-scale reactor systems with oscillations are not truly heterogeneous but have alternating homogeneous conditions and were therefore excluded. The mixing time is also easily adjustable in STR-STR systems by modifying the volumetric flow rate between the reactors.

The key objectives (defined in chapter 3.2) were the simulation of cultivations with non-ideal mixing conditions (research question #1) with a structured mathematical model and the model-based process design of processes with heterogenous conditions (research question #2). Therefore, the compatibility of the physical scale-down model with the mathematical model must also be considered for the selection of the physical model. The advantage of the STR-STR system over the other systems is that the two individual reactors can be assumed to be ideally mixed. This simplifies the simulation compared to systems with tubular reactors, as these can only be simulated with more effort (e.g. via partial differential equations or stirred tank cascades). Therefore, the simulation of tubular flow reactors usually requires smaller step sizes, resulting in more calculation iterations and longer simulation times, which would make model parameterisation more time-consuming and model-based process design difficult. For systems with built-in components that extend the mixing time, it is difficult to simulate the resulting heterogeneities.

For these reasons, the chosen experimental approach of two coupled stirred tank reactors was ultimately selected, which is presented in the next subchapter.

4.3.2.2 Specifications of the physical scale-down model

The experimental investigation of the heterogeneous dissolved oxygen and substrate concentrations in a physical scale-down model were conducted in an STR-STR. This physical scale-down model was developed and consists of two 1 L stirred tank reactors (MDX Biotechnik International GmbH, Germany) (Figure 16). A description of the single bioreactors can be found in section 4.3.1

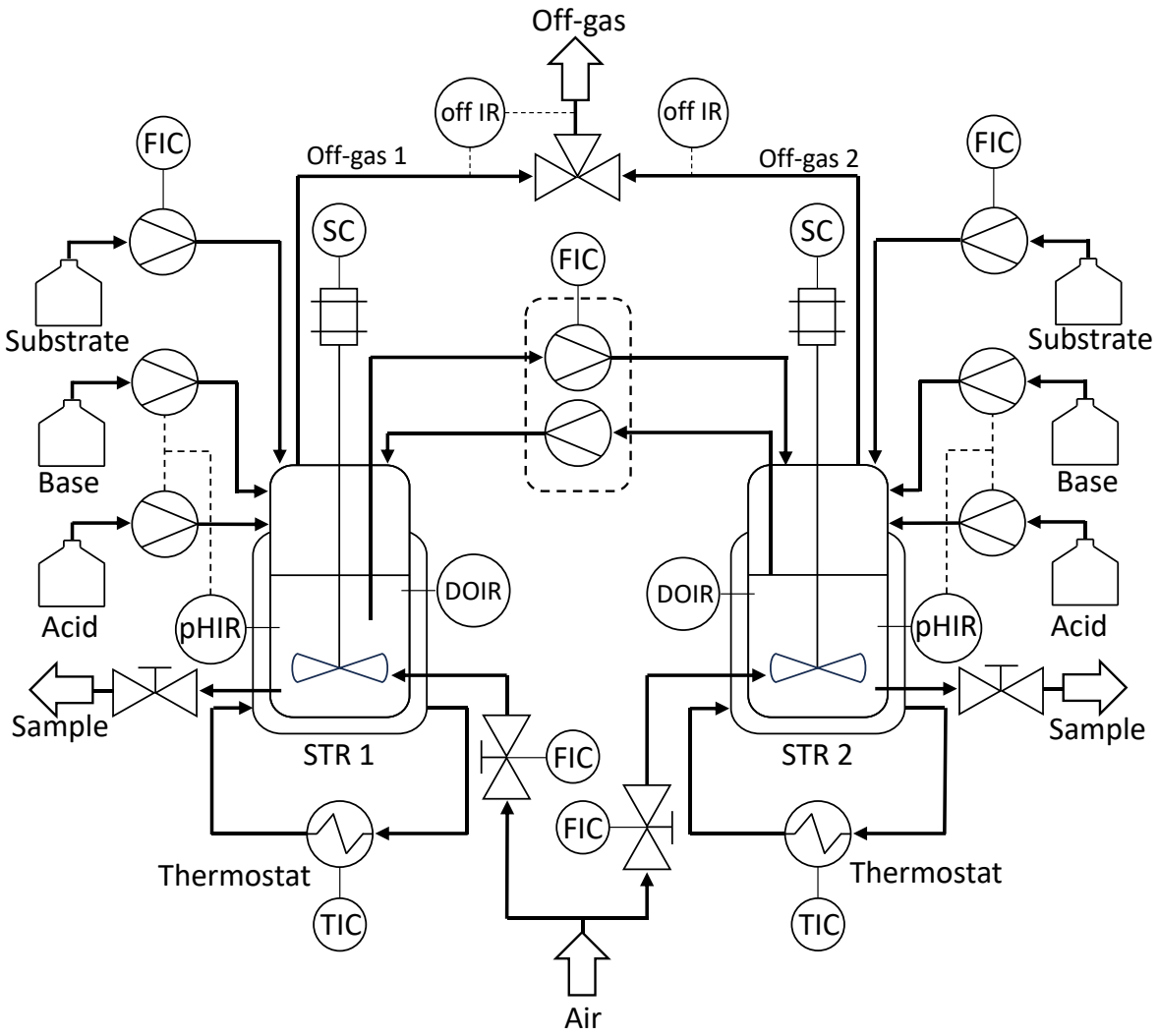


Figure 16: Physical scale-down model consisting of two connected 1 L stirred tank reactors (MDX Biotechnik International GmbH, Germany).

The feed was added into one of the STRs by a single-channel syringe pump (LSP 01-1A, Longer Precision Pump Co., Ltd., China).

The flow between the STRs was created via a two-channel peristaltic pump (TC/200, MDX Biotechnik International GmbH, Germany) with a maximum volumetric flow rate of 2 L min^{-1} per channel. The circulation pump provided a user-defined volumetric flow rate from STR 1 to STR 2 through the first channel of the pump. This first channel is referred to as “master channel” in the following. The second channel provided the flow from STR 2 to STR 1. The pressure on the piping was adjusted in order to provide a higher volumetric flow rate through the second channel. The inlet of the second channel was at a fixed height (at the liquid surface level) in STR 2 to ensure a constant volume during the SDM experiments. The inlet of the primary channel was located within the reactor medium during the experiments. This setup was able to maintain a constant volume in STR 2 without requiring any further volume and volumetric flow control. Calibration was only required for the volumetric flow rate

through the main pump channel (primary channel). In addition, this rate could be freely selected within the specified circulator parameters.

During the cultivations, the pH value in both reactors was measured in situ using amperometric pH probes (EasyFerm Plus PHI S8 225, Hamilton, US in MDX-Biotech reactors). The dissolved oxygen concentration (DO) in the SDM experiments was determined using an optical dissolved oxygen (DO) probe (VisiFerm DO ECS 225 H0, Hamilton, US). The DO was controlled manually by adjusting the stirrer speed and/or aeration rate. The temperature in the SDM was controlled via an external water bath (Huber CC-104 A Pilot One).

4.3.3 Determination of mixing times

The mixing time was determined experimentally by measuring the time course of the pH after the addition of pH media bolus feeds (chapter 4.3.3.1) or calculated through mathematical correlations using dimensionless numbers (chapter 4.3.3.2). The experimental method for determining the mixing time was chosen primarily because it can be easily implemented with the existing equipment.

4.3.3.1 Experimental determination of the mixing time

Mixing times were determined by pH measurements after the addition of an acid (0.1 M HCl or addition of 20% phosphoric acid) or a base (0.1 M NaOH or addition of 20% potassium hydroxide solution). The pH value was measured experimentally in the SDM from the start of mixing (y_0) until an equilibrium pH value y_{equil} is reached. Prior descriptions of this measurement method exist (Ascanio 2015; Poulsen and Iversen 1997). Here, y_1 is the pH value at which 90% of the total shift to equilibrium occurred.

$$y_1 = 0.9 \cdot (|y_{equil} - y_0|) \tag{4.2}$$

The mixing time t_{M90} is calculated as the time difference between the starting time t_0 of the circulation pump (or addition of pH media) and the time t_1 at y_1.

$$t_0 = t(y_0)\ and\ t_1 = t(y_1) \tag{4.3}$$

$$t_{M90} = t_1 - t_0 \tag{4.4}$$

The mixing time t_{M95} was also used in this work. t_{M95} describes the mixing time where 95% of the change to equilibrium occurred and is calculated analogously to t_{M90}.

Figure 17 displays a visualisation of the calculation of mixing times in a two-compartment system using the described method.

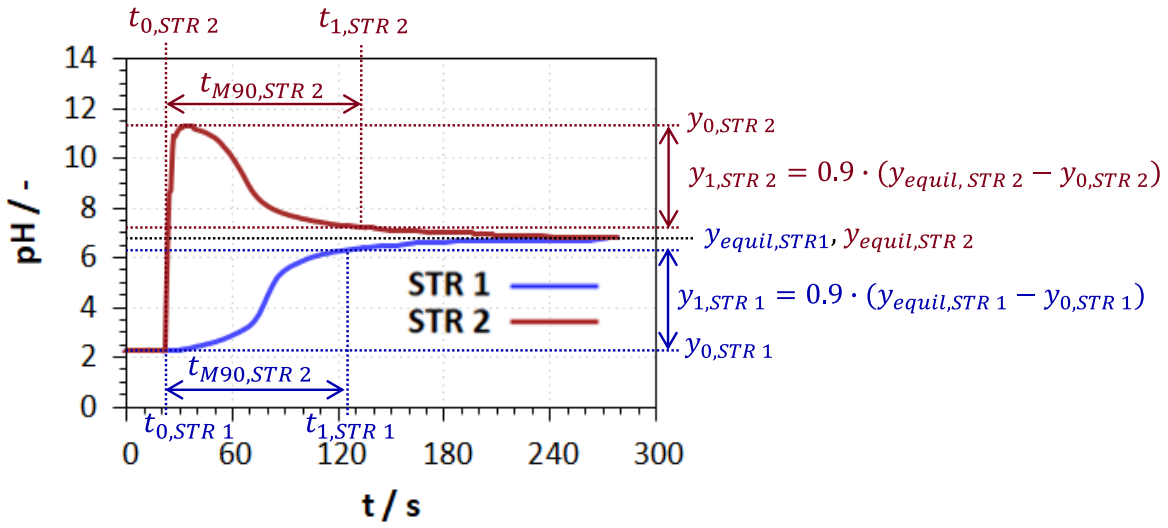


Figure 17: Mixing time determination for experimental data of a scale-down system consisting of two connected bioreactors.

The mixing times were determined at least three times to ensure reliable results (three replicates). To achieve the desired mixing time, the volumetric flow rate of the circulation flow was adjusted.

4.3.3.2 Theoretical calculation of the mixing time

The estimation of the mixing time t_M using dimensionless numbers and mathematical correlations is the second method used for determining the mixing time in systems. Nienow (2010) gives an approximation of the mixing time t_M under turbulent conditions (Nienow 2010).

$$t_M = 6 \cdot D \cdot \bar{\varepsilon}_T^{-\frac{1}{3}} \cdot (d/D)^{-\frac{1}{3}} \cdot (H_L/D)^{2.5} \tag{4.5}$$

The parameter d is the diameter of the stirrer, D the diameter of the reactor. H_L is height of the liquid medium in the reactor. For the calculation of the specific average energy dissipation $\bar{\varepsilon}_T$, the following equation is used.

$$\bar{\varepsilon}_T = \frac{P}{V \cdot \rho} \tag{4.6}$$

The medium has a volume V and the density ρ. The power input P can be calculated by rearranging the definition of the Newton number Ne (Chmiel et al. 2018).

$$P = Ne \cdot \rho \cdot N^3 \cdot d^5 \tag{4.7}$$

The speed of the stirrer is denoted as N. The Newton number for turbulent processes may be estimated using the following mathematical correlation proposed by Henzler (1982). (Henzler 1982).

$$Ne = z \cdot \frac{Ne_0 + 187 \cdot Q \cdot (Fr)^{-0.32} \cdot (d/D)^{1.53} - 4.6 \cdot Q^{1.25}}{1 + 136 \cdot Q \cdot (d/D)^{1.14}} \quad (4.8)$$

z represents the number of stirrer levels with $Ne_0 \approx 4.9$. Q refers to the aeration number and Fr is the Froude number. The aeration number is calculated as the gaseous volumetric flow rate divided by the product of the stirrer diameter d to the power of three and the stirrer speed.

$$Q = \frac{q_G}{d^3 \cdot N} \quad (4.9)$$

Equation (4.8) is only valid exclusively for turbulent conditions where the Reynolds number Re exceeds 10^4 and when the specified condition is fulfilled.

$$Fr < 0.07 \cdot (D/d)^3 \quad (4.10)$$

The Froude number is determined according to following equation and utilised to satisfy the previous condition. Furthermore, the Froude number is used for the calculation of the Newton number.

$$Fr = \frac{d \cdot N^2}{g} \quad (4.11)$$

The factor g is the acceleration due to gravity also known as the gravitational field strength. The Reynolds number is the product of stirrer diameter squared, stirrer speed and the medium density divided by the dynamic viscosity.

$$Re = \frac{d^2 \cdot N \cdot \rho}{\eta} \quad (4.12)$$

The dynamic viscosity η can be approximated to 1 mPa s according to van't Riet and Tramper (1991) (van't Riet and Tramper 1991). Furthermore, it is assumed that the dynamic viscosity remains unaffected by the cell density and cultivation state.

Additionally, equation (4.8) is only valid when the stirrer levels do not influence each other ($\Delta h/D > 0.75$, where Δh represents the distance between two stirrer levels) and when the quotient of the stirrer diameter divided by the reactor diameter d/D is between 0.2 and 0.42 (Henzler 1982). If these prerequisites are fulfilled, the Newton number specified in equation (4.8) can be calculated and utilised to approximate the power input and finally the mixing time t_M.

This method was only utilised when an experimentally performed measurement of the mixing time was not available or possible (chapter 9.1).

4.3.4 Process performance indicators.

Process performance indicators are calculated and used for evaluation of the individual cultivations. **Yield coefficients** in this work are describing a produced product divided by a

consumed substrate. The yield coefficient $Y_{X/S}$ ($Y_{X/SC}$ or $Y_{X/SN}$) is the mass ratio of produced biomass ($m_{X,produced}$) to consumed substrate ($m_{S,consumed}$).

$$Y_{X/S} = \frac{m_{X,produced}}{m_{S,consumed}} \tag{4.13}$$

$Y_{P/SC}$ describes the mass ratio of produced product $m_{P,produced}$ (ethanol) per consumed substrate $m_{S,consumed}$ (e.g. glucose).

$$Y_{P/SC} = \frac{m_{P,produced}}{m_{SC,consumed}} \tag{4.14}$$

Ethanol can also be consumed by yeasts for biomass growth. $Y_{X/P}$ is the yield coefficient of biomass ($m_{X,produced}$) per ethanol consumed ($m_{P,consumed}$).

$$Y_{X/P} = \frac{m_{X,produced}}{m_{P,consumed}} \tag{4.15}$$

For example, the maximum theoretically possible yield coefficient for glucose to ethanol is derived from the stoichiometric equation:

$$C_6H_{12}O_6 \rightarrow 2 \cdot C_2H_6O + 2 \cdot CO_2 \tag{4.16}$$

The yield coefficient $Y_{P/SC}$ for the stoichiometric equations can be calculated using the molar masses of glucose (M_{Glc} = 180 g mol^{-1}) and ethanol (M_{EtOH} = 46 g mol^{-1}).

$$Y_{P/SC} = \frac{2 \cdot M_{EtOH}}{M_{Glc}} = \frac{2 \cdot 46\ g\ mol^{-1}}{180\ g\ mol^{-1}} = 0.511\,\frac{kg}{kg} \tag{4.17}$$

Additional performance indicators are the **growth, production, or uptake rates**, which are calculated from experimental data. The growth rate μ_X is calculated with the produced biomass $m_{X,produced}$ observed over a time-period divided by the product of the average biomass $\overline{m_X}$ multiplied by the time difference Δt of the same time-period.

$$\mu_X = \frac{m_{X,produced}}{\overline{m_X} \cdot \Delta t} \tag{4.18}$$

The overall average biomass $\overline{m_X}$ is calculated by the sum of the average biomass for each interval $\overline{m_{X,i}}$ between samplings weighted by the quotient of time interval Δt_i divided by the complete total time interval Δt considered.

$$\overline{m_X} = \sum_{i=1}^{n} \overline{m_{X,i}} \left(\frac{\Delta t_i}{\Delta t}\right) \tag{4.19}$$

This was done to account for different time intervals between samples and non-linear growth. The rate of substrate uptake r_S as well as rate of product formation r_P are calculated using the used substrate $m_{S,used}$ and produced product $m_{P,produced}$ respectively.

$$r_S = \frac{m_{S,used}}{\overline{m_X} \cdot \Delta t} \tag{4.20}$$

$$r_P = \frac{m_{P,produced}}{\overline{m_X} \cdot \Delta t} \tag{4.21}$$

During many processes such as yeast cultivations the product (ethanol) can also be consumed, resulting in a calculable product uptake rate. Volume changes such as dilution, sampling and substrate input through feeding are considered in the calculation of the produced biomass and product and the consumed substrate. The equation for the produced biomass $m_{X,produced}$ is modified to:

$$m_{X,produced} = c_{X,end} \cdot V_{X,end} - c_{X,ini} \cdot V_{ini} + \sum_{i=1}^{n} c_{X,sample,i} \cdot V_{sample,i} \tag{4.22}$$

For the calculation of $m_{X,produced}$, the initial, inoculated biomass $m_{X,ini}$ (product of the initial biomass density $c_{X,ini}$ and initial volume V_{ini}) is subtracted from the final biomass $m_{X,end}$ (product of the final biomass density $c_{X,end}$ and the final volume V_{end}). The total biomass which is removed from the bioreactor through sampling needs to be considered for the calculation of $m_{X,produced}$. The biomass in samples $m_{X,sample}$, is the sum of the biomass density in each sample $c_{X,sample,i}$ multiplied by the corresponding sample volume $V_{sample,i}$.

The equation for the consumed substrate $m_{S,consumed}$ is extended by a term considering the additional substrate $m_{S,Feed}$ added by a substrate feed. $m_{S,Feed}$ is calculated by multiplying the substrate concentration in the feed $c_{S,Feed}$ with the feed volume V_{feed}, assuming the concentration in the feed is constant.

$$m_{S,consumed} = c_{S,ini} \cdot V_{ini} + c_{S,Feed} \cdot V_{feed} - c_{S,end} \cdot V_{end} - \sum_{i=1}^{n} c_{S,sample,i} \cdot V_{sample,i} \tag{4.23}$$

Here, the final substrate $m_{S,end}$ is subtracted from the initial substrate $m_{S,ini}$ and the substrate in the feed $m_{S,feed}$. From this difference, the combined substrate in all samples $m_{S,sample}$ is subtracted.

The modified equation for the mass of the total produced product $m_{P,produced}$ is generated analogous to the equation for the produced biomass.

$$m_{P,produced} = c_{P,end} \cdot V_{end} - c_{P,ini} \cdot V_{ini} + \sum_{i=1}^{n} c_{P,sample,i} \cdot V_{sample,i} \tag{4.24}$$

The initial product $m_{P,ini}$ (product of the initial product concentration $c_{P,ini}$ and V_{ini}) is subtracted from the final product $m_{P,end}$ (product of the final product concentration $c_{P,end}$ and V_{end}). The total product in all samples $m_{X,sample}$, which is the sum of the product concentration in each sample $c_{P,sample,i}$ multiplied by $V_{sample,i}$ is added to the previous difference.

4.4 Software

For this work and the tests of the research hypotheses, the application of different software tools was required, ranging from modelling and program development environments, programming languages to visualisation programs and data handling tools (see Table 6).

Table 6: Software (and version) used for process simulation, experimental evaluation, data processing, data visualisation etc. as well as their developers (Windows and Excel were updated automatically).

Software	Version	Usage	Developer / Publisher
C-eStIM	2021-11	Model code	Florian Kuhnen, VC Hass, City University of Applied Sciences, Bremen, Germany
Excel 365	-	Data processing, Data visualisation, Spreadsheet calculations	Microsoft Corporation
gnuplot	5.4 patch level 8	Data visualisation	Thomas Williams, Colin Kelley
Notepad ++	8.4.8	Text Editor / Code Editor	Don Ho
Python	3.9.12	Data processing, Model stability tests, data visualisation	Python Software Foundation, Guido van Rossum
R	4.2.2	Data processing, Parameter identification, mDoE, NMPC, model generation tool, data visualisation	R Core Team
R-Studio	2022.07.2 Build 576	IDE for R	Posit, PBC
Spyder	5.1.5	Integrated development environment (IDE) for Python	Carlos Córdoba, Pierre Raybaut
Visual Studio 2019	16.11.8	Installing C-eStIM	Microsoft Corporation
Windows 10 & 11	-	Operating system (OS)	Microsoft Corporation

C-eStIM is a simulation program written in C++ for developing models for dynamic systems (Hass et al. 2005; Kuhnen 2008). In principle, any system described by coupled ordinary

differential equations (ODE) can be simulated. In this thesis, C-eStIM is utilised to create the mathematical model code for the simulation of yeast cultivations under non-ideal mixing conditions. The base mathematical model is described in the following chapter 4.5 and adaptations to the model are presented in chapter 5.2. C-eStIM contains libraries (Table 20 in Appendix) for compiling the model code into an executable model program, identifying parameters, and more.

The model-code is written into several >*.cpp< files, which are compiled by a C++ compiler used by C-eStIM into an executable program (file extension >*.exe< or >*.so< for Unix systems). The model settings, parameters, and initial values are written into a configuration file with the filename extensions >*.cfg<. The profiles (trajectories) of control variables are defined in control files (>*.xct<). The model is initiated through the command line, and it retrieves the chosen configuration file and all control profiles. The solutions of the nonlinear differential equation of the mathematical model are approximated with the Runge-Kutta method (Runge 1895). The results of each ODE and selected variables are saved into separate text files with the file extension >*.sim< or in one overall simulation result file (>simulation.result<).

R (R Core Team 2023; Giorgi et al. 2022; Ihaka and Gentleman 1996) and algorithms within R packages were used for the development of the parameter identification algorithm (ParIde, chapter 4.6.1), the model-based design of experiments (mDoE, chapter 5.3), the process optimisation (NMPC, chapter 5.4) and for the model generation tool (chapter 5.2.2). Table 7 provides details about the R packages used and in which developed software tools the respective packages were used.

The creation of data visualisations was achieved through utilising the plotting software gnuplot, which is based on C. The code for gnuplot is written in >*.plt< files and interpreted by the software to display an output on screen. Additionally, the graphics can be saved in standard graphic files formats like >*.png< or >*.jpg<. (Thomas Williams and Colin Kelley 2021)

The Python programming language (van Rossum 2010) was employed for data processing, primarily for the conducting model stability tests and calculating process performance indicators (such as yield coefficients). Additionally, Python was utilised for visualising data. The Python packages “pandas” (The pandas development team 2020) and “numpy” (Harris et al. 2020) were utilised for the data handling and “matplotlib” for visualisations (Hunter 2007).

Table 7: List of R packages, their main usage, and the software tools in which they were utilised. The software tools are the parameter identification (ParIde), model-based design of experiments (mDoE) and nonlinear model predictive control (NMPC).

Package	Usage	ParIde	mDoE	NMPC	Source
lme4	Optimisation algorithms	X		X	(Bates et al. 2015)
nloptr	Optimisation algorithms	X		X	(Johnson 2022)
reshape2	Reshaping dataframes	X			(Wickham 2007)
lhs	Latin Hypercube Sampling		X		(Carnell 2022)
rsm	Calculation of response surfaces & visualisation		X		(Lenth 2009)
zoo	Replacing “NA” values in dataframes	X			(Zeileis and Grothendieck 2005)
matrixStats	New functions for matrices (e.g. quantile determination)		X		(Bengtsson 2022)
truncnorm	Random number generation around normal distribution	X			(Mersmann et al. 2018)
Data visualisation					
ggplot2	Data visualisation	X	X	X	(Wickham 2016)
plotly	Data visualisation		X		(Sievert 2020)
ggthemes	Themes for data visualisation	X	X	X	(Arnold 2021)
RColorBrewer	Colour palettes	X		X	(Neuwirth 2022)

4.5 Mathematical model

The first research question that shall be answered is, if a structured, mechanistic mathematical model, calculating several interconnected ideally mixed zones, is able to approximate the flow patterns in stirred tank reactors and describe the effects caused by non-ideal mixing in biotechnological processes with complex kinetics. In the state of the art (chapter 2.5.3), structured biokinetic models have been identified as potentially the most suitable biokinetic submodel since they offer a sufficient description of the dynamics of bioprocesses while computation times are low enough for the application in model-based process design methods.

For this reason, the already existing, structured “6-Compartment model” published by Brüning (2016) was selected as the basis and extended for the simulations within this thesis. This model is built upon a model based on four biomass compartments by Witte (Witte 1996), which has been extended to six biomass compartments by Brüning (Brüning 2016). This extension resulted in a generalised and versatile model capable of calculating growth on multiple substrates (carbon source and nitrogen source), including diauxic growth and maintenance metabolism (Witte 1996).

The structure allows the model to simulate cultivations of various organisms, such as yeast (*S. cerevisiae*), bacteria (*E. coli* & *L. delbrueckii*), fungi (*C. striatus*), as well as mammalian cell lines (hybridoma and CHO cells), solely by adapting the model parameter set (Brüning 2016).

The model was extended to use multiple simultaneously active (sub-)models to simulate the biological behaviour of cells (e.g. cellular properties), physicochemical quantities such as the dissolved oxygen concentration (DO), temperature or pH, as well as reactor and product properties (Pörtner et al. 2017a). These three submodels namely the biokinetic, physico-chemical and reactor submodel are interconnected to form a submodel framework (Blesgen 2009; Brüning 2016; Blesgen and Hass 2010; Gerlach et al. 2013). The interactions between the submodels are illustrated in Figure 18 (Brüning 2016; Witte 1996).

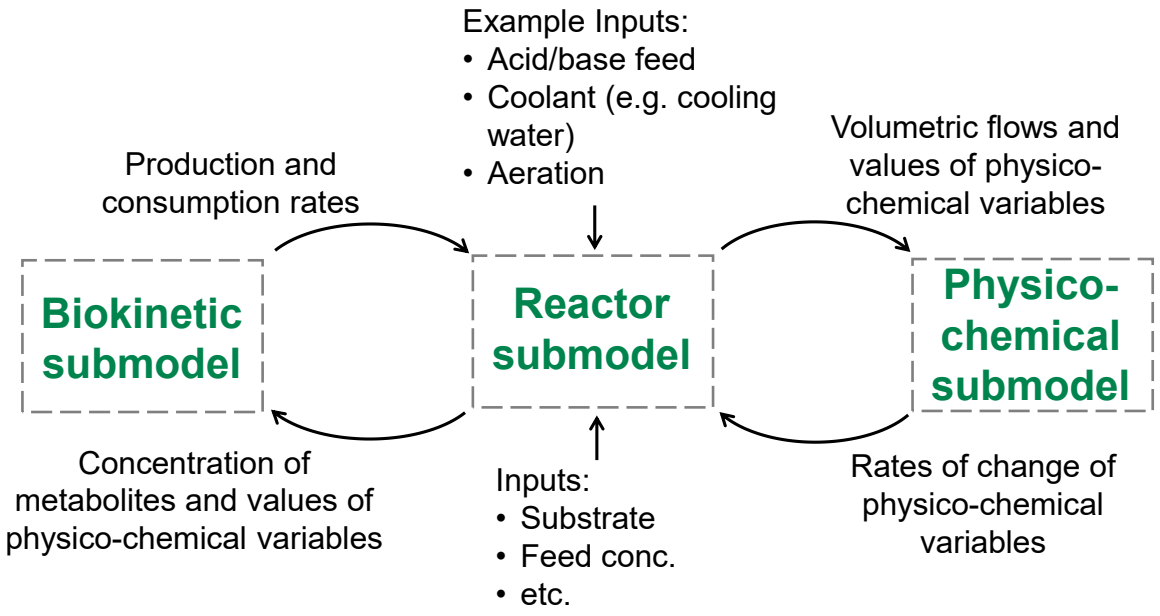


Figure 18: Interactions between “biokinetic”, “reactor” and “physicochemical” submodels.

The individual submodels can be parameterised and modified independently. This facilitates a rapid and simple parameter identification of submodel parameters, as well as a swift substitution and expansion of models within the framework.

4.5.1 Reactor submodel

The reactor model forms the central component in the submodel framework, calculating the concentrations of metabolites which are derived from general mass and energy balances (Chmiel et al. 2018; Himmelblau and Riggs 2012). The change of the medium volume is the difference between the volumetric flow rates of the inputs $F_{v,in}$ and outputs $F_{v,out}$.

$$\frac{dV}{dt} = F_{v,in} - F_{v,out} \tag{4.25}$$

The mathematical model uses the volume balance as well as mass concentrations and partial densities (kg m^{-3} = g L^{-1}) for further balances which are directly derived from the mass balance, with the assumption that the medium has a constant density throughout the process.

$$\frac{dc_i}{dt} = r_i^+ \cdot c_{Xv} - r_i^- \cdot Xv + c_i \cdot \frac{F_i}{V} - c_i \cdot \frac{F_{v,in}}{V} \tag{4.26}$$

The production rates r_i^+ and the uptake rates r_i^- are each multiplied with the viable biomass density c_{Xv}. The differential equations also include the concentrations c_i and volumetric flow rate F_i of feed streams. Furthermore, dilution effects are included into the model by multiplying the sum of the volumetric feed rates $F_{v,in}$ with the medium concentration c_i

divided by the medium volume V. The resulting concentrations are transferred to the biokinetic submodel.

4.5.2 Physico-chemical submodels

Physico-chemical submodels extend this structure and calculate the rates of change and values of physico-chemical variables, including dissolved oxygen concentration (DO), temperature (T), pH-value (pH), exhaust gas-concentrations. These rates are transferred back to the reactor submodel which in return calculates the resulting values (process states). The values of physico-chemical variables may subsequently affect the kinetics of the biological submodel (Blesgen and Hass 2010).

4.5.3 Biokinetic submodel

The structured, biokinetic submodel is based on the “6-compartment model” by Brüning (2016) and splits the biomass into six linked, separate compartments each with different metabolic roles (Brüning 2016). The uptake rate r_{S} for each substrate are the rate-limiting steps of the mathematical model and are calculated according to equation (4.27).

$$r_{\mathrm{S}} = r_{\mathrm{S,max}} \left(\frac{c_{\mathrm{S}}}{Ks + c_{\mathrm{S}}} \right) \cdot \left(\frac{c_X}{c_{Xv}} \right) \cdot \left(\prod_{i=1}^{n} f_{\mathrm{Dsig}}(x_i) \right) \tag{4.27}$$

The maximum uptake rate $r_{\mathrm{S,max}}$ is multiplied by a quotient which encompasses the substrate concentration c_{S} divided by the sum of c_{S} and the half-saturation constant Ks (Monod-kinetic). Since there are various biomass compartments with individual uptake rates, the differential equations (see equation (4.26)) multiply all rates with the viable biomass density c_{Xv}. As a result, the uptake rates are multiplied with the quotient of the density of the considered biomass compartment (here, c_X) divided by the viable biomass density c_{Xv}. The addition of the resulting uptake rates for each compartment yields the total uptake rate $r_{\mathrm{S,total}}$.

$$r_{\mathrm{S,total}} = \sum_{i=1}^{n} r_{\mathrm{S,i}} \tag{4.28}$$

The individual rates may also be influenced by metabolite concentrations or physico-chemical variables. Changes in the cellular metabolism are computed by multiplying the uptake rates with the product of multiple double sigmoid functions f_{Dsig} of the value of a state variable x such as a metabolite concentration or physico-chemical variables, for example the pH.

$$f_{Dsig}(x) = \left(Y_l + \frac{Y_{\mathrm{mid}} - Y_{\mathrm{l}}}{1 + e^{-K_{\mathrm{sl}}(x - X_{50,\mathrm{l}})}} \right) \cdot \left(1 + \frac{Y_{\mathrm{h}}/Y_{\mathrm{mid}} - 1}{1 + e^{-K_{\mathrm{sl}}(x - X_{50,\mathrm{h}})}} \right) \tag{4.29}$$

The result of f_{Dsig} becomes Y_l for low x and Y_h for high x. Y_mid is the result of the function between $X_{50,\mathrm{l}}$ and $X_{50,\mathrm{h}}$. The parameters $X_{50,\mathrm{l}}$ and $X_{50,\mathrm{h}}$ define the halfway point of the slope on the low and high side. K_sl defines the slope of the gradient. Y_mid corresponds to the standard result of the double sigmoid function and is equal to one and therefore the result of f_{Dsig} without an influence on the result. This applies when Y_h and Y_l are both equal to one, or when x is between $X_{50,\mathrm{l}}$ and $X_{50,\mathrm{h}}$ outside of the slopes. Four examples of different double sigmoid functions with different sets of coefficients (parameters) are illustrated in Figure 19 (parameter values listed in Table 8).

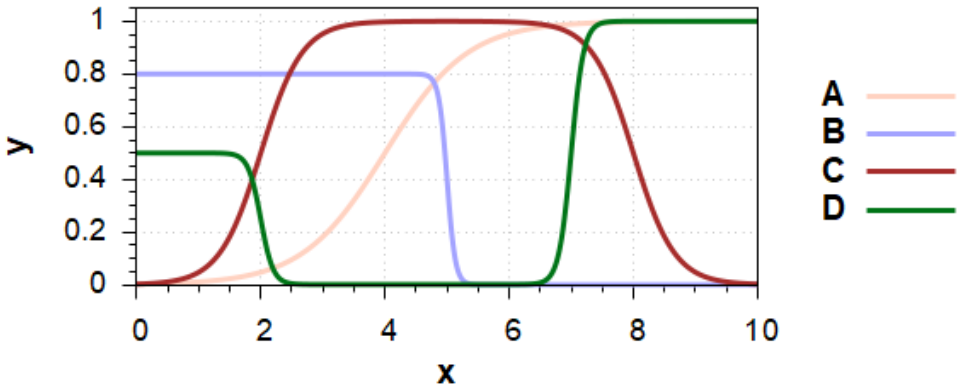


Figure 19: Example of double-sigmoid functions with different sets of coefficients (parameter sets listed in Table 8)

Table 8: Parameter sets of the double sigmoid functions in Figure 19.

	Y_l	Y_{mid}	Y_h	$X_{50,\mathrm{l}}$	$X_{50,\mathrm{h}}$	K_sl
A	0.0	1.0	1.0	4.0	10.0	1.5
B	0.8	0.8	0.0	5.0	0.0	15.0
C	0.0	1.0	0.0	2.0	8.0	3.0
D	0.5	0.0	1.0	2.0	7.0	10.0

The function is additionally utilised to describe effects of state variables on various rates such as activation, inactivation, or death rate, as well as on yield coefficients. Additionally, double-sigmoid functions calculate switch values and transitions between metabolic states such as aerobic and anaerobic metabolism or between diauxic and non-diauxic growth.

The following generalised stoichiometric function is representing every modelled metabolic pathway.

$$\mathrm{C_xH_yO_z} + \nu_\alpha \mathrm{O_2} + \nu_\beta \mathrm{H_gO_hN_i} \rightarrow \nu_\gamma \mathrm{C_aH_bO_cN_d} + \nu_\delta \mathrm{CO_2} + \nu_\varepsilon \mathrm{H_2O_t} + \nu_\zeta \mathrm{C_jH_kO_lN}_m \qquad (4.30)$$

The stoichiometric coefficients ν_i are defined as parameters and are employed in the equations to determine the mass ratio of the products relative to the substrate used.

$$Y_{i/S} = \nu_{i,S} \frac{MW_i}{MW_S} \tag{4.31}$$

The molecular weight for the metabolites involved in the pathway (e.g. biomass, O_2, byproduct) is defined as MW_i, while the weight of the substrate of that pathway is named MW_S. The individual rates for every substance involved are calculated by multiplying the yield coefficients $Y_{i/S}$ with the uptake and production rate of the corresponding substrates. The resulting rates are used according to equations (4.26) and (4.27). (Gerlach et al. 2013; Brüning 2016)

4.6 Computer-assisted evaluation of simulations

Multiple methods were used to evaluate the simulations of the developed mathematical model (see chapter 5.2). The parameter identifications employed a weighted squared mean deviation (WSMD, described in chapter 4.6.1). Simulation results were also evaluated using the coefficient of determination (R^2) as described in chapter 4.6.2. For the evaluation of the mDoE experiments, a desirability score is calculated (chapter 4.6.3). The results might be visualised and assessed using response surfaces based on the quadratic functions introduced in chapter 4.6.4.

4.6.1 Identification of model parameters

The model parameter identifications are performed by minimising the weighted mean squared deviation (WMSD).

$$WMSD = \sum_{i=1}^{n} \frac{\left(y_{s,i} - y_i\right)^2}{n} k_{i,Weighting} \tag{4.32}$$

WMSD is determined by computing the squared difference of the determined value y_i and the simulated value $y_{s,i}$, which is then multiplied by a factor for weighting individual data points $k_{i,Weighting}$. The resulting value is divided by the number of data points n in the data set. The weighting accounts for dimensional differences among data points. WMSD was minimised in all parameter identifications using the downhill simplex method (Nelder Mead method) provided in the R package *nloptr* (Nelder and Mead 1965; Johnson 2022).

4.6.2 Coefficient of determination

Simulations are quantitively evaluated and compared to experimental data using the coefficient of determination (R^2), which includes the differences between the simulated $y_{s,i}$ and the experimental data y_i, as well as the differences between the experimental data and their mean value $\bar{y}$.

$$\bar{y} = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i \tag{4.33}$$

R² is calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i - y_{s,i}\right)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{4.34}$$

Possible values for *R²* range from negative infinity to one. If the value of *R²* is equal to one, the data points perfectly align with the model solution. However, if *R²* is below zero, the mean of the determined data points is closer to the experimental data than the model solution (Colin Cameron and Windmeijer 1997; Sachs and Hedderich 2006).

R² is calculated in this work either for single concentrations and the biomass density, or as a total *R²* for all offline data together.

4.6.3 Desirability score

The desirability score D_i is utilised to evaluate the results of the mDoE, when the mDoE is performed with an uncertainty evaluation. D_i indicates how desirable individual experimental settings are during the application of the mDoE. The algorithm used in this work for the calculation of D_i is based on Möller et al. (Möller et al. 2019) and was realised using the programming language R (R Core Team 2023; Giorgi et al. 2022; Ihaka and Gentleman 1996). A higher desirability score indicates a more desirable experimental setting. D_i is the combination of a result of an optimisation function while incorporating the variability of this result into the score (Candioti et al. 2014; Möller et al. 2019). Prior to the calculation of D_i an mDoE run has to be performed where each setting in the design space must be simulated multiple times (n-times) with different parameter sets (see section 5.3) (McKay et al. 1979; Gargalo et al. 2016).

For each point in the design space, the average result $\bar{x}_i$ and the variability v_i of the optimisation function are calculated. The variability is the difference between the 10% and 90% quantile of the optimisation function utilising the function *rowQuantiles* of the *matrixStats* package in R (Bengtsson 2022). A small variability could indicate a robust experimental setting and is therefore desirable.

For this reason, the objective is to maximise $\bar{x}_i$ while simultaneously minimising v_i. The lowest $\bar{x}_i$ and the highest variability v_i of all simulated settings are assigned normalised desirability scores of zero (worst setting). (Candioti et al. 2014; Möller et al. 2019; Moser et al. 2021)

$$L(\bar{x}) = \min(\bar{x}_i) \text{ with } d\big(L(\bar{x})\big) = 0 \tag{4.35}$$

$$L(v) = \max(v_i) \text{ with } d\big(L(v)\big) = 0 \tag{4.36}$$

The highest $\bar{x}_i$ and lowest v_i are assigned to normalised desirability scores of one (best setting).

$$U(\bar{x}) = \max(\bar{x}_i) \text{ with } d\big(U(\bar{x})\big) = 1 \tag{4.37}$$

$$U(v) = \min(v_i) \text{ with } d\big(U(v)\big) = 1 \tag{4.38}$$

In the next step, the individual desirabilities for the average result $d(\bar{x}_i)$ and the variability $d(v_i)$ are determined (Candioti et al. 2014; Möller et al. 2019).

$$d(\bar{x}_i) = \left(\frac{\bar{x}_i - L(\bar{x})}{U(\bar{x}) - L(\bar{x})}\right) \tag{4.39}$$

$$d(v_i) = \left(\frac{v_i - L(v)}{U(v) - L(v)}\right) \tag{4.40}$$

Since the desirability of the variability $d(v_i)$ shall be minimised, it is inversely calculated to $d(\bar{x}_i)$ (Candioti et al. 2014; Möller et al. 2019).

Both individual desirabilities $d(\bar{x}_i)$ and $d(v_i)$ are combined into the desirability score D_i.

$$D_i = w_x \cdot d(\bar{x}_i) + w_v \cdot d(v_i) \tag{4.41}$$

The impact of the mean result and the variability on the desirability score is influenced via the weighting factors w_x and w_v. The sum of both weighting factors must equal one.

$$w_x + w_v = 1 \tag{4.42}$$

In this thesis the average simulated biomass density $d(\bar{x}_i)$ was weighted with 0.8 and $d(v_i)$ with 0.2 for the calculation of the desirability score (chapter 7.1.2). This should put the emphasis on maximising the biomass density (or optimisation criterion) while the variability is still important (Moser et al. 2021).

4.6.4 Quadratic response surface

The parameters of quadratic functions are adapted to the results from the model-based design of experiment. The resulting quadratic functions are then graphically visualised and can assist the evaluation of the results of the designs. In the first case with a single variable x (univariate case) the quadratic function is (Anderson 2016):

$$f(x) = ax^2 + bx + c \tag{4.43}$$

The resulting curve is a parabola, which is symmetrical parallel to the y-axis. With two variables (bivariate case) the quadratic function is modified and contains bilinear terms:

$$f(x, y) = ax^2 + by^2 + cxy + dx + ey + f \tag{4.44}$$

This function is capable of describing circles, ellipsoids, or parabolas. In cases with three factors or variables the function is extended to the following form:

$$f(x, y, z) = ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz + j \tag{4.45}$$

This three-dimensional quadratic function is employed and adapted to represent three-factorial design of experiments. At least one of the coefficients of the quadratic terms of each function must be non-zero. (Anderson 2016)
The coefficients are adapted by maximising the coefficient of determination with the application of the Nelder-Mead algorithm (Nelder and Mead 1965). The calculation is executed using the *rsm* package in R (Lenth 2009).

4.7 Sensitivity analysis and plausibility check

The sensitivity analysis is performed to determine the sensitivity of the mathematical model. Different approaches for sensitivity analyses exist, where different components, variables or factors of the model are varied/modified and the influence on the simulation results is quantified and/or visualised. These factors, components and variables include:

- Values of model parameters
- Initial values of variables
- Trajectories of control functions and variables
- Step size of the integration algorithm of the model
- Model equations and kinetics

As commonly suggested, these factors are varied one at a time (OFAT – one-factor-a-time) and the factor levels are either user-defined or determined by Monte-Carlo methods (Sin et al. 2009).
The sensitivity analysis in this work was based on parameters and model settings specific to the newly developed model. The sensitivity analysis performed includes:

- Values of control variables in reactors of the SDM (chapter 6.2 and 6.4)
- Different reactor designs (chapter 6.3.1)
- Discretisation of the model (number of reactor zones and models) (chapter 6.3.2)

The simulation results of the sensitivity analysis are displayed and interpreted. The results of this analysis are used to check plausibility of the model simulations.

5 Model generation, utilisation, and validation workflow

Two research questions were introduced in chapter 3. The first research question hypothesises that a structured mathematical model with multiple interconnected stirred tank reactor models and an incorporated flexible structure might be suitable for the description of effects caused by heterogeneous conditions in biotechnical processes with complex kinetics. These heterogeneous substrate or dissolved oxygen concentrations are caused by non-ideal mixing in production reactors. Alternatively, they can be controlled and adjusted in scale-down systems on the laboratory scale. The second question asks how the process operation of biotechnological processes under controlled heterogenous conditions can be systematically designed. To investigate these research questions, new software tools and a workflow had to be created and evaluated.

The new workflow assists the application of the mathematical model and model-assisted software tools. It was developed with the aim to be able to answer the research questions and in order to obtain a validated model (chapter 5.1). This workflow was followed and experimentally evaluated (chapters 7 to 9).

The mathematical model is intended to describe yeast cultivations on a laboratory scale in ideally mixed reactors and a scale-down system consisting of two connected stirred tank reactors. The existing mechanistic, mathematical (biokinetic) model introduced in chapter 4.5 was modified and extended to be capable of simulating processes under heterogeneous conditions (see chapter 5.2). This was achieved by utilising the option to simulate any number of ideally mixed STR models (maximum number is limited by the computing power) that are interconnected with each other to form a network (network of zones model) (Moser et al. 2024).

Even small modifications to the model such as a change of the number of STR models were initially too time-consuming. A short development time was necessary since different model configurations were needed throughout this thesis, especially for the simulation (and sensitivity) study in chapter 6. For this reason, a software tool for the automatic generation of user-defined process models was created (chapter 5.2.2) which reduces the development time from multiple days of manual work to a few minutes (Moser et al. 2024).

The new mathematical model was utilised for model-assisted process operation by calculating predetermined fixed feeding profiles (a) and adaptive feeding profiles (b) (2nd research question). The fixed feeding profiles were determined with a modified model-assisted design of experiments (mDoE) software. This method was used for model development and process optimisation (chapter 7.2) and for the calculation of experimental conditions of a model validation experiment (chapter 8.3.2.3). The modified mDoE method is described in chapter 5.3. An adaptive feeding profile was determined by an NMPC to

recreate data of an industrial pilot reactor process on the laboratory scale. (chapter 9.3). The developed NMPC is based on the OLFO-control algorithm and is presented in chapter 5.4.

5.1 Process development and model validation workflow

To answer the two research questions (chapter 3), a new mathematical model and software tools in combination with a physical experimental system were required, which were incorporated into a workflow (Figure 20). The experimental investigations on the laboratory scale were performed in ideally mixed bioreactors as well as in a bioreactor system consisting of two connected stirred tank reactors (physical scale-down model / SDM).

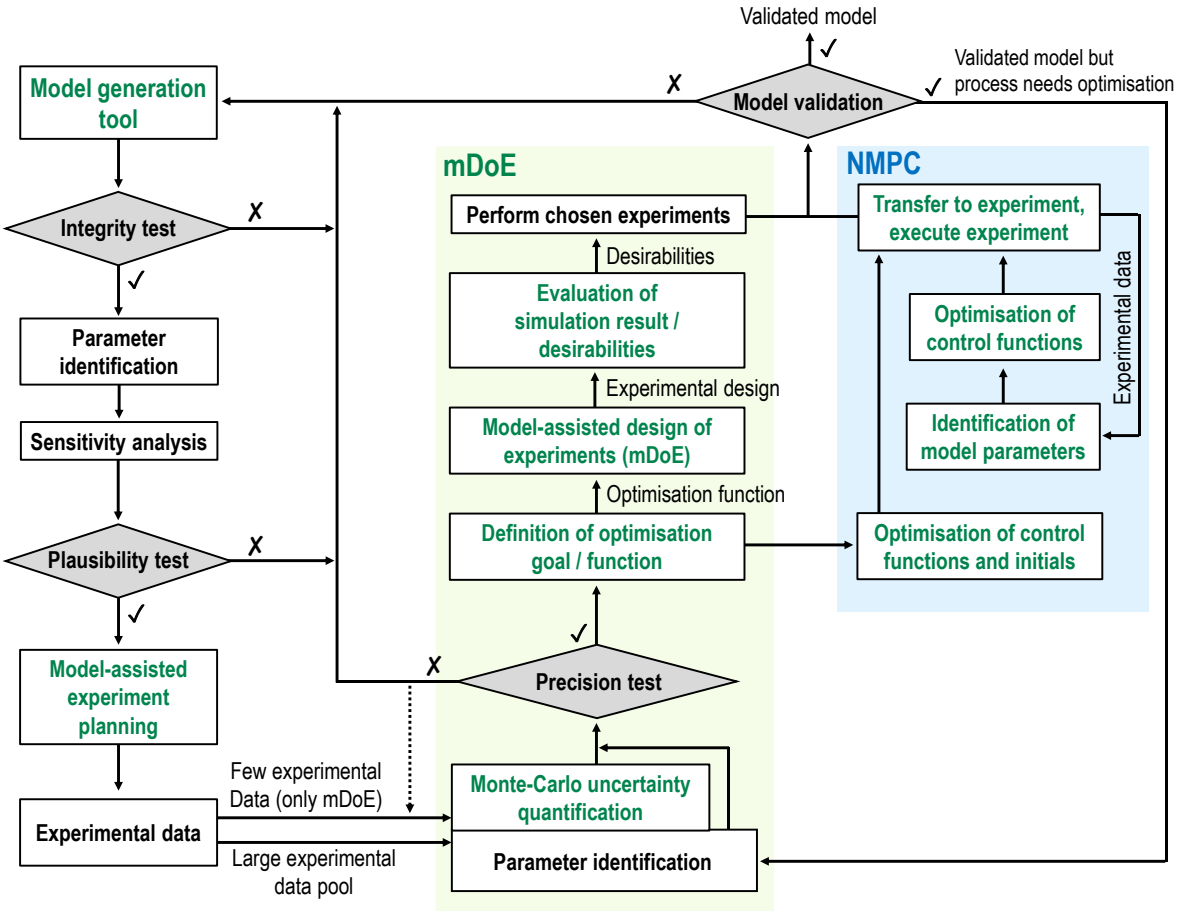


Figure 20: Workflow for simultaneous process and model development as well as optimisation for processes under non-ideal mixing conditions under heterogenous conditions (distributions of concentrations in space, varying in time).

The mechanistic, mathematical model development is assisted by a model generation tool (chapter 5.2.2) (Moser et al. 2024). This tool generates customised user-defined models, which are applicable to systems containing one single or multiple connected bioreactors. Furthermore, the user is able to select the desired submodels. The model generation tool

needs a preliminary configuration file with a parameter set. This parameter set is only used for the model generation and should include all double sigmoid function, which shall be implemented into the model.

This is followed by the model integrity test which is performed in two stages. First stage is the successful compilation of the model; the second stage checks if the model is functional and produces reasonable simulation results. Both these stages are supported by the model generation tool since only functioning models should be created. If the model is nevertheless failing to produce any simulation results, the model equations and initial parameter set must be checked for errors and be corrected. After a successful integrity test, a first parameter identification is conducted with available historic, literature or heuristic data.

These steps, until a functioning model is obtained, take only a few minutes instead of days; the parameter identification takes a few hours. This made the upcoming sensitivity analysis and simulation study feasible. For the developed model, the simulation study represents the sensitivity analysis and includes simulations of different reactor designs, discretisation of the model and varying control variables. The simulation results are analysed and interpreted.

This is accompanied by a plausibility analysis. This analysis has the goal of checking the plausibility of the mathematical model. It simultaneously improves understanding of correlations and relationships within the simulation results, which is important for an effective parameter identification. Usually the sensitivity of model parameters, initial variables or process variables is visualised and quantified (chapter 6).

The plausibility is usually checked by interpreting and comparing the simulation results against expectations, experimental data, and/or other prior knowledge of the process.

If the model is found to be plausible, new experiments may be designed and conducted with the model's assistance. The new experimental data are used to identify the model parameters, followed by testing the precision of the model simulations against experimental data. For this thesis, values of R^2 of above 0.8 are deemed good, while those above 0.9 are desirable. The quality of the model parameterisation is additionally evaluated based on qualitative criteria and the evaluation is supported by graphical comparisons. The criteria include characteristics of the process, such as similar values of process variables, similar times for extremes, comparable gradients (production or uptake rates of substances or growth rate of the biomass).

When only limited experimental data are available, conducting a Monte-Carlo uncertainty analysis might be advantageous during parameterisation. During this analysis, multiple model parameter identifications are performed to quantify the model and parameter

uncertainty. Simulating the process multiple times with different parameter sets derived from the parameter uncertainty yields a predicted process robustness (Möller et al. 2020).
However, when a large experimental data pool is available, it might be better and faster to perform one parameter identification with all data. With an increase in available, reliable data representing different experimental settings and effects, the parameter identification generally results in a more general and robust parameter set. A parameter set producing reliable and robust simulation results is essential for the application of the model in model-based process control and design methods.
After the parameter identification, two model-assisted process design methods are available. The first option for the model-based process design is a modified version of model-assisted design of experiments (mDoE), the second option the application of an adaptive nonlinear model predictive control (NMPC).
In this work, mDoE was performed with the aim of reducing experimental effort during process development and optimisation, and with the aim of designing a model validation experiment in the area of the desired process performance. The application of mDoE offers the advantage of being a multidimensional sensitivity analysis regarding the chosen factors besides being applicable for process operation design. Information is obtained about the location and shape of the area with desirable process performances. Based on this information, experiments are selected that should either optimise the process, improve the model and its parameterisation, or preferably accomplish both. Furthermore, all model simulations are performed before the experiments are carried out which simplifies the application of mDoE. But, because of this, the model must be well adjusted to the process since no adaption of the model is performed during the experiment. The mDoE was modified during this work to simulate detailed design spaces, with many (> 1000) factor combinations (simulated experimental settings) within the selected boundaries. This increases the reliability of the mDoE, as the resulting design space is not so strongly dependent on the selected boundaries (see chapter 7.2.1). In addition, the increased resolution can be used directly for the graphical assessment and enables an innovative approach of evaluating the predicted process robustness (see chapter 5.3).
The experiments determined by the modified mDoE can be simultaneously used for the validation of the model, as new experimental conditions are determined. The model is validated if the new experimental data and the predicted simulation agree with a satisfactory accuracy ($R^2 > 0.8$). The application of the mDoE (see chapter 8.3.2.3) could therefore facilitate the answering of both research questions (chapter 3.1) of this thesis.
An NMPC was utilised in this work to obtain data of desired process conditions and with the aim of simultaneously reproducing data of a pilot scale experiment on the laboratory scale.

The application of an NMPC (OLFO-control) involves repeated parameter identifications during the experiment. This allows for the usage of process models with parameter sets that are not yet well fitted, since the parameter sets are improved during the experiment. However, this requires modelling effort during the running cultivation, as both a parameter identification and a model-based control (optimisation of control trajectories) are performed. These simulations should be carried out as fast as possible to minimise the time between the last sample being taken, all relevant values being determined, the identification of the model parameters and finally the transfer of the optimised control trajectories to the process. A great advantage of the application of an NMPC is, that the operational strategies of the processes are periodically recalculated since disturbances might occur that might influence the process performance. Furthermore, the NMPC is utilised to optimise the process and simultaneously improve the identification of the model close to the expected optimum.

Newly acquired experimental data, which was not used for parameter identification, can be used for the validation of the model. The precision of the experimental data is checked against the prediction of the mathematical model. This is done by calculating the R^2 ($R^2 > 0.8$, desirably $R^2 > 0.9$) and additionally comparing the experimental and simulated data visually. If the validation is not successful, the model must be modified, or the model parameters identified again. If the process optimisation is not satisfactory or incomplete, the mDoE or the NMPC should be repeated.

The two methods could also theoretically complement each other. For example, the mDoE can be carried out initially, prior to cultivation, and the resulting cultivation experiment can be controlled with the utilisation of an NMPC.

This workflow aims at validated models that are well parameterised for desirable process performances. Furthermore, the process is optimised in parallel with the mathematical model.

This workflow was tested and validated for *S. cerevisiae* processes in lab-scale bioreactors (chapter 7), as well as for processes with special distributions of the dissolved oxygen and substrate concentration (chapter 8 and chapter 9). Chapter 6 comprises the simulation study and model plausibility study for the newly developed mathematical model.

5.2 Adaptations to the mathematical model

The mathematical model must be able to simulate the effects of heterogeneous conditions caused by non-ideal mixing. In the case of industrial *S. cerevisiae* cultivations, the dissolved oxygen concentration (DO) or/and the substrate concentration were identified to potentially exhibit spatial distributions in the reactor medium that might have a relevant influence on

the process performance. These heterogeneous concentrations might be simulated with a reactor compartment model / network of zones model.

For an effective application of the mathematical process model, three challenges and bottlenecks have been identified and need to be overcome:

1) Mathematical representation of cultivations under heterogenous conditions and in scale-down models caused by non-ideal mixing.
2) Fast, flexible adaptations, additions, or removals of submodels without laborious work to e.g. change the number of reactor models or reduce the simulation time.
3) Fast computation times are needed for an effective application of the model in parameter identification as well as in process design and operation.

These challenges are addressed by three components, an extended model structure and a versatile, innovative model generation tool:

1) A model using a network of zones (STR models) allows for the representation of different reactor configurations. This enables the simulation of scale-down systems consisting of multiple connected stirred tank reactors including the connection between the reactors. Furthermore, this structure should be capable of modelling non-ideal mixing conditions and flow patterns in bioreactors, as these might significantly affect the kinetics, performance, and dynamics of bioprocesses. It is important to incorporate flexibility to represent various reactor types and/or non-ideal mixing through modifications in the reactor submodel. These modifications to the model are described in chapter 5.2.1.
2) A tool for the automatic generation of the mathematical model was developed which is presented in chapter 5.2.2. The tool generates mechanistic models by automatically modifying model equations and implementing submodels from a model library, based on user input. This tool reduces the time needed for model creation to a few minutes and minimises the risk of errors during the model creation (Moser et al. 2024). Without an automatic model generation tool, each adaption to the model structure and equations would have required at least a few hours up to several days of manual coding work. This duration is required since the modifications necessitate adaptations to model equations throughout the model code such as modifying the number of reactor compartments or adjusting, adding, or removing submodels. Throughout this work more than 50 different model configurations were created, e.g. during the simulation study (see chapter 6).
3) The model generation tool reduces computational workload automatically by pre-solving computationally intensive equations (double sigmoid functions), accompanied by an option to activate a more stable and faster simulation mode

(chapter 5.2.2). The reduction of the computation time was crucial to support and facilitate parameter identification and model-based process optimisation. Originally, the numerical solution of the mathematical model, which consists of a high number of nonlinear differential equations, required extensive computational effort (>20 min per simulation).

5.2.1 Structure of the reactor model (network of zones)

The reactor model (see chapter 4.5.1) was extended to a network of zones model to enable the simulation of a freely selectable number of interconnected ideally stirred tank reactor models (STR models) (see chapter 2.5.4.2). This network of STR models should be able to simulate the influence of non-ideal mixing on cultivations when combined with a biokinetic submodel. The full model employs one structured, biokinetic (6-compartment model, chapter 4.5.3) and one physico-chemical submodel (chapter 4.5.2), with one single set of parameter values.

Figure 21 presents the general structure of this new mechanistic, mathematical model.

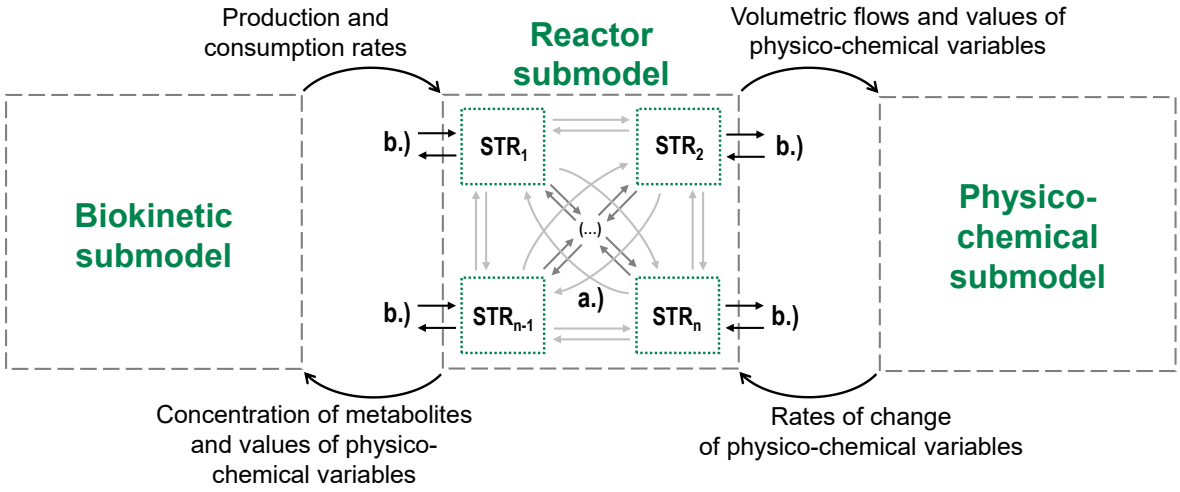


Figure 21: Structure of the new mathematical model, based on one biokinetic and one physico-chemical submodel. The reactor submodel is split up into several ideally mixed tank reactor model (zones), which are all connected to each other and have individual inputs and outputs forming a network of zones model.

Each single STR model in the network is connected to every other STR model, with individually adjustable inputs and outputs. This enables the possibility for simulations of numerous different reactor configurations through a network of zones (of interconnected ideally mixed STR models). The STR models are variable in size and arrangement (variable over time), unlike CFD approaches, where the finite elements are often uniquely sized, constant in their size and their position throughout the simulation process.

The mass balance is inherently closed over all volume elements. This contrasts with systems of partial differential equations where the mass balance is not implicitly closed. Each STR model requires one set of ordinary differential equations containing the mass balances, as well as the rates of change for biokinetic and physico-chemical variables. In the model the density ρ of the medium is assumed to be constant. This implies that the overall mass balances can be replaced by volume balances (and volume flows).

The ODEs were extended by terms describing the volumetric and mass flow between reactor compartments. The calculation involves totalling volumetric flows leaving ($F_{out,Comp}$) and entering ($F_{in,Comp}$) compartment i from every other compartment j.

$$F_{in,Comp,i} = \sum_{j=1}^{n_{Comp}} F_{j,i} \tag{5.1}$$

$$F_{out,Comp,i} = \sum_{j=1}^{n_{Comp}} F_{i,j} \tag{5.2}$$

$F_{j,i}$ and $F_{i,j}$ might be for example the volumetric circulation flow rates between the two stirred tank reactors in the physical scale-down model. The ordinary differential equation (ODE) for the medium volume in each compartment is modified accordingly:

$$\frac{dV_i}{dt} = F_{in,i} - F_{out,i} + F_{in,Comp,i} - F_{out,Comp,i} \tag{5.3}$$

The equations representing the concentrations of each substance k are expanded by a term to calculate the substance input from all other STR models. In addition, the dilution term is extended to consider the input through feeds $F_{in,i}$ and volumetric flow rates $F_{in,Comp,i}$ from other STR models.

$$F_{v,in,i} = F_{in,i} + F_{in,Comp,i} \tag{5.4}$$

$$\frac{dc_{k,i}}{dt} = production - uptake + inputs - dilution + inputs\ from\ other\ STR\ models \tag{5.5}$$

$$\frac{dc_{k,i}}{dt} = r_{k,i}^{+} \cdot c_{Xv,i} - r_{k,i}^{-} \cdot c_{Xv,i} + c_{input,k,i} \cdot \frac{F_{input,i}}{V_i} - c_{k,i} \cdot \frac{F_{v,in,i}}{V_i} + \sum_{j=1}^{n_{Comp}} c_{k,i,j} \cdot \frac{F_j}{V_i} \tag{5.6}$$

The differential equations of concentrations for every component k in each STR model i can be derived from the general mass balances if the density of the culture broth does not change with time.

The differential equations of the concentrations considering the production rate $r_{k,i}^{+}$ and uptake rate $r_{k,i}^{-}$ multiplied with the viable cell density $c_{Xv,i}$. The change of the concentration caused by inputs (e.g. feeding) on the concentration are calculated with the concentration

of k in the input flow $c_{input,k,i}$ multiplied by the volumetric flow rate $F_{input,i}$ divided by the volume of medium in the STR model V_i. The dilution caused by feeding is calculated with the concentration of substance k within the medium $c_{k,i}$ of the STR model i multiplied by the combined volume flow rate of all inputs $F_{v,in,i}$ divided by the volume V_i.

At this stage, the gas phase was not considered separately from the liquid phase. Dissolved oxygen and carbon dioxide is exchanged between the individual STR models as solutes in the medium. The aeration and feeds can be positioned individually in each STR model.

By incorporating a biokinetic model into the zone model, concentration gradients can be simulated in a variety of non-ideally mixed reactors (chapter 0). To date, simpler unstructured models have mainly been integrated into network of zones models to calculate concentration gradients within reactors or growth rates as a function of the location. The most commonly used models to account for the hydrodynamics are even pure CFD models without consideration of biokinetic reactions. The focus of the other research groups was primarily on a realistic representation of the hydrodynamics of the system and less on the biokinetics (see chapter 2.5.4). In addition, these hydrodynamic models were utilised to obtain knowledge and insight into the formation of concentration gradients in bioreactors and their realistic recreation with simulations (Delafosse et al. 2014). The biokinetics were more often considered by Lagrangian simulations rather than by Network of Zones models based on Eulerian CFD simulations (Delvigne et al. 2017). The computing time of the models only played a minor role, as the models were not used during the process or within numerical optimisation algorithms.

The new model developed in this study should be able to represent entire cultivation processes that pass through several growth phases, considering non-ideally mixed behaviour. Furthermore, this new model structure should be able to model processes occurring under ideally mixed and heterogeneous conditions with only one set of parameters. This model has also developed with the aim of being applicable in process control and optimisation in order to support process development, process optimisation and model validation. To the author's knowledge, such a mathematical model with a structured biokinetic submodel and a flexible reactor model does currently not exist.

Utilisation options for the new reactor submodel

The model structure of a user-defined number of connected ideally stirred tank reactor models enables a multitude of options for simulations. Figure 22 presents different utilisation options of the novel model structure.

The first two options involve simulating single bioreactors. The model retains the ability of simulating cultivations in one single ideally stirred tank reactor (Figure 22 a.).It is possible

to simulate reactor zones (Figure 22 b.), with each zone containing distinct effects or combinations of effects, for example, a zone at the top of the reactor with low DO and high substrate concentration or a zone around the air inlet at the bottom with high DO and low substrate concentration. It is also possible to create detailed network of zones models with a high number of reactor compartments (Figure 22 c.), allowing for a detailed representation of non-ideally mixed bioreactors at the expense of increased computation times.

The heterogeneous conditions of scale-down models can be simulated by connecting multiple stirred tank reactor models (Figure 22 d.). A comparison between simulations of these three model types (b., c., and d.) is presented in the sensitivity analysis in chapter 6.3.1.

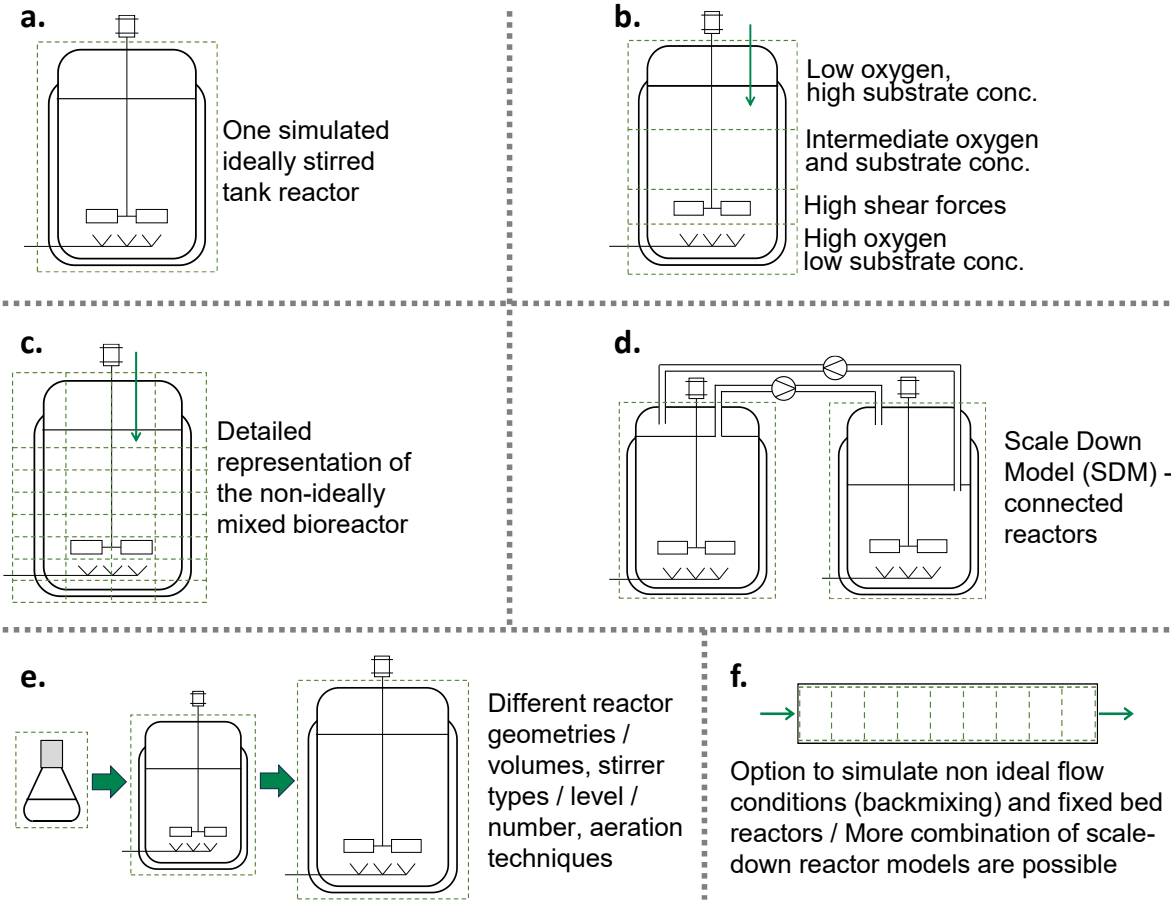


Figure 22: Reactor types, configurations, and combinations capable of being simulated with the new model structure. a.) Single ideally stirred bioreactor; b.) Reactor zones; c.) Detailed flow regimes of non-ideally mixed stirred reactors; d.) Reactor configuration of connected reactors; e.) Seed-Trains with consecutive cultivations; f.) Tubular flow reactors

Further options are the simulation of seed-trains (Figure 22 e.) and tubular flow reactors (Figure 22 f).

Seed-trains are utilised to provide the required dry cell weight concentration for a production reactor. They can be simulated by connecting multiple STR models, each model representing one stage. A transfer flow between STR models is only activated at the end of each stage. It is possible to consider individual scale-effects for each stage in the simulations for seed-trains.

Tubular flow reactors are simulated using a series of connected STR models, also optionally comprising the simulation of non-ideal flow conditions such as back mixing between STR models. Additionally, other combinations of scale-down models are possible to model (e.g. combinations of stirred tank reactors and plug flow reactors). In this thesis, for example, the piping between reactors for the circulation of the medium, were simulated as series of STR models.

5.2.2 Model generation and development tool

During the development of the mathematical model, the sensitivity analysis and simulation studies (chapter 6), several model versions (more than 50) were needed to be created with different numbers of STR models. Initially a few days of manual coding work were necessary to change the number of ideally stirred tank reactor submodels (reactor zones) which would have resulted in months of manual coding work. In addition, fast computation times of a few seconds at most for one model simulation were necessary for the application in model-based bioprocess design methods.

For this reason, a novel software tool was developed with the objective of accelerating the generation of mathematical models (from days to minutes). The second reason for the creation of this tool was the reduction of the initially high computation time. As the number of STR models increases, the time required to solve the mathematical model also increases because each STR model has its own set of differential equations that must be solved. Furthermore, the number of intermediate volumetric flows between compartments is equal to the number of STR models squared. A model comprising 12 STR models initially required approximately half an hour to be solved for a process with a duration of 48 hours (specifications of the utilised computer hardware listed in Table 21 in the Appendix). Parameter identification and process optimisation are not feasible with these high computation times. Therefore, it is crucial to increase the simulation speed.

The computation time can be reduced by streamlining the mathematical model. This can be achieved by reducing the number of submodels, by reducing the number of equations or by replacing nonlinear differential equations with equations that can be solved more rapidly (e.g. algebraic equations). Consequently, the generation tool is capable of adjusting, incorporating, and removing submodels, while only including necessary equations in the

model. Furthermore, a faster (although less accurate) model can be generated. Priority is given to the computing time of the mathematical model, which must be short enough for the intended purpose of the model (e.g. a few seconds for parameter identification). Afterwards, it must be checked whether the model fulfils the set requirements in terms of calculation accuracy (e.g. $R^2 > 0.8$).

Figure 23 presents the structure and workflow of the model generation tool.

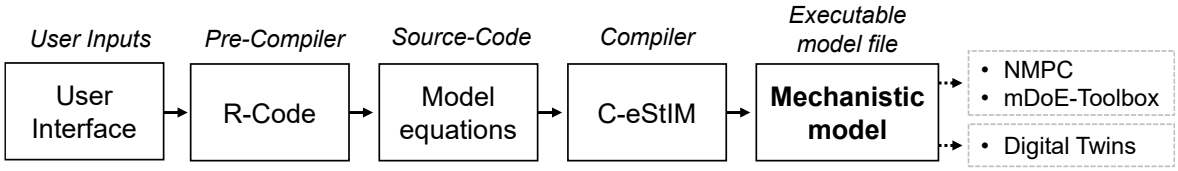


Figure 23: Simplified structure of the model generation tool. User-Inputs are transferred to the pre-compiler, which writes the model equations, and the equations are compiled into the executable model file.

The user inputs are processed in a user interface and transferred to a pre-compiler (both coded in R). This pre-compiler creates the desired model code, according to the user-inputs, and inserts the code into the individual >*.cpp< source code files of the model. After successful implementation, the model is generated via a C++ compiler utilised for C-eStIM models. This results in the compilation of an executable file of the mechanistic model, which can then be utilised in model-based methods, such as model-based optimisation and control.

The available options for the user inputs are presented in Figure 24, together with the sub-model at the same level to which these options belong.

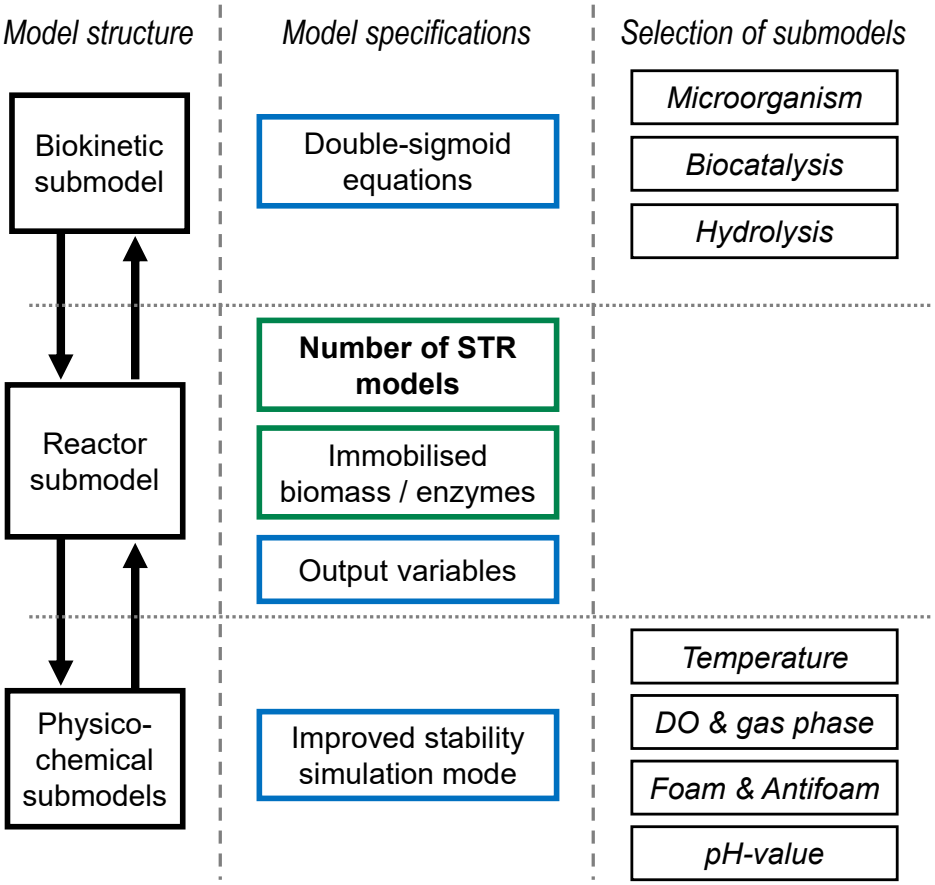


Figure 24: Model specifications and selectable submodels for the biokinetic, reactor and physico-chemical submodel. The centre column shows model specifications; green boxes indicate new model functions and blue boxes technical modifications to accelerate calculation times and improve model stability. On the right selectable submodel options are shown.

The user selects which submodels should be included into the mathematical model. There are submodels regarding the biocatalysis and hydrolysis which belong to the biokinetic submodel. The temperature, DO & gas phase, foam & antifoam, and pH model are physico-chemical submodels and can be included or excluded from the mathematical model. Furthermore, there are selectable model specifications. The user defines which double-sigmoid functions are included and which variables shall be exported into files. Further options are the definition of the number of STR models and if the biomass shall be immobilised. The volumetric flow rates between the STR models are defined in separate control files. There is also an option for a mode with improved simulation stability at the expense of calculation accuracy. These options facilitate and enable a fast, reliable, and knowledge-based model reduction.

5.2.2.1 Reduction of double-sigmoid equations

In the biokinetic submodel a total of 94 double sigmoid functions f_{Dsig} for each reactor compartment must be solved, but only a fraction of these functions is active depending on the model parameters. An active double sigmoid function is defined when either Y_l, Y_{h} or Y_{mid} are not equal to one. This condition is reduced to Y_l and Y_{h}, since Y_{mid} is equal to one in all equations. If both Y_l and Y_{h} have a value of one, the result of f_{Dsig} will always be one regardless of the other variables of the function.

$$if\ Y_l\ and\ Y_{\mathrm{h}} = 1\ then\ f_{Dsig}(x) = 1 \tag{5.7}$$

However, the model still calculated the whole double-sigmoid function. The mathematical solution of a double sigmoid function is resource intensive because exponential functions must be solved numerically. Since the result for the specific case presented in equation (5.7) is known, f_{Dsig} can be set to one without having to compute the whole function.

The reduction of 94 double sigmoid functions to the 26 necessary functions for the description of the biokinetics of *S. cerevisiae* decreased the simulation time by 15%.

5.2.2.2 Selection of output variables

A considerable amount of time needed for the model execution was required for exporting the time courses of state variables as calculated from the ODE's and the time courses of selected additional variables, such as specific rates as well as results from algebraic calculations (i.e. DO) into simulation data files.

Modification and implementation of the model code to change the additional exportable variables was time consuming. To facilitate and accelerate this process, the model generation tool is capable of automatically implementing and modifying the model source code required for variable export based on user-inputs. This enables a swift creation of models that only export desired process variables. For example, reducing of the number of output variables from 20 to 5 reduces the computation time by approximately 5%.

5.2.2.3 Improved stability simulation mode

The lower the number of simulation steps (corresponding to a larger simulation step size), the shorter the simulation time, but this may also lead to non-stable simulations. A stable model is important for an effective parameter identification, during a Monte Carlo-based uncertainty quantification, and for mDoE applications (chapter 7.1.2). To calculate the model stability, the minimum number of computation steps is determined that is required for a successful simulation. A simulation is defined as not successful when the simulation results significantly change (assessed with changes in R^2) without modifying inputs (except

computation steps). A lower step count (larger simulation step size) corresponds to less robust simulations but faster possible computation times.

Initially, the model simulations required high step numbers (small step sizes) which was mainly caused by stiff differential equations of the gas-phase submodel. This submodel calculates both the dissolved oxygen concentration (DO or pO_2) and the composition of the exhaust gas (y_{O_2} and y_{CO_2}). This instability was compensated by a high number of simulation steps, which resulted in high computation times. One way to reduce the computing time for the solution of such systems is to convert the differential equations with the shortest time constants into algebraic equations (see (Witte 1996)), as presented in Table 9.

Table 9: Comparison between the differential equations in the model and the implemented algebraic equations adapted from (Witte 1996).

Differential equations		Algebraic equations	
$\frac{dO_2^{out}}{dt} = \frac{O_2^{in} \cdot f_{g,in} - f_{O2,gl} - O_2^{out} \cdot f_{g,out}}{V_H}$	(5.1)	$O_2^{out} = O_2^{in} - r_{O_2} \cdot X_v \cdot f_{g,in} \frac{V_m \cdot 100}{M_{O_2}}$	(5.2)
$\frac{dCO_2^{out}}{dt} = \frac{CO_2^{in} \cdot f_{g,in} - f_{CO2,gl} - CO_2^{out} \cdot f_{g,out}}{V_H}$	(5.3)	$CO_2^{out} = CO_2^{in} - r_{CO_2} \cdot X_v \cdot f_{g,in} \frac{V_m \cdot 100}{M_{CO_2}}$	(5.4)
$\frac{dO_{2,L}}{dt} = k_L a_{O_2} \cdot (O_{2,eq} - O_{2,L}) - r_{O_2} \cdot X_v + \frac{(O_{2,L,max} \cdot f_{L,in} - O_{2,L} \cdot f_{L,in})}{V_M}$	(5.5)	$DO = \left(\frac{O_2^{out}}{H} - \frac{r_{O_2} \cdot X_v}{k_L a_{O_2}}\right) \cdot \frac{100}{O_{2,L,max}}$	(5.6)
$DO = \frac{O_{2,L}}{O_{2,L,max}} \cdot 100$	(5.7)		

This modification of model equations could substantially affect the outcome of the DO results, potentially rendering the subsequent model parameter identifications only valid for the newly implemented algebraic equations. The faster calculation time makes up for this disadvantage.

Another possibility to reduce the simulation time is to decrease the potential minimum number of simulation steps, by artificially reducing the substantial differences in the model's time constants (stiff differential equation). Biokinetic differential equations may have much higher time constants than those related to temperature equations, often differing by many orders of magnitude. This results that temperature equations such as those for heat transfer dictate the minimum number of computational steps. To address this, pre-factors were introduced to artificially decrease the heat transfer coefficient by 99%. This only affects the

calculation of the medium temperature. The model is then no longer suitable for identifying the parameters of the temperature equations, but the medium temperature is typically controlled at a constant value in cultivations anyway.

Figure 25 illustrates a comparison between the original and modified versions of the mathematical model for computation time and R^2. The R^2 calculation was performed to be able to assess the influence of the number of simulation steps on the simulation result with one simulation step per second as the baseline. The calculation was based on the simulation results of the maximum step number (one step per second). The average R^2 was calculated using eight simulation variables, the concentrations of biomass, glucose, ethanol and O_2 in the exhaust gas, as well as DO, medium temperature, volume, and pH value. The differential equation system of the model is solved numerically using the Runge-Kutta algorithm (Runge 1895).

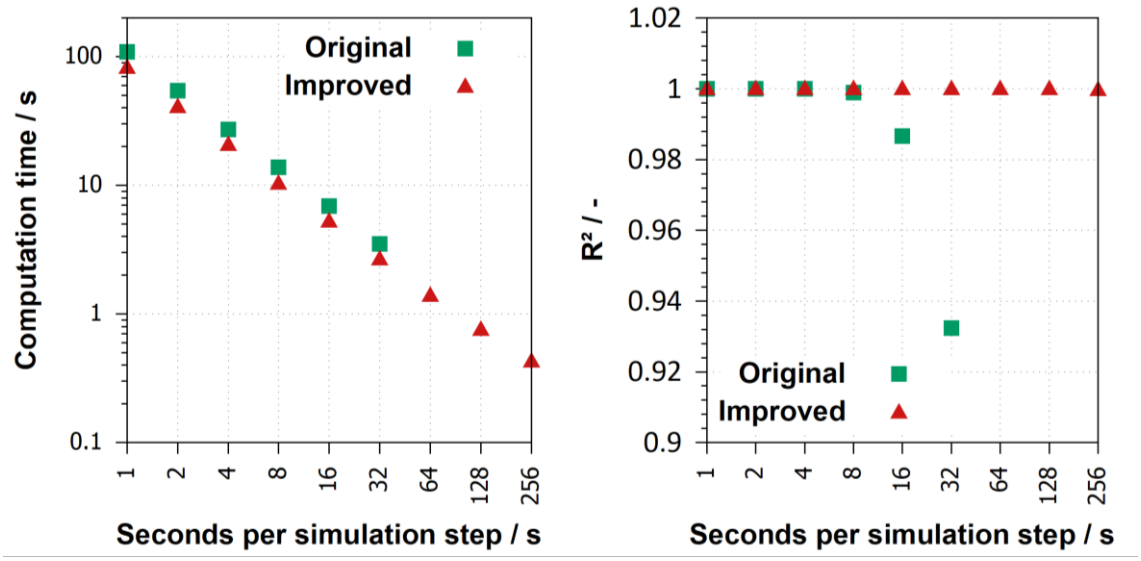


Figure 25: The initial mathematical model in comparison to the improved version with algebraic equations and slower heat transfer for R^2 computed via the difference between the experiment at one step per second compared to the simulations with less steps as well as the simulation time for a 48 h simulation experiment.

The simulation speed of the improved version is increased by 25% compared to the original version for the same simulation step sizes.

The data clearly demonstrate that the improved model remains more stable at lower step numbers compared to the original model. Based on the average R^2, the minimum number of steps without changing the results for the original model version would be an increment of 4 simulated seconds for each solver iteration. The differential equations of the original model become unstable at 64 simulated seconds per iteration and the model simulation does no longer produce meaningful results. However, the improved version can handle

even 64 or 128 simulated seconds per iteration step without resulting in any significant alterations of the simulation result.

In summary, the model improvements led to a 95% reduction in computation time from 30 seconds (4 s per simulation step) to 1.5 seconds (64 s per simulation step) without removing submodels, for the considered case of a single ideally stirred tank reactor model and a cultivation time of 48 h. Simultaneously, the computational stability was also improved.

Moreover, the model generation tool offers another way of reducing the simulation time if certain physico-chemical variables do not need to be calculated. It is possible to integrate only user-selected submodels into the generated mathematical model. The physico-chemical variables are then either set to a constant value or a user-specified time-dependent profile using control files which could potentially save over another 95% of the time if these physico-chemical submodels are speed-determining. This is particularly beneficial during the parameter identification of the mathematical model as the profiles / time courses of the variables are known.

5.3 Model-assisted design of experiments software

The mDoE application utilised in this thesis is based on the work of an academic working group led by Prof. R. Pörtner, Prof. V.C. Hass, Prof. B. Frahm and Dr.-Ing. J. Möller in collaboration with the author of this thesis. The mDoE concept was initially developed in MATLAB R2018b (The MathWorks Inc. 2018) (The MathWorks Inc., USA) (Möller et al. 2015; Möller et al. 2019; Möller and Pörtner 2017; Moser et al. 2021). It was subsequently translated to R by the author and was modified, adapted and extended for this work (Moser et al. 2021).

The application of model-assisted design of experiments (mDoE) aims to reduce the number of experiments necessary for optimising the process and improving the model parameterisation. The high number of experiments usually required are one of the main factors contributing to the time and cost intensity of bioprocess development (Scarlat et al. 2015; Guo and Song 2019). The mDoE approach combines mathematical modelling with statistical DoE-methods. Figure 26 illustrates the workflow for the application of the modified mDoE version (Moser et al. 2021).

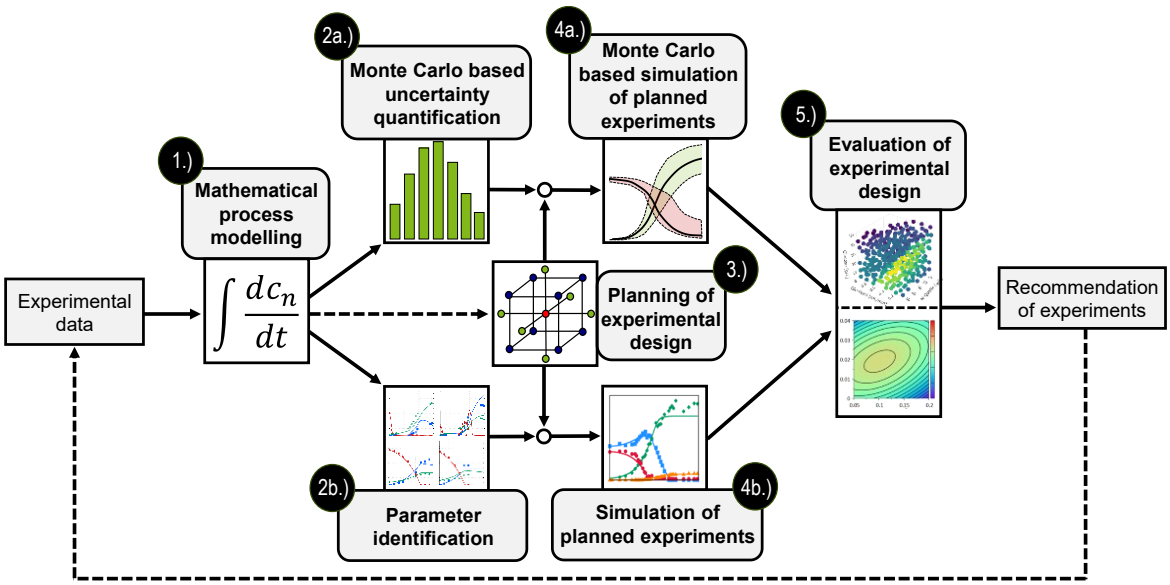


Figure 26: Workflow of the mDoE-concept as developed by Möller et. al (2015) and further developed for this work: 1.) Mechanistic mathematical model considering non-ideal mixing; 2a.) Monte-Carlo based uncertainty quantification with low available data or 2b.) Parameter identification with a high amount of available data; 3.) Planning of experimental design; 4a.) Monte-Carlo based simulation of planned experiments or 4b.) Simulation of planned experiments with one parameter set; 5.) Evaluation of experimental design. This workflow results in recommendations for new experiments (Moser et al. 2021; Möller et al. 2015).

Available experimental data such as historical data and/or literature are utilised in the development of the mathematical model (Figure 26-1). These data are used to identify the mathematical model parameters once (Figure 26-2b) or multiple times to quantify model uncertainties (Figure 26-2a). The Monte-Carlo based uncertainty quantification is especially useful when experimental data is limited. It involves multiple identifications of the selected model parameters. For each parameter identification, the experimental data, control variables and initial values are varied by Monte-Carlo sampling based on experimental uncertainties (e.g. measurement inaccuracies or standard deviation). Experimental data points are considered to be independent from each other and to be normally distributed. Individual control variables and their time courses including feeding rates, set temperature, set pH value, DO, concentrations of feed streams, are collectively varied. The standard deviations are determined by typical standard deviations in bioprocesses based on expert knowledge or with experimental data (Wechselberger et al. 2013; Hernández Rodríguez et al. 2019). This method leads to multiple identified parameter sets.

The parameter identifications are performed by minimising the weighted mean squared deviation (WMSD) (see section 4.6.1). The resulting parameter sets are utilised to quantify

the model uncertainties and are subsequently analysed by visualising with correlation matrices, distribution histograms or cumulative curves. (Möller et al. 2020)
When a sufficient number of experimental data points are available, it is also possible to just determine one set of parameters without Monte Carlo-Sampling (Figure 26-2b).
Next, an experimental design is planned, which might be classic DoE designs such as a Box-Behnken design or a full factorial experimental design (Figure 26-3). The designs can be imported or automatically created. While the number of dimensions is freely selectable, the evaluation and analysis of more than three dimensions is challenging.
The automatic creation of designs allows a freely selectable number of factor combinations (points in the design space). This enables detailed runs, where many points (>1000) are simulated. This potential of simulating detailed design spaces represents a modification and further development from previous versions of the mDoE-Toolbox (Möller et al. 2019; Möller et al. 2015; Möller and Pörtner 2017; Gassenmeier et al. 2022). The modification offers the novel possibility of accurately locating the position of the optimum as well as the location, shape, and rotation of the response surface (desirability function).
The visual analysis of the detailed design space enables a new evaluation of the predicted process robustness that was previously inaccessible. Experimental settings located far away from steep gradient or edges, in the centre of “plateaus” of high desirabilities, can be assumed to be more robust. The increased robustness is expected since small deviations of the factors do not lead to vastly different results. Moreover, it is theoretically possible to obtain and observe multiple optima. This is not possible with quadratic functions, which are typically used for the analysis of DoEs. Multiple optima are also not reliably detectable by traditional statistical design. The version of the toolbox programmed in R, which was utilised in this thesis, was extended by that feature.
The points of the automatically created designs were distributed as evenly as possible in the design space (maximising the distances between points), to reduce the number of points required for a good coverage. A numerical algorithm was developed and implemented in R to achieve this distribution, as this is rarely possible to be performed analytically in multidimensional spaces. The workflow of this algorithm is explained in the following section.
Initially every factor is normalised in the range between 0 and 1 before being converted to actual values at the end of the design creation workflow. The factor combinations are determined with the aim to maximise the distance between points of the design. This is done by first distributing many points (ideally >10^6 points) randomly in the design space, drawn from a uniform distribution. These points are then clustered into k clusters using the *kmeans* clustering algorithm in R with the Lloyd algorithm (Hartigan and Wong 1979; Lloyd

1982). The centres of these clusters are the factor combinations of the DoE design. The outermost centres are shifted onto the borders of the design space (0 or 1). Finally, all determined points are converted to their corresponding real values, creating the finished design.

In the next step, each point in the design is then simulated multiple times with n parameter sets (Figure 26-4a). The parameter sets between different points are identical. Latin Hypercube Sampling is used to draw n new parameter sets from the identified parameter sets (McKay et al. 1979). When only one set of parameters is identified, each point of the design space is simulated once (Figure 26-4b). Desirability scores are calculated (when n > 1) based on the simulation results and are graphically presented as 3D-plots and as response surfaces of quadratic functions (Figure 26-5). These plots are then utilised for the recommendation of new experimental settings. Overall, the mDoE-Toolbox supports a risk and knowledge-based evaluation of experimental setting and designs. (Moser et al. 2021)

5.4 Non-linear model predictive control software

The NMPC realised and utilised for this thesis is the Open-Loop-Feedback-Optimal strategy (or OLFO-strategy). The OLFO-strategy is realised in R, as outlined in chapter 4.3.4.

The process model utilised in the OLFO-strategy is the network of zones model of multiple interconnected stirred tank reactor models described in chapter 5.2. The parameter identification model is the same parameter identification algorithm as described in the previous chapter 5.3 without the quantification of the model uncertainty. The OLFO-strategy is consisting of four building blocks, the experimental system, the parameter identification algorithm, the process model, and the control function optimisation (see Figure 27).

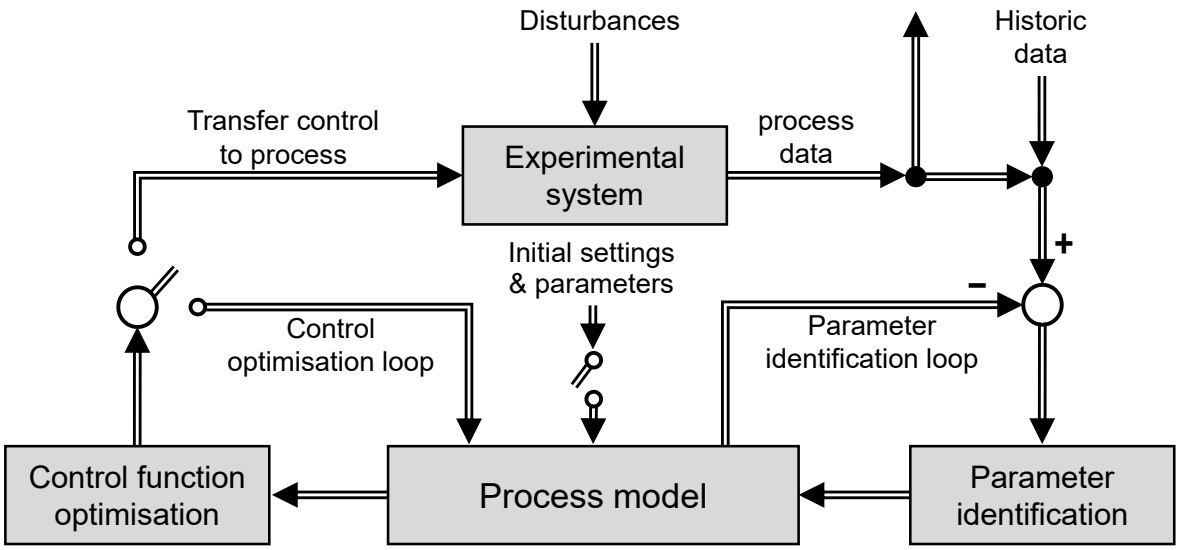


Figure 27: Modified structure of the OLFO-strategy consisting of the experimental system, the parameter identification algorithm, the process model, and the control function optimisation. The OLFO strategy is based on (Luttmann et al. 1985) and (Witte 1996).

Prior to a particular process run, available, historic data is used to identify parameters and initial values of process variables. During the process run, historic data is used together with new process data for the parameter identification. The weighting of the new data was incrementally increased to boost their influence. This approach of using already available data is an extension of the OLFO-strategy and was utilised and tested in this work. The inclusion of historic data is possible because the computation time of the process model was reduced to less than 2 s per simulation (chapter 5.2).

The inclusion of historic data might be important, since at the beginning of the NMPC cultivation the parameter set is not perfectly identified. If only data from the NMPC experiment would be utilised for the parameter identification, the risk of overfitting would be increased. This issue was previously addressed by limiting the number of identifiable parameters and constraining the identification range. The utilisation of historical data has the potential to enhance the quality of the identified parameter sets during the NMPC application and simultaneously reduce the risk of overfitting, without imposing constraints on the parameter range.

The control function optimisation algorithm is calculating control trajectories of one or more selected actuating variables. Different control trajectory profiles can be selected including steady state, linearly increasing or exponential functions. This reduces the number of values that need to be optimised and therefore accelerates the optimisation compared to previous versions of the OLFO-control. In previous versions, constant values of defined time periods were optimised (Frahm et al. 2005).

The optimisation criterion can either be minimised or maximised, user-defined optimisation criterions can also easily and quickly be implemented into the algorithm. When the optimisation process is complete, the resulting control trajectories are exported, transferred, and executed by the experimental system. Within the investigations for this thesis, the OLFO-control was used for the first time to control a process in a scale-down model in that way that the experimental offline data of a lab-scale process are reflecting the data of a large-scale process as close as possible.

6 Simulation study

The simulation study was performed to evaluate the model sensitivity and plausibility with the focus on the newly developed reactor model and the potential for simulating various reactor configurations, mixing times and heterogeneous concentrations. The focus was on these aspects since they were planned to be investigated experimentally in chapters 7 - 9. Four main topics were investigated with simulations of *S. cerevisiae* cultivations using different model configurations:

1. Assessment of the influence of different mixing times on the simulation results for a mathematical model of a two-compartment SDM (chapter 6.2).
2. Evaluation, if different model arrangements resembling single stirred bioreactors and a physical two-compartment SDM lead to comparable simulation results when the mixing time is constant across simulations (chapter 6.3.1).
3. Determination, if the number of STR-models for the connection between the reactors of a two-compartment SDM influences the simulation results (chapter 6.3.2).
4. Comparison of simulation results regarding different DO setpoints in the individual reactors of a two-compartment SDM (chapter 6.4).

The parameters employed for this sensitivity analysis are the resulting mean parameters presented in chapter 7.1.2 identified for *S. cerevisiae* cultivations in a single bioreactor system (listed in Table 22 in the appendix).

6.1 Baseline simulations for an ideally mixed system

To investigate the main objectives, baseline simulations were performed for ideally mixed systems. Two simulations in batch mode and two with glucose feeding were carried out for cultivations under aerobic and anaerobic conditions. The initial concentrations for the simulations for the batch cultivations were set to 50 g L^{-1} for glucose, 20 g L^{-1} for nitrogen source and 2 g L^{-1} for biomass. Figure 28 presents the simulation results of two batch processes under aerobic (left side) and anaerobic conditions (right side).

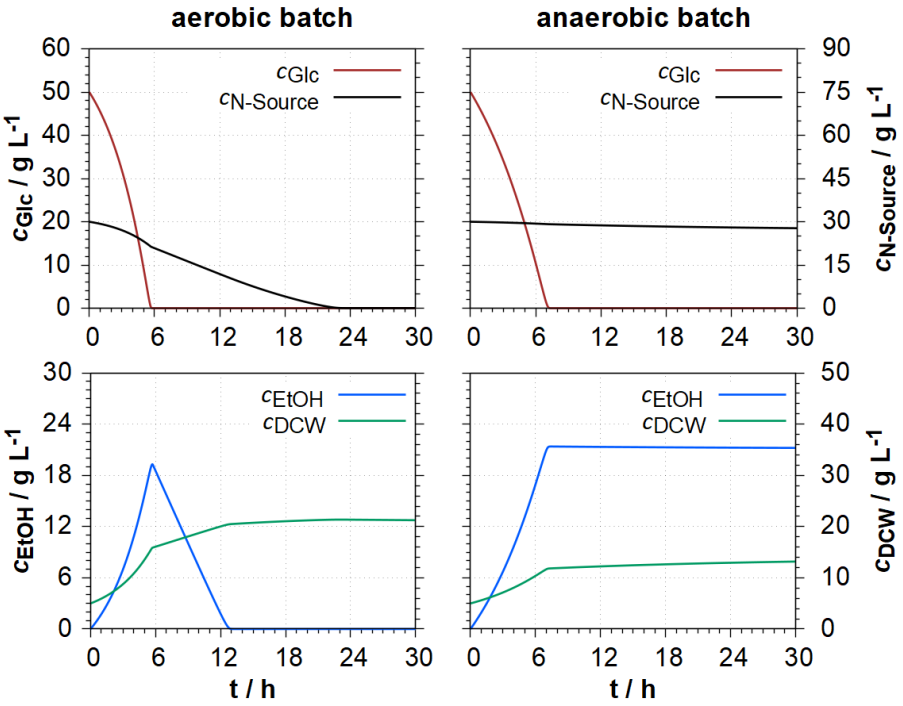


Figure 28: Simulations of two batch processes in an ideally mixed bioreactor under aerobic (left column) and anaerobic conditions (right column).

In the simulation of the cultivation under aerobic conditions, glucose is completely consumed at 5.5 h, whereas in the anaerobic process glucose is depleted at 7 h. Biomass is produced up to a density of 21 g L^{-1} in the aerobic process and 13 g L^{-1} in the anaerobic process. Ethanol is produced up to 20 g L^{-1} under aerated conditions and even 21 g L^{-1} under anaerobic conditions reaching its maximum concentrations at 5.5 h (aerobic) and 7 h (anaerobic). As expected, ethanol is only consumed under aerobic conditions and is completely consumed at 13 h.

A fictitious feeding strategy was designed which was used for all simulated fed-batch processes in this simulation study. A linearly rising feeding strategy was selected since it should initially lead to increasing glucose concentrations. These concentrations should be above the limit for the incomplete oxidation of glucose to ethanol (Crabtree effect). This glucose concentration above which this incomplete oxidation of sugar to ethanol takes place is specified in the literature between 3.6 mg L^{-1} to 210 mg L^{-1} (Woehrer and Roehr 1981; Rose 1993; Leuenberger 1972); the model parameter for this concentration was set to 155 mg L^{-1}. Since the yeast production is expected to be close to exponential growth, glucose concentrations are expected to decrease after some time until they reach values close to 0 g L^{-1} which should cause ethanol uptake (diauxic metabolism). In the model, diauxic metabolism starts at glucose concentration below 0.06 g L^{-1}. Both parameter values and the associated double-sigmoid functions have not been changed during the parameter identifications (chapters 7.1.2, 8.3.1 and 9.3.2).

The feeding rate was linearly rising with adaptions every 4 hours. The initial volumetric feed rate was set to 0.025 L min^{-1} and the final rate to 0.3 L min^{-1} at 48 h. The glucose and nitrogen concentrations of the feed were set to 280 g L^{-1}.

Figure 29 presents the simulation results of two fed-batch processes under aerobic (left side) and anaerobic conditions (right side).

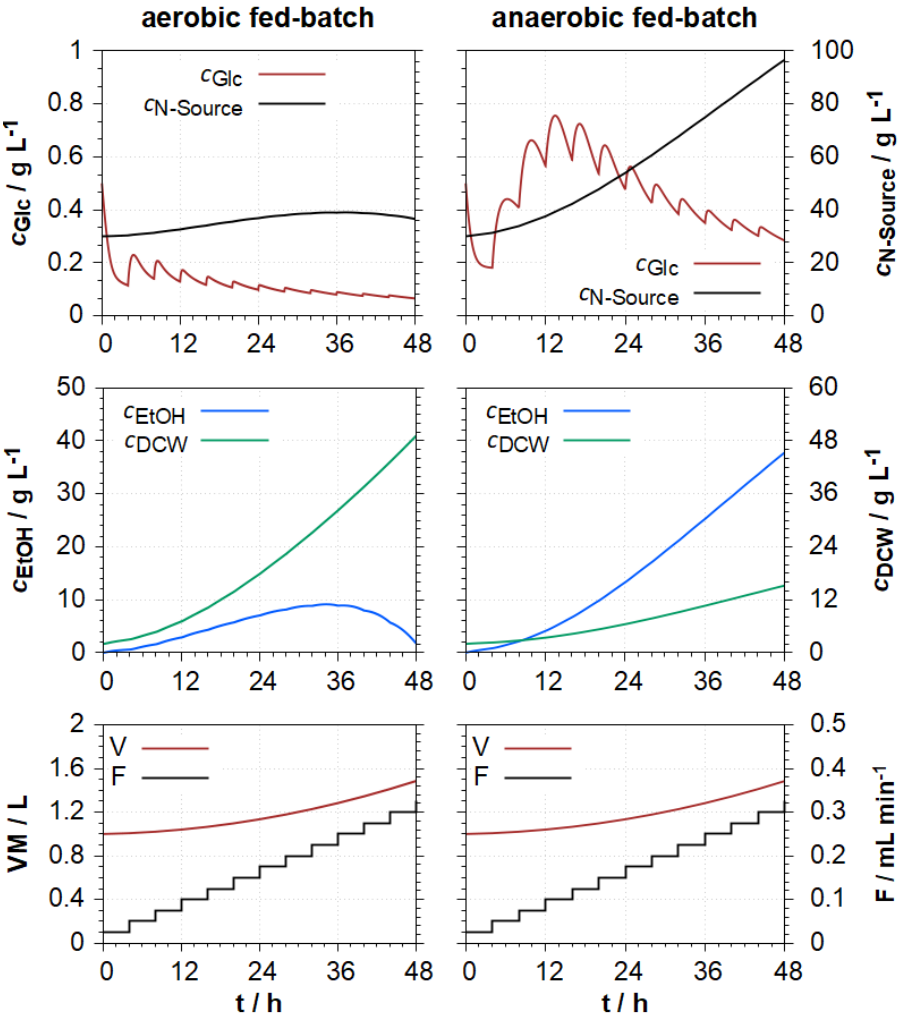


Figure 29: Simulations of two fed-batch processes in an ideally mixed bioreactor under aerobic (left column) and anaerobic conditions (right column).

Glucose concentrations are lower in the aerobic process. The biomass is increasing up to a density of 50 g L^{-1} under aerobic conditions and only to 15 g L^{-1} in the anaerobic process. The ethanol concentration reaches a maximum of 9 g L^{-1} at 34 h under aerated conditions and is consumed to 2 g L^{-1} at the end of the process simulation (at 48 h). In the simulation for anaerobic conditions 38 g L^{-1} of ethanol is produced until the end of the process.

The results of these four simulations are plausible.

6.2 Influence of different mixing times

Three simulations with identical initial settings were performed for fully aerobic batch experiments in the SDM to determine the influence of different mixing times. The model is considering the piping between the reactors as two connected STR models per direction resulting in six STR-models. The initial concentrations were set to 50 g L^{-1} for glucose, 20 g L^{-1} for nitrogen source and 2 g L^{-1} for biomass. Three simulations are compared with mixing times $t_{M,95}$ of 26 s, 108 s and 552 s (Figure 30).

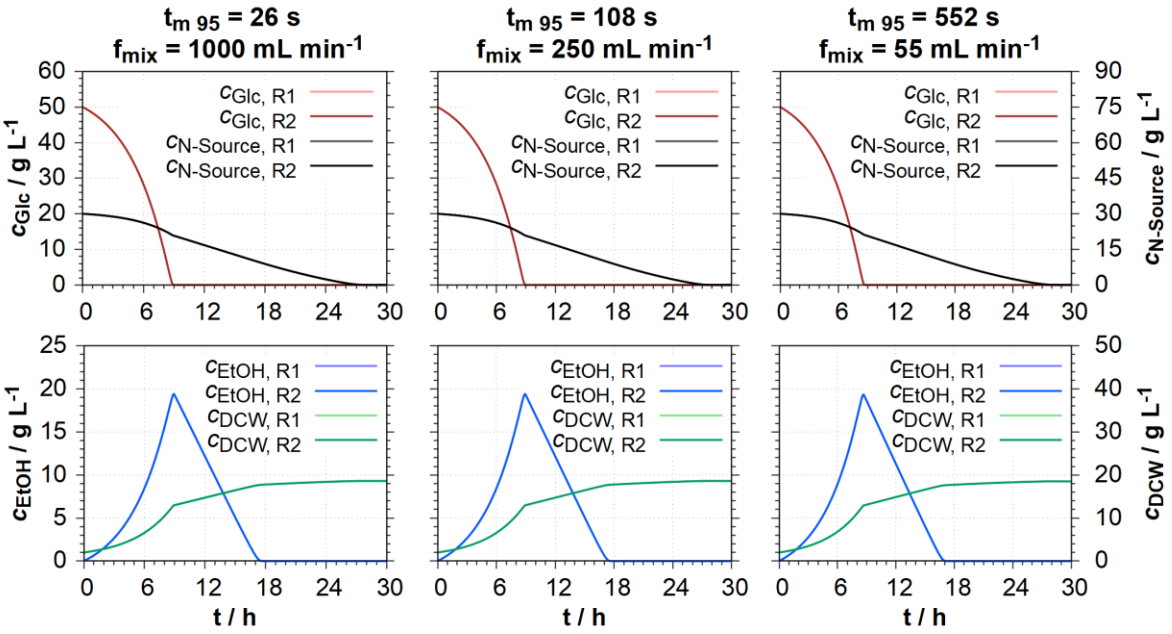


Figure 30: Simulations of three batch experiments in a two-compartment system at different mixing times ($t_{m,95}$) of 26 s, 108 s and 552 s.

The concentrations for all metabolites and the biomass densities are identical in both reactors. Therefore, different mixing times have no influence on the simulation results for batch processes. The concentration profiles are also identical to the aerobic batch simulation of chapter 6.1, which is expected and plausible, since no heterogeneity is induced to the system. Since there are no gradients in the system, the SDM behaves like one ideally mixed stirred tank reactor or system.

The next step was the introduction of a substrate feed, which feeds into STR 1, while the volume in STR 2 was kept constant (see Figure 31). The feeding strategy is identical to chapter 6.1 starting at 0.025 mL min^{-1} and rising every four hours until reaching 0.3 mL min^{-1} after 44 h. The glucose and nitrogen concentrations in the feed were set to 280 g L^{-1}. The initial glucose concentration was set to 0.5 g L^{-1} for these simulated fed-batch processes.

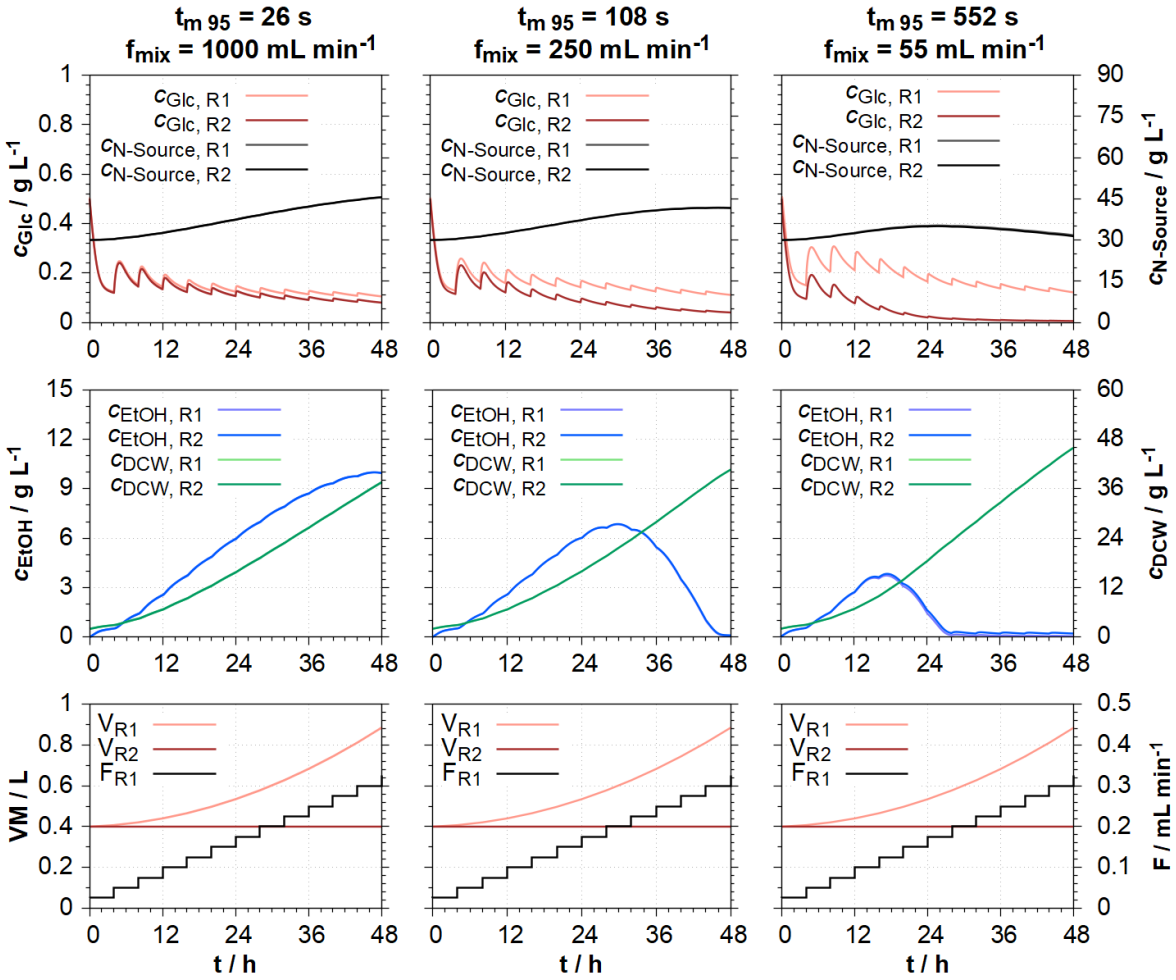


Figure 31: Simulations of three fed-batch experiments in a two-compartment system at different mixing times ($t_{m,95}$) of 26 s, 108 s and 552 s.

The difference of glucose concentrations between the compartments is increasing with longer mixing times. The simulated ethanol production is lower in systems with a longer mixing time while producing a higher biomass density. Ethanol is produced up to a concentration of 9 g L^{-1} in the simulation with the lowest mixing times ($t_{M,95}$ = 26 s) and only up to 4 g L^{-1} when the mixing times is twenty times longer ($t_{M,95}$ = 552 s). The final biomass density is 10% higher in the system with the longest mixing time (52 g L^{-1}) compared to the shortest mixing time (48 g L^{-1}). The simulated concentration curves of the simulation with the shortest mixing time of 26 s are similar to the courses of the ideally mixed fed-batch process in chapter 6.1, which is both expected and plausible, as the system approaches an ideally mixed system when the mixing time decreases.

The increasing glucose difference between the compartments causes a simultaneous uptake of glucose and ethanol in the simulation which resulted in lower ethanol concentrations at longer mixing times. This effect could occur in cultivations performed in the physical scale-down model as well as in large-scale systems (Figure 32).

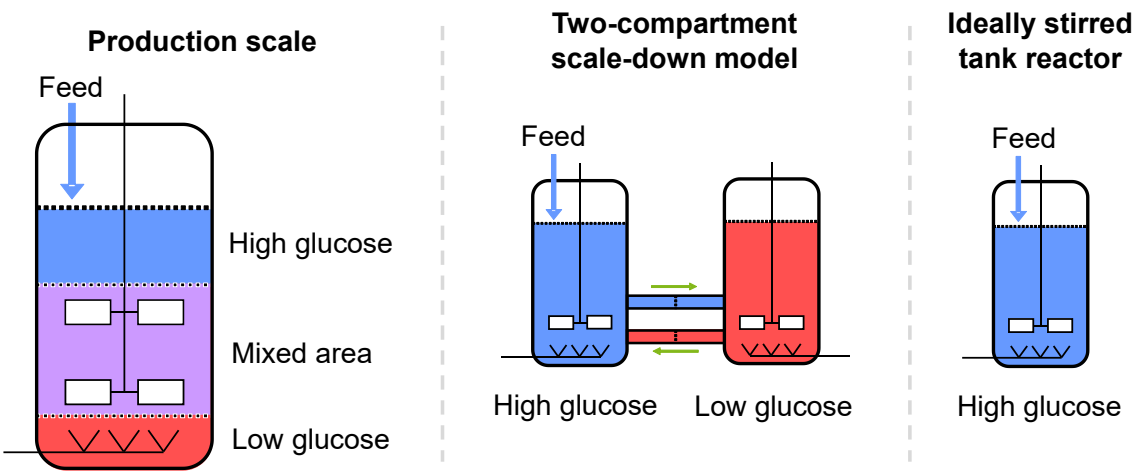


Figure 32: Simultaneous consumption and production of ethanol in systems with heterogenous glucose concentration in zones with overflow metabolism (ethanol production, blue), oxidisation of ethanol (ethanol uptake, red) and possible mixed areas where both metabolisms could simultaneously occur (purple).

The glucose concentration in the production scale is expected to be above the limit of overflow metabolism (ethanol production) in the top area where the feed inlet is located and below the limit in areas at the bottom, which might lead to ethanol uptake (Larsson et al. 1996; Bylund et al. 1998). Around the stirrers is a well-mixed area, where ethanol production and uptake might be possible. The SDM might have one reactor with glucose concentrations above the overflow limit (resulting in ethanol production) and the other one with lower concentrations below the limit needed for the consumption of ethanol, whereas one ideally stirred tank reactor either produces or consumes ethanol. It is also possible that ethanol is produced or consumed in both reactors of the SDM. Both the production scale and the SDM might result in simultaneous uptake and production of ethanol.
In the simulations of the SDM, longer mixing times are causing increasing glucose concentration differences between the reactors. This induces an earlier start of decreasing ethanol concentrations (at 14 - 16 h for 552 s mixing time), since the glucose concentration in the reactor without feeding decreases below the Crabtree limit earlier in systems with longer mixing times. In the process with a mixing time of 26 s the ethanol concentration peak is reached later at 34 h.

6.3 Comparison of reactor configurations

For the investigation of the newly developed reactor submodel consisting of multiple ideally stirred tank reactor models (STR models), simulations of different reactor configurations were compared.

First, three different reactor configurations were simulated and compared (chapter 6.3.1). In chapter 6.3.2, simulations of three model configurations consisting of two, six and 12 STR models resembling a scale-down system consisting of two connected stirred bioreactors were compared.

6.3.1 Reactor configurations

With the developed model structure (chapter 5.2) it is possible to approximate the hydrodynamics within various reactor configurations and types. The aim of this simulation study is to determine if different reactor designs and configurations lead to comparable simulation results of *S. cerevisiae* cultivation when the mixing times are identical. Three models with different arrangements of STR models but identical mixing times are simulated and compared:

1. 24 STR models (zones) connected to resemble the flow characteristics in a stirred bioreactor (rotational and vertical flow, model 1).
2. 12 STR models vertically arranged in a stirred bioreactor (vertical flow, model 2).
3. 12 STR models connected to simulate two-connected stirred bioreactors and the connections between the reactors (two reactors, model 3).

The volumetric flows between STR models of the first model correspond in principle to those in a bioreactor with one Rushton turbine, as they were described, for instance, in the work of Delafosse et al. (Delafosse et al. 2014). Figure 33 presents the flow directions of this model consisting of 24 interconnected ideally STR models (Figure 33; left side). The stirrer would be located in the lowest layer. The right side of Figure 33 shows the configuration of the second model consisting of 12 vertically arranged zones.

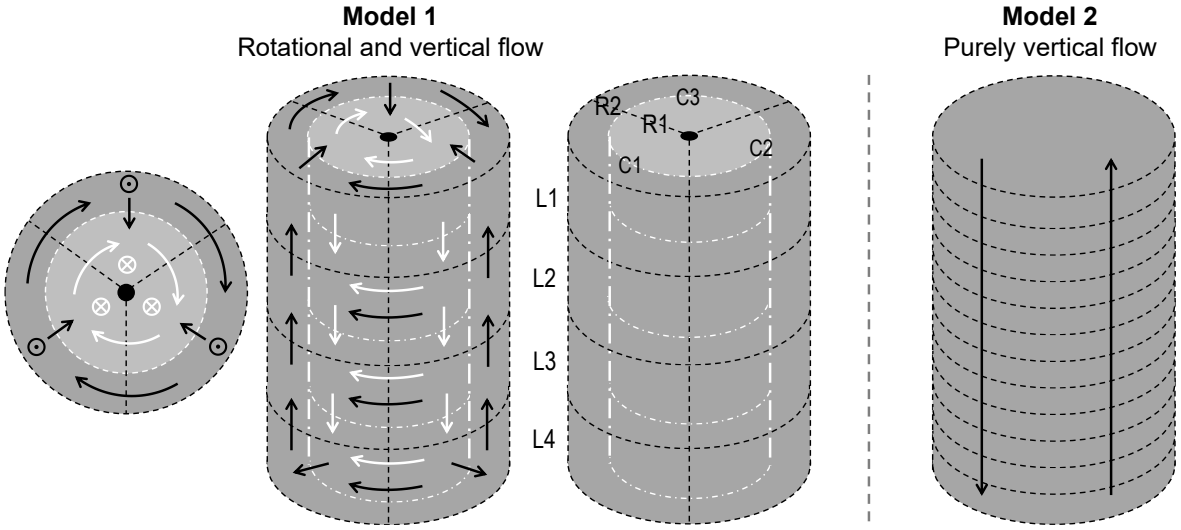


Figure 33: Structure of two network of zones models of interconnected STR-models. Left side: 24 ideally mixed tank reactor submodels (24 cells). Inner ring in light grey, flow is indicated with white arrows. Outer ring in dark great, flow is indicated with black arrows. Rotational flow is ten times higher than radial and horizontal flow (model 1). Right side: 12 connected STR models (reactor zones) in a purely vertical arrangement (model 2).

The model with rotational flow consists of two rings with 12 reactor cells each (Figure 33, left side). The general axial flow direction in the inner ring is down, and in the outer ring upwards, which would be the case for Rushton turbines in a stirred tank reactor without baffles (Desouza and Pike 1972). The rotational flow is assumed to be in the direction of the impeller rotation and ten times greater than the radial and axial flow. The simulation of this model is compared to a simulation of model 2 with 12 vertical reactor zones without rotation and only vertical flow (Figure 33, right side) as well as a simulation of a physical two-compartment scale-down model (model 3). Model 3 has one ideal STR model for each reactor and five for each piping simulating the circulation flow, resulting in 12 STR models in total (Figure 34).

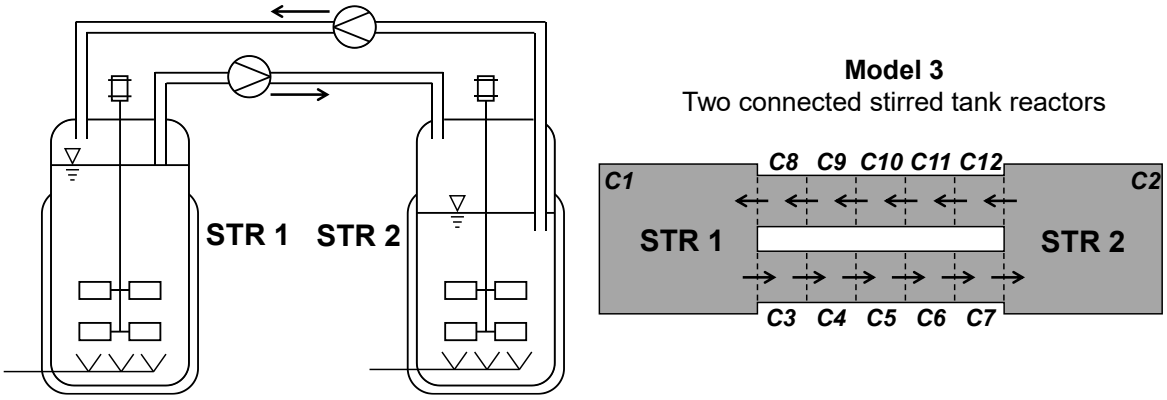


Figure 34: Left side: Schematic of the scale-down model (SDM) with two stirred tank reactors (STR 1 and STR 2). Right side: Structure of the network of zones model with 12 STR-models, one for each STR (C1 & C2) and five for the piping (C3 – C7 & C8 – C12) representing the physical SDM (model 3).

The initial total volume of model 1 and model 2 was chosen to be 1 m³, the volume of all STR models for the scale-down model simulation was chosen to be 1 L, which corresponds to the volume of the physical scale-down model used in this work (chapters 8 and 9). The STRs in the SDM model have a volume of 0.4 L each, each pipe 0.1 L, corresponding to 0.02 L per STR model. All individual STR models in the other two model configurations have the same initial volume (0.0625 m³ for model 1, 0.125 m³ for model 2). In model 1 with rotational flow, all six STR models at the top (L1) increase in volume due to feeding, in models 2 and 3 only the zones where the feeding is located increase in volume (see Figure 35).

The initial biomass density was set to 2 g L^{-1}, the initial glucose concentration to 0.5 g L^{-1}. The feeding strategy is identical to chapter 6.1 with an initial volumetric feed rate of 0.025 m³ min^{-1} (0.025 L min^{-1} for model 3) and a final rate of 0.3 L min^{-1} (0.3 mL min^{-1} for model 3) at 48 h. All simulations were performed, assuming fully aerobic conditions throughout the vessel (DO was set to 20% in all zones). The mixing time is identical for each simulation and is set to about 552 s. This mixing time was deliberately set so high to increase the substrate concentration differences between STR models and amplify the influence of the heterogeneous concentrations on the simulation result. Figure 35 presents a comparison of the resulting simulations.

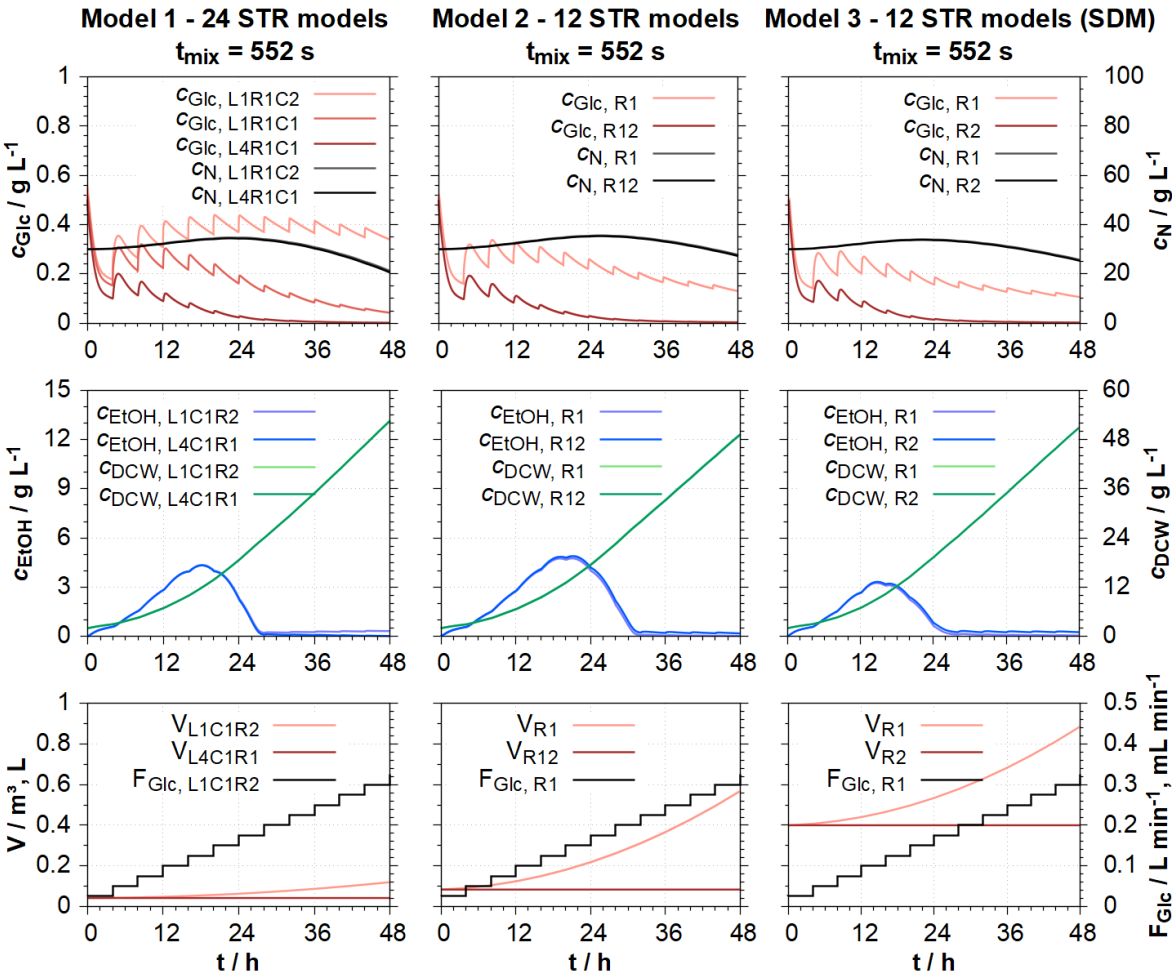


Figure 35: Comparison of three simulations of fed-batch cultivations with different model configurations. 1st column: Network of zones model with 24 compartments situated on four levels, two rings and three cells per ring (model 1). 2nd column: Reactor zone model with 12 compartments and only vertical flow (model 3). 3rd column: Scale-down model with two compartments (model 2).

The glucose concentration differences between the simulations for zones with low overall concentrations are small, whereas the differences in zones with higher concentrations (with the feeding inlet) are more pronounced. The glucose concentration in the zone with the feeding inlet in model 2 is about 20% higher than in model 3 but over 50% lower compared to model 1. The zone with the feeding inlet in model 1 has a higher glucose concentration compared to every other zone of all three simulations. This increase is mainly caused by the smaller relative volume of this STR-model *L1R1C2* (upmost, outer ring).

After 16 h to 24 h, the zones with the highest glucose concentration are above the limit needed for ethanol production, the zones with the lowest glucose concentrations are below the limit necessary for ethanol uptake in all simulations. Consequently, there is a simultaneous ethanol production and uptake in the simulations.

The ethanol peaks at concentrations of 3.5 - 5 g L^{-1} in all simulations which is about 50% of the maximum ethanol concentration in the baseline simulation of an ideally mixed bioreactor (chapter 6.1).

The different maximum ethanol concentrations (Figure 35) are caused by the differences of the glucose concentrations. The lowest peak ethanol concentration and earliest time is in the SDM simulation with 3.5 g L^{-1} at 16 h, the highest peak ethanol concentration with 5 g L^{-1} occurred in the model with purely vertical flow at 20 h. In each simulation, the biomass density increases to about 49 – 52 g L^{-1} (model 1: 52 g L^{-1}, model 2: 49 g L^{-1}, model 3: 50 g L^{-1}).

Overall, the impact of the different designs on the simulation results is relatively low when the mixing time is kept constant. These simulation results support the hypothesis, that heterogeneities occurring in large-scale reactors might be replicable with a scale-down model consisting of two compartments if identical mixing times are used.

6.3.2 Number of reactor submodels for the SDM

Three simulations were performed to assess the impact of the number of STR models used for the piping between the two reactors of the scale-down model. If the simulation results are not impacted, this saves computing time, since there is direct correlation between computing time and the number of STR models (compartments).

The two tank reactors were simulated with one STR model for each reactor. For the assessment of the initial situation, model 3 of the previous chapter 6.3.1 was selected with a total of 12 STR models (five for each connection, Figure 36 left). Two further models were selected, a model with 6 STR models (two for each connection, Figure 36 centre) and a model without the zones for the connection between the reactors (Figure 36 right).

The total volume of the system remained at 1 L; the initial volume of each reactor was set to 0.4 L (0.5 L in the model with 2 STR models). The flow between reactors was set to 250 mL min^{-1} (1/4 of the initial total volume per minute) which resulted in a realistic mixing time (see Vrábel et al. 1999) $t_{M,95}$ of 108 s, which is about one fifth of the mixing time used in the previous section. The volumetric feeding rates started at 0.025 mL min^{-1} and were increased every four hours until reaching 0.3 mL min^{-1} after 44 h (identical to the strategy applied in (chapter 6.3.1)). This feed was fed into STR 1; the volume in STR 2 was constant. The feeds were simulated with glucose concentrations of 280 g L^{-1} (Figure 36).

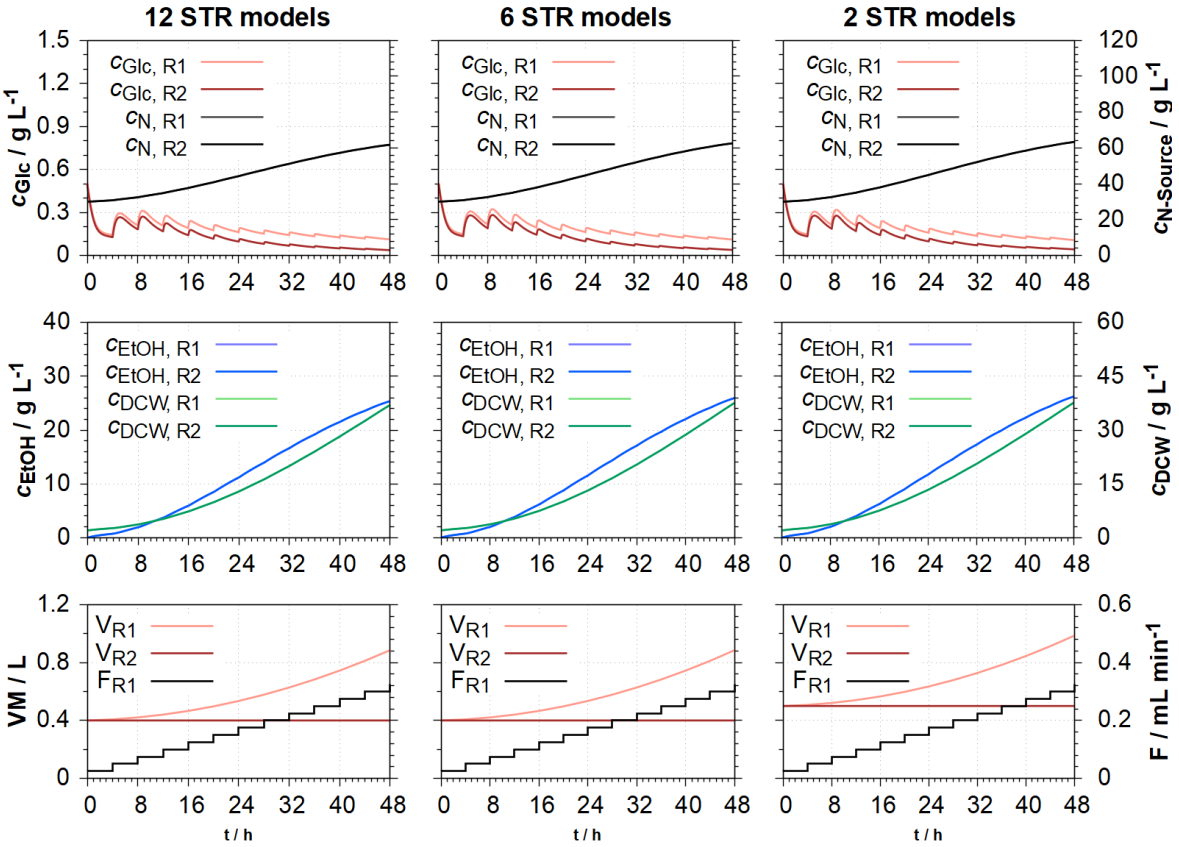


Figure 36: Comparison of three simulations of fed-batch cultivations of the SDM with one STR model for each stirred tank reactor and different numbers of STR models for the piping between reactors. 1st column: 12 STR models (five for each connection). 2nd column: Six STR models (two for each connection) 3rd column: Two STR models (without a simulation of the connection).

The volumes in R1 and R2 are lower for six and twelve compartments since part of the medium is in the piping between the reactors. The concentration courses are almost identical between the simulations. Therefore, if the volumetric flow rates of the circulation flow between the reactors and the mixing times are identical, the STR models for the connections can be omitted without influencing the simulation results.

6.4 Comparison of heterogenous DO concentrations in the SDM

The first experimental study (chapter 8) with the scale-down model utilises an aerobic and an anaerobic reactor to imitate heterogeneous dissolved oxygen concentration. The influence of the heterogeneous DO concentration on the process performance might be relevant for cultivations in industrial reactors (chapter 2.3.1).

The simulations with this DO heterogeneity were performed with two STR models (zones); one for each reactor of the SDM. For the simulation, the flow rate between reactors was set to the rates from the previous chapter 6.3.2 at 250 mL min^{-1}. The initial volume was set to

0.5 L in each reactor which resulted in a mixing time of 108 s. This is the median mixing time of the previous chapter 6.2. The initial biomass density for the batch processes was set to 5 g L^{-1} and for the fed-batch processes to 2 g L^{-1}; the initial glucose concentration to 50 g L^{-1} for the batch experiments and to 0.5 g L^{-1} for the fed-batch processes.

Heterogeneous DO concentrations are also experimentally investigated in a scale-down system consisting of two STRs (chapter 8). Three simulations of batch processes with different DO setpoints (30% for aerobic conditions, 0% for anaerobic), for an aerobic-aerobic, an aerobic-anaerobic, and an anaerobic-anaerobic scale-down system were performed (Figure 37).

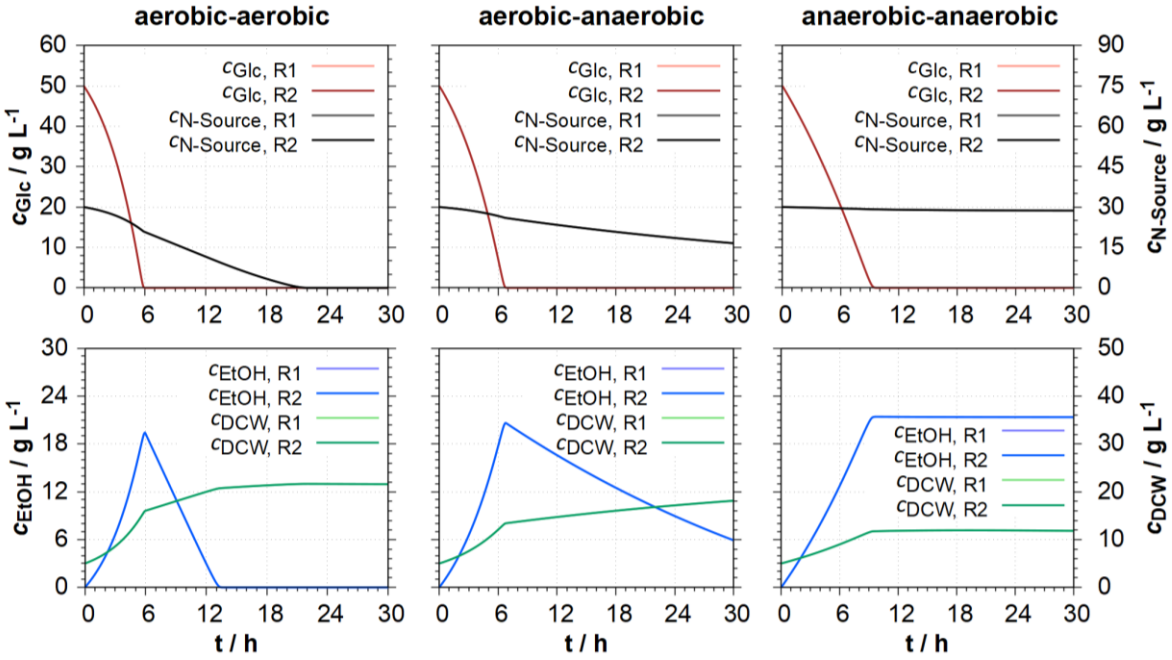


Figure 37: Comparison of three simulations of batch cultivations. 1st column: aerobic-aerobic two compartment scale-down model. 2nd column: aerobic-anaerobic configuration. 3rd column: anaerobic-anaerobic configuration.

The concentrations between the two reactors are identical in system with a homogenous DO (left and right side of Figure 37). For the process in the aerobic-anaerobic system this difference is about 0.1%.

The simulations demonstrate that approximately 20% less biomass is produced in the system with heterogeneous DO concentration compared to fully aerobic conditions. This reduction was also shown by Sweere et al. for fluctuating oxygen profiles in a single reactor setup without considering heterogeneous conditions within the reactor (Sweere and Mesters et al. 1988). In the anaerobic process only half the biomass is produced compared to fully aerobic conditions. In the aerobic-aerobic simulation the ethanol is fully consumed after 8 h, whereas the cultivation with heterogeneous DO concentrations takes about

5-times as long. This model hypothesis was subsequently tested experimentally (see chapter 8.1.1).

In addition, three fed-batch processes were simulated with a combined glucose and nitrogen-source feed (Figure 38). The initial glucose and nitrogen source concentrations were reduced for the fed-batch simulations. All other initial conditions remained unchanged.

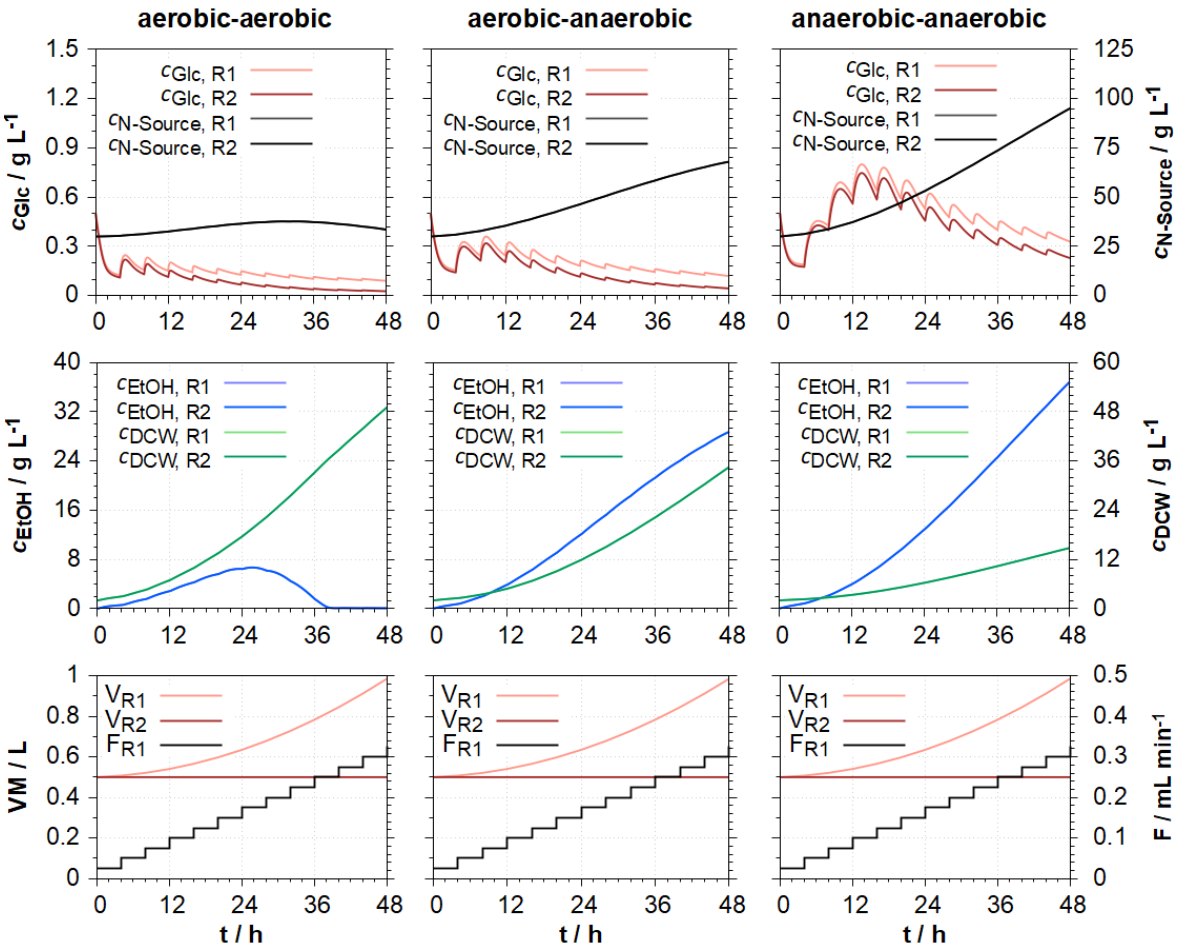


Figure 38: Comparison of three simulations of fed-batch cultivations with a stepwise linearly increasing feed. Feeding is performed into reactor R1. 1st column: aerobic-aerobic two compartment scale-down model. 2nd column: aerobic-anaerobic configuration with the feed located in the aerobic reactor. 3rd column: anaerobic-anaerobic configuration.

The glucose concentration difference between R1 and R2 is about 0.1 g L^{-1} in all simulations, the relative differences for the biomass density and ethanol concentrations are below 0.2%.

In the fully aerobic configuration, glucose concentration drops below the 0.1 g L^{-1} in R1, resulting in ethanol uptake. Ethanol concentration is decreasing until reaching 0 g L^{-1} at 38 h of the fully aerobic process. However, ethanol is not consumed in the system with different DO setpoints since the glucose concentration is not decreasing below the concentration

needed for ethanol uptake (diauxic metabolism). The final biomass density in the fully aerobic configuration is more than 30% higher compared to the aerobic-anaerobic combination (feed in aerobic compartment), and 3.5 times higher compared to the fully anaerobic system.

The simulation results suggest that DO differences resulting from non-ideal mixing conditions may influence the process performance. The experimental verification for batch and fed-batch processes is carried out in chapters 8.1 and 8.2 where experimental data from the SDM with heterogeneous DO concentration are compared to data from single bioreactors.

7 Single bioreactor investigations

The developed mathematical model was first utilised with processes performed in a single bioreactor system, in an ideally mixed, homogeneous environment. The modified mDoE method utilising the mathematical model was tested and evaluated before its application with processes under non-ideally mixed conditions.
The historical data of three processes were used for parameter identification (chapter 7.1). Since the available experimental data is limited, the parameter identification was performed multiple times to quantify the parameter uncertainties (chapter 7.1.2).
Multiple mDoEs with different designs were carried out and compared to a detailed design using 2000 different experimental settings (chapter 7.2.1). The desirability score was calculated targeting biomass production while considering process robustness (chapter 7.2.2). The experimental settings of four experiments were selected based on the desirability scores and experimentally carried out (chapter 7.2.3).
Furthermore, during the application of the mDoE, the mathematical model was tested for its descriptive and predictive capabilities for single bioreactor systems. By applying the mDoE method a simultaneous improvement of the process and the model was achieved.

7.1 Experiments for modelling (A1 – A3)

The experimental data of three cultivations were used for the model parameter identifications and the Monte Carlo based uncertainty quantification. The initial volumes and the compositions of the medium, as well as the feed concentrations, are displayed in Table 10.

Table 10: Initial volume (V) and medium composition as well as feed concentrations. A single feed containing the carbon source glucose (Glc.) and nitrogen source consisting of yeast extract (YE) and peptone (Pep.) was used.

	Initial medium					**Feed**		
	V	Yeast	Glc.	YE	Pep.	Glc.	YE	Pep.
	L	g L^{-1}	g L^{-1}	g L^{-1}	g L^{-1}	g L^{-1}	g L^{-1}	g L^{-1}
A1	1.0	3.0	6.0	1.4	2.3	290.0	50.0	50.0
A2	1.0	3.0	6.0	1.4	2.3	290.0	50.0	50.0
A3	1.0	20.0	30.0	2.9	4.5	290.0	85.0	135.0

The temperature was set to 30 °C in the cultivations (A1 – A3) and the pH was automatically controlled at 5. The volumetric gassing rate was manually adjusted to values between 1

and 2 vvm, with a maximum value of 2 L min^{-1}. The stirrer speed was set to values between 500 and 800 rpm to maintain a DO above 10% (aerobic conditions).

Cultivation A1 utilised a feeding strategy, which was expected to lead to a high biomass density. A2 was performed with an initially low glucose feeding rate expected to lead to no ethanol production in the first 20 h. This feeding rate was increased during the process to reach ethanol inhibition. In A3, a high initial biomass density of 20 g L^{-1} was inoculated to achieve a high final biomass density. The feeding strategies were designed based on process knowledge and the process model developed by S. Brüning (2016) (utilising a parameter set of a different *S. cerevisiae*) as well as adapted during the experiments (Brüning 2016).

These three experiments had different objectives and thus generated different experimental data. These data show several effects that are process-relevant and therefore suitable for parameterising the model. The resulting parameterisations were employed for the applications of the modified mDoEs.

7.1.1 Experimental results

Figure 39 displays the offline data for the biomass density, as well as the glucose and ethanol concentrations of the experiments (A1 – A3). The composition of the exhaust gas and the RQ, as well as the feeding rate, volume, and the pH (controlled at pH 5) are presented.

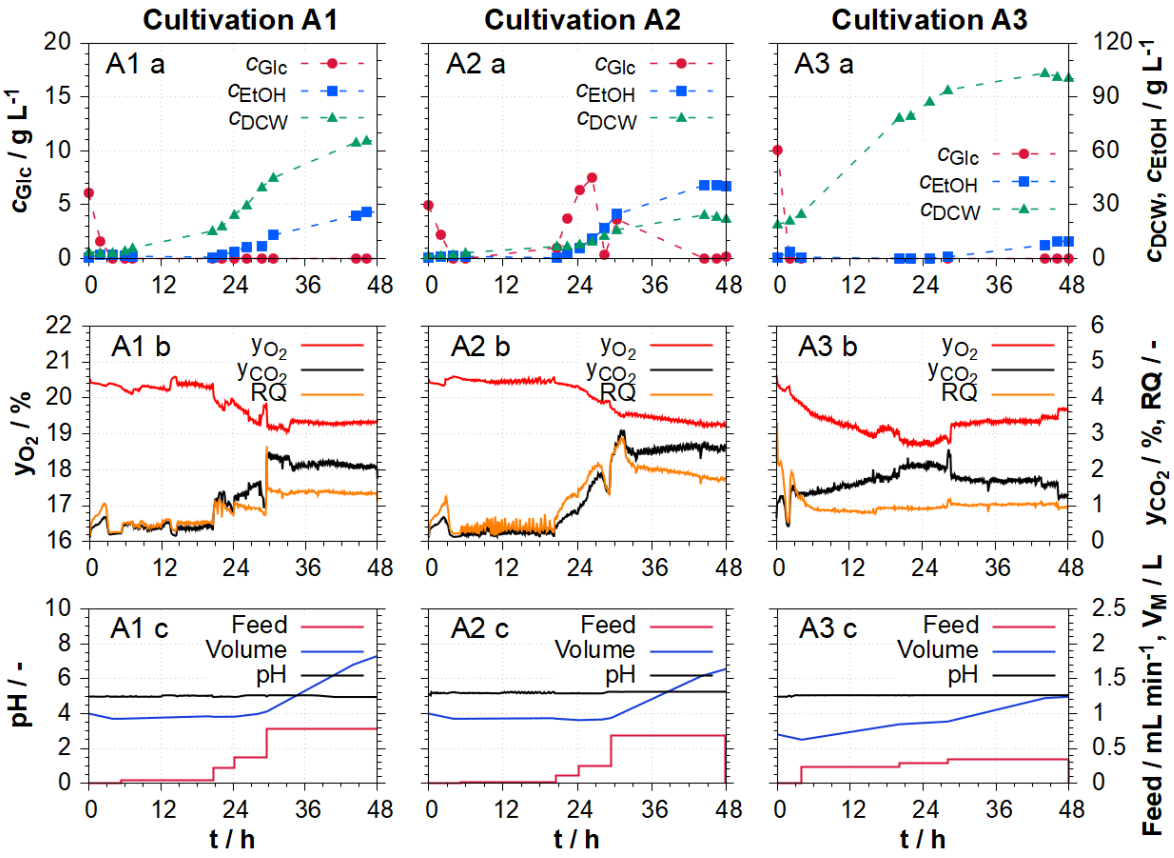


Figure 39: Experimental data of cultivations A1 - A3. 1st row: Offline data for biomass density, glucose, and ethanol concentration. 2nd row: Online data of the off-gas composition (y_{O2} and y_{CO2}) as well respiratory quotient (RQ) calculated from the offgas-data. 3rd row: Theoretical calculated feeding rate and volume, as well as the online measurement of the pH. Experimental settings are explained in chapter 7.1 and the used reactor in chapter 4.3.1.

At the end of the **cultivation A1**, a dry cell weight density of 64 g L^{-1} was achieved. The ethanol concentration reached 28 g L^{-1}. The initial glucose in the medium was fully depleted after 3 h, afterwards the determined glucose concentration was always close to 0 g L^{-1}. The glucose concentrations were below the limit of determination of the used enzymatic test kits. This was the case despite feeding from 3h onwards. The initially low ethanol production rate is reflected by the RQ value, which remains below one for the first 21 h; afterwards, ethanol production starts and the RQ stays above one for the last feed rate increasement at 30 h.

Cultivation A2 was performed with the aim of producing ethanol and reaching ethanol inhibition. Experiment A2 utilised identical initial conditions as A1, but with a feeding strategy designed to initially lead to no ethanol production and after 20 induce overflow metabolism through overfeeding. This should result in ethanol inhibition towards the end of the

cultivation. The feed was chosen to start at a lower rate of about 0.01 mL min^{-1} until 21 h which resulted in a lower biomass density at 24 h compared to A1. After increasing the feed, 40 g L^{-1} of ethanol was produced. This observation was supported by the RQ which exceeded 1 from 21 h and by a glucose concentration peak which reached approximately 8 g L^{-1}. Dry cell weight density reached 24 g L^{-1} at the end of the cultivation.

Cultivation A3 aimed to reach a high biomass density. The dry cell weight increased from initially 18 g L^{-1} to over 100 g L^{-1} at the end of the cultivation. This was achieved until the end of the process within a comparatively short process duration. Glucose concentrations were below the limit of determination once the initial glucose was consumed. The ethanol concentration remained lower compared to the other cultivations reaching a maximum of 10 g L^{-1} at the end of the cultivation. The RQ was constantly close to 1 which is consistent with the observed low ethanol production rate in this cultivation.

7.1.2 Parameter identification and uncertainty quantification

The experimental data of the three processes (A1 – A3) were used for the identification of the model parameters. Since the data amount at this early stage of process development is low, the model parameters were identified 116 times with the application of the Monte-Carlo based uncertainty quantification. This enables a prediction of the process robustness even with low amount of available data at the cost of an increased computational effort.

For each of the 116 parameter identifications, 3000 simulations of the three processes (A1 – A3) were performed, totalling over one million model simulations. This was only possible because the simulations could be performed before the next experiments and with the developed and accelerated model (under 1 s per simulation, see chapter 5.2). Furthermore, this identification step could be parallelised on multiple PCs since each parameter identification is independent from each other. In this work, the parameter identification was performed on three PCs simultaneously.

The experimentally determined offline data, initial concentrations, and control functions of processes A1 - A3 were varied by using (assumed) standard deviations for each of the 116 parameter identifications.

The offline data and the initial concentrations were varied by a standard deviation of 5% around the determined value. Since the biomass compartments cannot be directly determined, a variation of 10% for the initial value for each compartment was assumed. Physico-chemical measurements including pH, DO and temperature, alongside feed volume flows, feed concentrations (glucose and nitrogen source) were also varied by a 5% standard deviation close to the set or determined value.

A total of 18 parameters were selected for repeated parameter identifications. The selected parameters include the description of uptake rates, yield coefficients for metabolic pathways, transfer rates of cell compartments, as well as boundary conditions and strengths between effects like overflow metabolism or ethanol inhibition. The initial values and identified mean values, as well as a short description of each parameter are supplied in Table 22 (in appendix) as well as reasons why they were chosen for the parameter identification (see chapter 12.2.1.1).

Evaluation of parameter identification

The evaluation of the parameter identification is performed graphically (Figure 40) and with the coefficient of determination R^2.

The arithmetic mean of each parameter was calculated, and a mean parameter set of the 116 resulting parameter sets was generated. The simulation with the mean parameter set (continuous black line in Figure 40) and the uncertainty band of the simulated data courses after identification (dashed lines) are displayed and compared to the experimental offline data for cultivations A1 – A3. The uncertainty bandwidth is the area between the 10% and 90% quantiles of the simulations results.

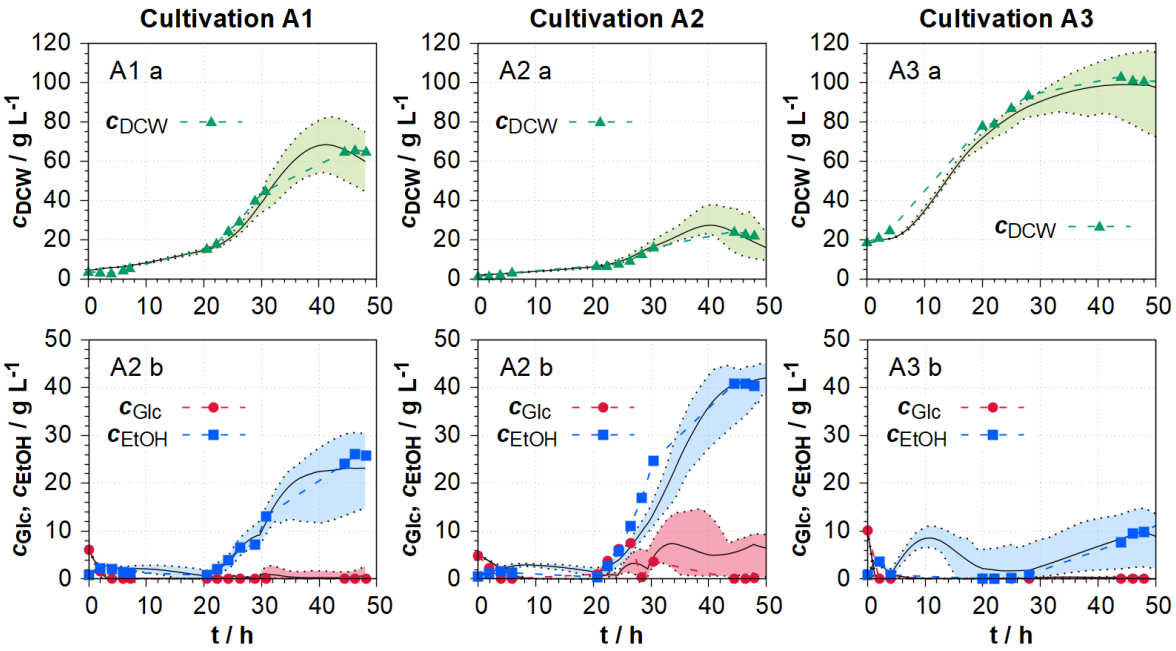


Figure 40: Experimental offline-data (dry cell weight density, glucose concentration, ethanol concentration) of cultivations A1 - A3 in comparison to results after model parameter identification. The continuous line represents the mean of 116 Monte-Carlo parameter identifications, the dashed lines represent the 10% and 90% quantiles of these identifications (based on already published graphics in (Moser et al. 2021)).

The model parameters were adapted successfully to the data of the three experiments with good R^2 obtained with the mean parameter set (A1: R^2 of 0.98, A2: R^2 of 0.85 and A3: R^2 of 0.97). The simulated data for biomass density and ethanol concentration have a high agreement throughout the experiments. The largest difference between the simulated and experimental data occurred in cultivation A2 for the glucose concentration during the phase with ethanol production resulting in a slightly lower R^2. However, the glucose concentration of A2 lies within the uncertainty band.

Overall, the model parameters could be reliably and robustly identified both quantitatively and qualitatively, as the R^2 is above 0.85 and the uncertainty bands are relatively narrow for all three experiments.

7.2 Optimisation of a single reactor fed-batch process

Before the operating conditions for a process under non-ideal mixing conditions were calculated and optimised with mDoE, this optimisation method was first tested on processes performed in a single 1 L bioreactor.

Table 11 displays the planned initial medium composition selected for the mDoE experiments (B1 – B4) along with the compositions of the glucose and nitrogen source feed.

Table 11: Initial volume and medium composition as well as feed concentrations for the mDoE experiments. Separate feeds for the carbon source (glucose) and nitrogen source (yeast extract and peptone) were calculated.

	Medium					F_{Glc}	F_N	
	Vol.	Yeast	Glc.	YE	Pep.	Glc.	YE	Pep.
	L	g L^{-1}	g L^{-1}	g L^{-1}	g L^{-1}	g L^{-1}	g L^{-1}	g L^{-1}
B1 – B4	0.7	2.0	15.0	1.4	2.3	400	130	270

All experiments were inoculated with 2.0 g L^{-1} of dry yeast into an initial medium volume of 0.7 L. The initial medium consisted of 15.0 g L^{-1} glucose, 1.4 g L^{-1} yeast extract and 2.3 g L^{-1} peptone. The glucose feed was planned and carried out with a concentration of 400 g L^{-1}. The nitrogen-source feed was planned and carried out with a yeast extract concentration of 130 g L^{-1} and a peptone concentration of 270 g L^{-1} for all for experiments. The temperature was controlled at 30°C, the stirrer speed was fixed at 800 rpm for all four experiments.

Three factors were selected for the optimisation of a *S. cerevisiae* fed-batch cultivation with the modified mDoE:

1) The set-point of the pH value control
2) The volumetric feeding rate F_{Glc} at the end of the cultivation (t_{end} = 48 h) for a linearly rising glucose feed.
3) The volumetric feeding rate F_N at the end of the cultivation (t_{end} = 48 h) for a linearly rising nitrogen source feed.

The rates were calculated with the following equation (7.1):

$$F_k(t) = \frac{F_{k,end}}{t_{end}} \cdot (t - t_{start}) \quad (7.1)$$

Using this equation, linearly rising feeding strategies starting at 1 h into the cultivation are designed for the glucose and nitrogen source feeds. Linear feeding strategies offer an advantage over exponential feedings since they allow for a more stable process operation and are safer to handle. Although exponential feeds could theoretically lead to better results, due to the expected exponential cell growth at the start of the cultivation. However, exponential feeding rates that are slightly too high can quickly result into overfeeding and overflow metabolism. As a result, cell growth might be inhibited, reinforcing overfeeding.

The parameter sets used for the following simulations during the execution of the mDoE were drawn from the 116 identified parameter sets of the previous chapter with the usage of Latin-Hypercube sampling.

The desirability score D of the experimental conditions in the mDoE was calculated. The calculation consists of the maximisation of the biomass density and the minimisation of the variability of the simulation results regarding the biomass density (see chapter 4.6.3). The maximisation of the biomass density was weighted by 80%, the minimisation of the variability by 20%.

7.2.1 Evaluation of different model-based designs

The process of selecting the right design space is crucial since the choice of the design boundaries can greatly impact the outcome of the DoE. For this reason, multiple different experimental designs were tested and compared. These designs were two Box-Behnken designs, two central composite designs and one design with 27 different factor combinations created with the mDoE software tool.

Each point within the design space was simulated 20 times with different parameter sets. The 20 different parameter sets were drawn by Latin-Hypercube sampling from the 116 identified parameter sets used as the basis. All simulations in this chapter utilised the experimental factors (glucose and nitrogen feeding rates as well as pH), along with the initial conditions introduced in chapter 7.2.

For the first run of each design, higher maximum volumetric feeding rates of up to 1.0 mL min^{-1} for the glucose and 0.6 mL min^{-1} for the nitrogen source feed were chosen. The maximum rates were halved for the second design. The lowest feeding rate were set at 0.1 mL min^{-1} for both design spaces. The simulations within first design space were simulated with pH values ranging from 3 to 7, whereas the second design had narrower boundaries ranging from 4 to 6.

Figure 41 presents the calculated desirability scores for the two Box-Behnken designs for 3D-plot, 2D-plots, and response surfaces of a quadratic function.

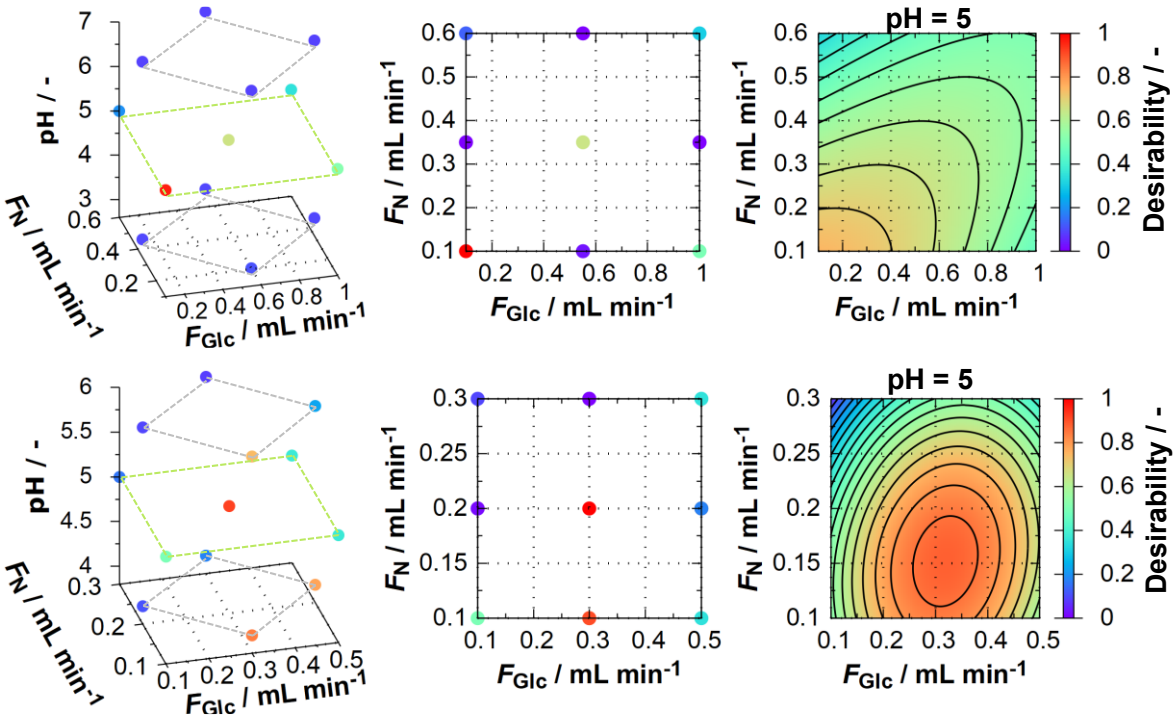


Figure 41: Desirability scores for two mDoE-Runs using Box-Behnken designs are presented for pH value, as well as the final feeding rates F_N and F_{Glc}. The top row has wider boundaries than the bottom row. 1st and 2nd column: 3D and 2D scatter plots; 3rd column: Visualisation of the response surfaces for the fitted quadratic functions for pH 5.

The simulation results and calculated desirability scores are highly dependent on the chosen boundaries of the designs. In the first design the best process is expected to be found in the lower left corner, while in the second design it is expected to be located in the lower centre. The limited number of simulated experimental settings renders the localisation of the optimum (within the design space boundaries) of the experiment (or simulation) difficult to impossible.

Figure 42 presents two mDoE-runs with the same boundaries for central-composite designs.

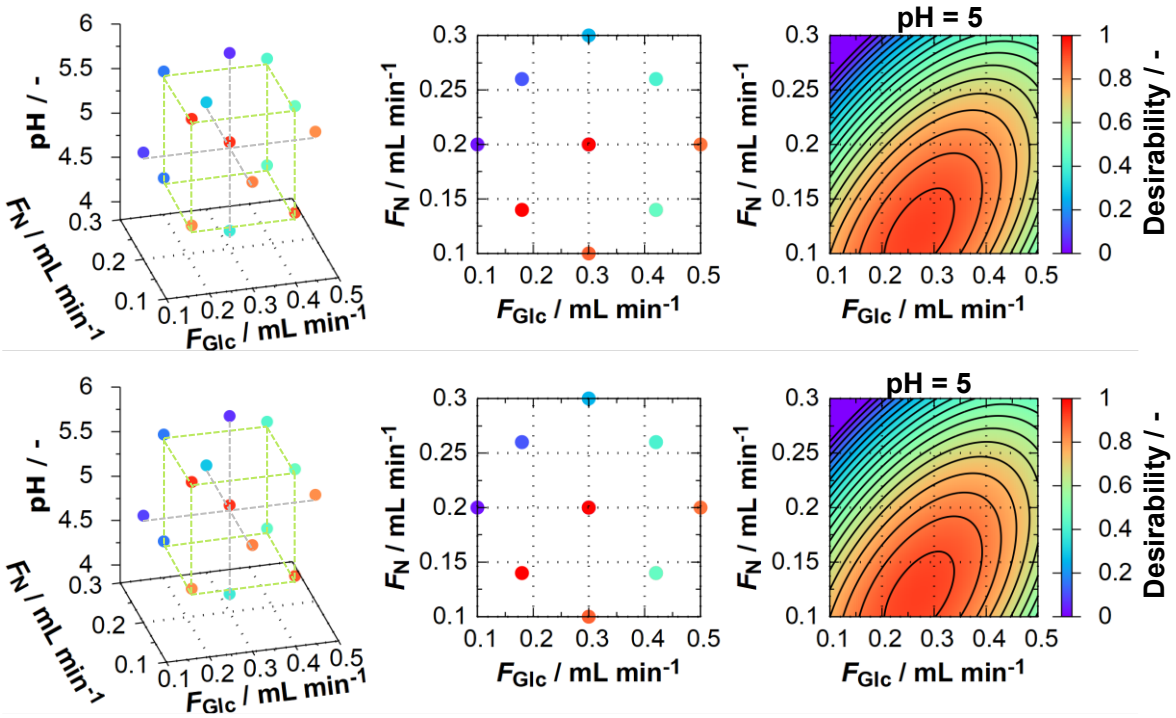


Figure 42: Desirability scores for two mDoE-Runs using inscribed central-composite designs are presented for pH value, as well as the final feeding rates F_N and F_{Glc}. The top row has wider boundaries than the bottom row. 1st and 2nd column: 3D and 2D scatter plots; 3rd column: Visualisation of the response surfaces for the fitted quadratic functions for pH 5.

The central composite design produced more distinct simulated optimum areas in the response surfaces, as evidenced by the stronger gradients and smaller area of the optimal area. This more pronounced optimum region compared to the Box-Behnken design could be attributed to the different type of DoE design or on the chosen boundaries. The true optimal region could still be missed by the central-composite design, since the empty space between design points is still large.

A design space consisting of 27 different factor combinations was created with the mDoE software tool (chapter 5.3). This corresponds to a full factorial design space with three levels (3 k factorial design). Each point was again simulated 20 times using parameter sets determined by Latin-Hypercube sampling. Figure 43 presents the resulting desirability scores for each point in the design space and quadratic response surfaces displayed for pH 5 (a), as well as the final volumetric rate (at t = 48 h) for the nitrogen source (F_N = 0.18 mL min^{-1}) (b) and glucose feeds (F_{Glc} = 0.41 mL min^{-1}) (c).

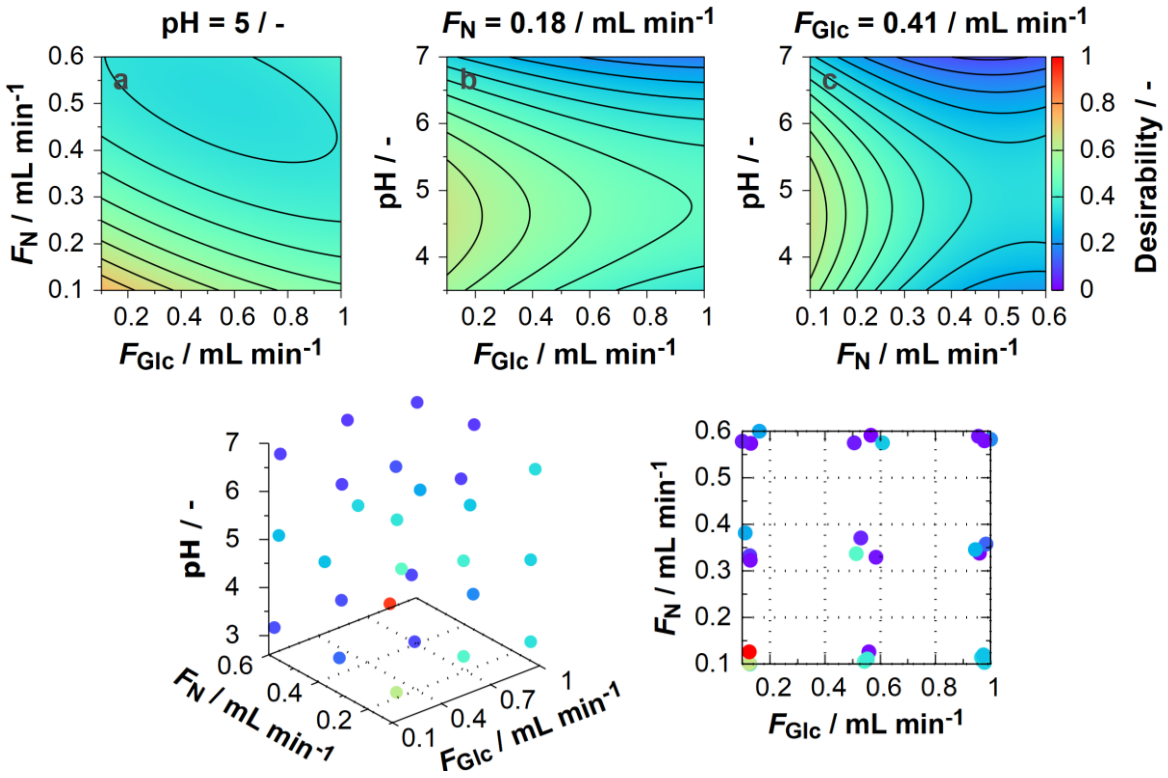


Figure 43: 1st row: Visualisation of the quadratic surface response function shown for a) pH = 5, b) F_N = 0.18 mL min^{-1} and c) F_{Glc} = 0.41 mL min^{-1}. 2nd row: Desirability scores for the mDoE run with 27 factor combinations (corresponding to a full-factorial design space) for pH value, as well as the final feeding rates F_N and F_{Glc}.

The optimal factor combination (experimental setting) is predicted to be pH 5 and at the border of the design space at low nitrogen and glucose feeding rates. This is evident in both the direct graphical visualisation of the desirability score of each point as well as in the quadratic response surfaces. The application of the full-factorial design (together with the results of chapter 7.2.2) was part of the publication that focused on the application of the mDoE-Toolbox (Moser et al. 2021).

In all presented designs the empty space between points of experimental settings was leading to uncertainties since the real optimum area could be located between points. Furthermore, the response surfaces and simulated optimum areas are clearly impacted by the selected design. Since the resulting response surfaces and their optimum areas are depending on the chosen design, a selection on this basis would lead to vastly different experiments being selected and carried out. Therefore, using classical designs, optimal experimental factor combinations can only be determined with limited reliability.

The uncertainty of identifying factor combinations leading to optimum process conditions could potentially be reduced by the simulation of an increased number (> 1000) of

experimental settings in the design space (chapter 7.2.2). This is now feasible since simulations can be completed in about a second.

7.2.2 Experimental design space and evaluation

Since the traditional experimental designs exhibited various positions of the optimal area depending on the design, a comprehensive mDoE with a total of 2000 experimental settings was carried out.

For each point 20 simulations with differing parameter sets were performed, resulting in a total of 40000 simulation runs. This required a computation time of about 14 hours, which is the main disadvantage of this method. However, this method results in an approximation of the response surface in the hyperspace by simulating many factor combinations. The main advantage is the provision of a clear view of the optimal area of the design space including its shape, rotation, and gradients. Optimal settings within the design space are defined as settings that lead to high results of the desirability function (D > 0.8).

Figure 44 illustrates the desirability scores of each point for the detailed run in a 3D view on the left side. The right side displays a 2D section of the desirability scores between pH levels of 4.5 to 5.5, which corresponds to the location of the optimal area.

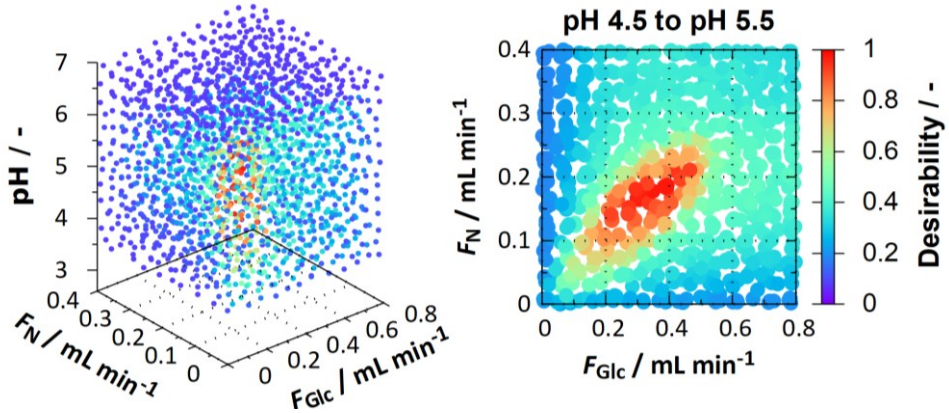


Figure 44: Desirability scores of 2000 experimental settings in a 3D view and a 2D slice between pH 4.5 and 5.5.

The area comprising experimental settings with high predicted desirability scores (red coloured points with scores close to one) is clearly visible. It can therefore be concluded that the optimum of the mDoE can be identified more reliably if a significantly higher number of factor combinations are calculated within the design space.

The best single simulated experiment is at F_{Glc} = 0.40 mL min^{-1}, F_N = 0.19 mL min^{-1} and pH 5.1. This experiment displays the highest average final viable biomass and a relatively low variability. This strongly corresponds to the optimum of the fitted quadratic function (equation 4.45) which has nearly the same settings (F_{Glc} = 0.41 mL min^{-1},

F_N = 0.18 mL min^{-1}, pH 4.9) (Figure 45). The parameters of the quadratic function were identified with the desirability scores of the 2000 different simulated experiment settings.
Since Figure 44 visualises the location and shape of the optimum area with higher detail and accuracy, the calculation of the quadratic function for the visualisation of the response surface is not necessarily needed. For comparison purposes, the response surface is shown below in Figure 45.

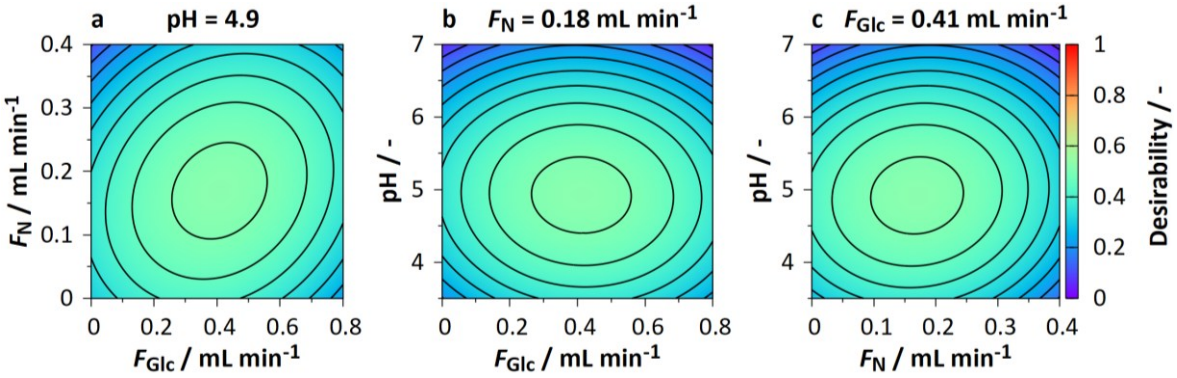


Figure 45: Response surfaces for determined optimum process conditions at pH 4.9, max. nitrogen feed at 0.18 mL min^{-1} and max. glucose feed at 0.41 mL min^{-1}. 2000 different experimental settings were simulated in the design space and used for the identification of the parameters of the quadratic function.

The location of the optimal area of the response surface is close to the area of high desirability scores (see Figure 44). The area with the highest desirability score has lower maximum scores (D ≈ 0.6) compared to the results of the individual factor combination where the highest scores reach values of about 0.92.
The response surfaces (Figure 45) exhibit lower maximum values than those observed in the response surfaces of the Box-Behnken and Central-Composite designs (chapter 7.2.1), which both reached maximum desirability scores of approximately 1.0. This discrepancy may be attributed to the high number of experimental settings which impacts the fitting process of the quadratic function.
The rotation and shape of the optimum area differs greatly from Figure 44. With a large number of points, the detailed representation of the response surface is therefore preferable. It's preferable since the location and shape of the optimum area can be derived directly from the results of the individual experimental settings without alterations (see Figure 44).
Based on Figure 44 (and Figure 45) four factor combinations were chosen for experiments. The corresponding desirability scores (ranging from 0.48 to 0.92) of these experiments are listed in Table 12.

Table 12: Chosen factor combinations determined with the mDoE. F_{Glc} and F_N are the final feed rates of a linearly rising feed at the end of the process simulated accordingly to equation (7.1) as well as the desirability score of each planned experiment.

	F_{Glc} mL min^{-1}	F_N mL min^{-1}	pH -	Desirability score -
B1	0.41	0.18	5	0.92
B2	0.28	0.12	6	0.88
B3	0.28	0.12	4	0.89
B4	0.54	0.25	5	0.48

The settings of the first experiment B1 are chosen at the optimum of the quadratic function (F_{Glc} = 0.41 mL min^{-1}, F_N = 0.18 mL min^{-1}, pH 5.0) with the highest desirability score of 0.92. For the next two experiments (B2 and B3), settings were selected in the centre of the optimal region. In this region, the average final biomass density is expected to be lower compared to B1. However, the variability is also lower which should lead to a more robust process operation, given the lower impact of small deviations and disturbances on the simulated process performance. This results in slightly lower desirability scores of 0.88 and 0.89 compared to B1. Two experiments with different controlled pH levels (pH 4 and pH 6) at lower feed rates (F_{Glc} = 0.28 mL min^{-1} and F_N = 0.12 mL min^{-1}) compared to B1 were selected and performed.

The final experiment B4 was performed at higher glucose and nitrogen-source feeding rates (F_{Glc} = 0.54 mL min^{-1} and F_N = 0.25 mL^{-1}) at the edge of high desirability scores resulting in a lower desirability score of 0.48. These experimental settings of B4 were chosen to assess if the edge of the optimum area was predicted accurately (see Figure 44).

7.2.3 Performed experiments (B1 – B4)

The experimental settings (F_{Glc}, F_N, pH value) of four cultivations (Table 12) were selected utilising mDoE and carried out in the single MDX bioreactor. Figure 46 presents the comparison between experimental and simulated data for the offline data of c_{DCW}, c_{Glc} and c_{EtOH} including uncertainty prediction bands as well as the online data for the exhaust gas composition. The fourth row of Figure 46 presents the mDoE factors, the pH value (online data) and the time courses of F_{Glc} and F_N (control values) as well as the calculated medium volume.

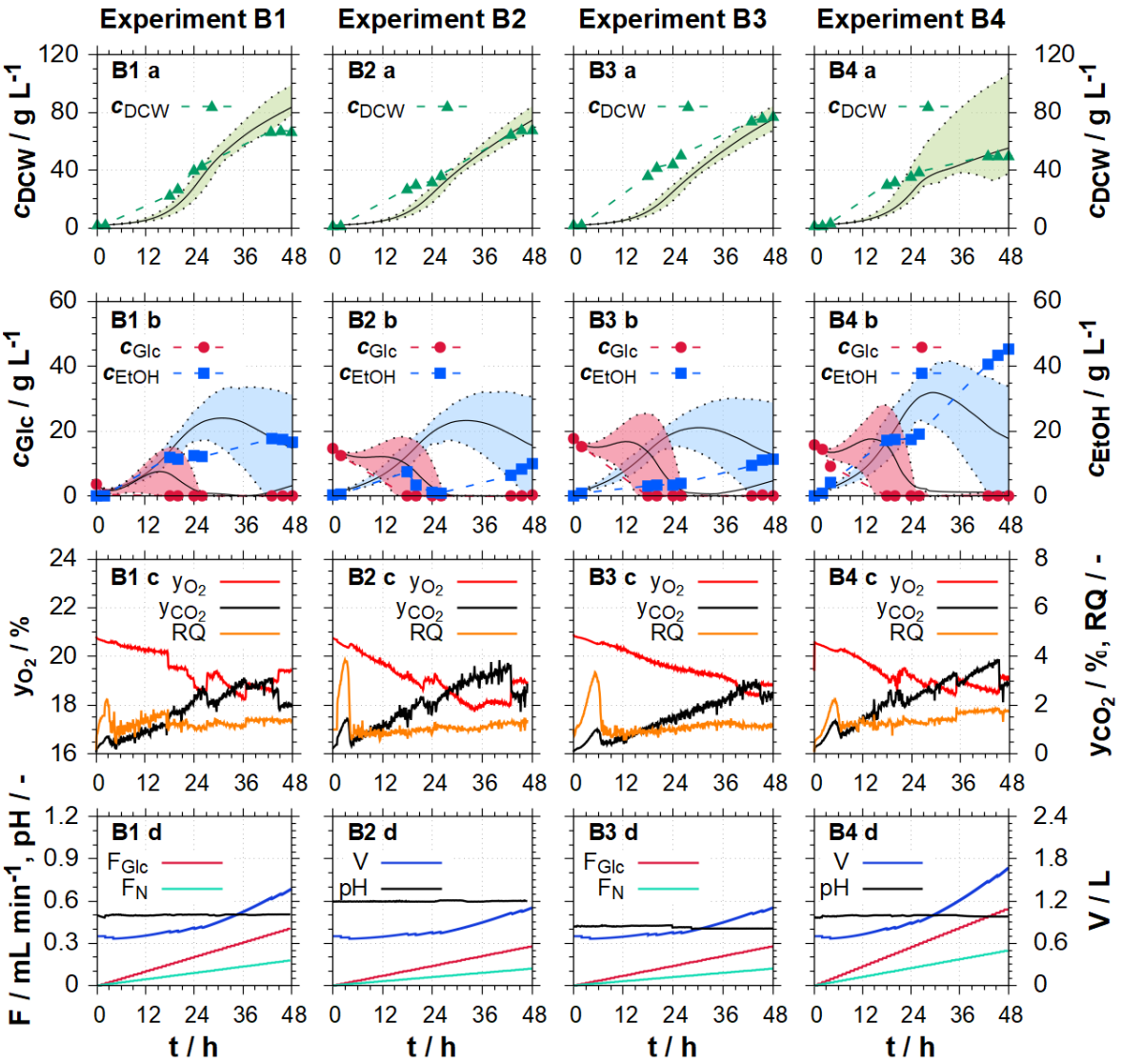


Figure 46: Experimental data for mDoE cultivations B1 – B4 compared to simulated data from the mDoE-Toolbox. Plots B1 a/b – B4 a/b show the data for c_{DCW}, c_{Glc} and c_{EtOH}. The continuous lines in plots B1 a/b – B4 a/b represent the arithmetic mean of 30 simulations with different parameter sets determined by Latin-Hypercube sampling. The dashed lines represent the 10% and 90% quantiles of these 30 simulations, the area between the dashed lines is the predicted uncertainty band. Plots B1 c – B4 c are the online data of the exhaust gas data composition (y_{O2} and y_{CO2}) and the calculated RQ. Plots B1 d – B4 d are displaying the pH, the calculated feed rates (F_{Glc} and F_N) as well as the calculated medium volume V.

B1 achieved the highest desirability score of 0.92 (Table 12) and highest simulated average biomass density while also having a narrow uncertainty band. B2 and B3 show slightly narrower uncertainty bands, but their simulated mean biomass density is slightly lower compared to B1. It is expected that simulations with narrow uncertainty bands are indicating

cultivations that are robust to changes in the model parameters and to the process control since the experimental data and process control of experiments A1 – A3 used for parameter identification were varied.

In contrast, B4 shows the widest uncertainty band but the highest potential biomass density in the prediction among all experiments (upper boundary of uncertainty band). The wide uncertainty band indicates a high variability in the simulations of B4. This is due to the risk of overfeeding, which potentially results in the highest ethanol concentrations.

The final dry cell weight density (at t = 48 h) is determined between 45 g L^{-1} (B4) and 80 g L^{-1} (B3). The initially provided glucose was fully depleted after 14 hours in all experiments. In B1 the glucose feed increased the glucose concentration in the medium above the limit for overflow metabolism which resulted in an ethanol concentration of almost 20 g L^{-1}. In cultivations B2 and B3 the lower glucose feeding rate caused less ethanol (c_{EtOH} = 12 g L^{-1}) to be produced.

In experiment B4, with the highest maximum glucose feeding rate (0.54 mL min^{-1}) and the highest nitrogen source feeding rate (0.25 mL min^{-1}), a final ethanol concentration of more than 45 g L^{-1} was reached. This high ethanol concentration most likely inhibited biomass growth resulting in the lower final biomass density of about 50 g L^{-1} compared to the other experiments.

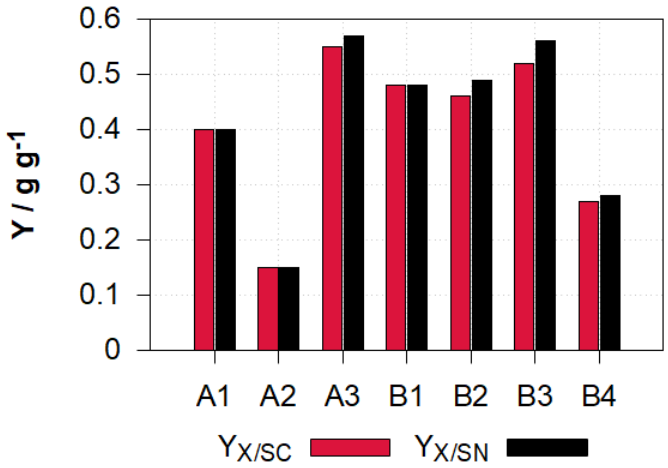
The **oxygen and carbon dioxide levels** in the exhaust gas changed nearly linearly, as theoretically expected with linear rising feeds when the glucose fed is immediately used up (see equation (4.16) for ethanol production). The steps in the exhaust gas online data were caused by adjustments of the aeration rate to ensure aerobic conditions.

The **RQ** reached its maximum in the first hours of the cultivation (at about t = 3 h), shortly before the initial glucose concentration was fully depleted. Afterwards, the RQ subsequently increased gradually, matching with increasing ethanol concentrations. Cultivations B2 and B3 exhibit low maximum ethanol concentrations of about 12 g L^{-1}, respectively. This corresponds to RQ values closer to and below 1 after the initial RQ peaks. In contrast, B1 and especially B4, showed higher RQ values above 1.5, indicating an ethanol production in both experiments, which is evident in the higher final ethanol concentrations (20 g L^{-1} in B1 and 45 g L^{-1} in B4 compared to B2 and B3).

The biggest difference between the simulated **uncertainty bands** and the determined dry cell weight density occurs in experiment B3. The experimental data exceed the simulations by over 20 g L^{-1} in experiment B3 at around 24 h. This was also the case in the other three cultivations but less pronounced. The final biomass densities are predicted well and are located within the uncertainty band, except for B1 which is about 10% less than the lower boundary of the uncertainty band.

The shape of the ethanol uncertainty bands differed qualitatively from the experimental data, exhibiting a peak concentration, whereas during the experiments, the ethanol concentration increased throughout the complete cultivation.

Experiment B3 with the highest final biomass density of 80 g L^{-1} showed simultaneously the highest **yield coefficient** $Y_{X/S}$ of 0.52 of the mDoE experiments B1 – B4 (Figure 47).



Exp.	$Y_{X/SC}$ $g_{DCW}\ g_{Glc}^{-1}$
A1	0.40
A2	0.15
A3	0.55
B1	0.48
B2	0.46
B3	0.52
B4	0.27

Figure 47: Yield coefficient for glucose $Y_{X/SC}$ for the experiments used for parameter identification (A1 – A3) and the mDoE experiments (B1 – B4).

The yield coefficient $Y_{X/SC}$ of experiments B1 and B2 were close with values of 0.46 and 0.48 respectively; only $Y_{X/SC}$ of B4 was considerably lower at 0.27. The yield coefficients of B4 and A2 are lower compared to the other experiments since most of the glucose is metabolised into ethanol instead of biomass resulting in lower final biomass densities. With the application of the modified mDoE the yield coefficient of B3 was increased by 25% and the final biomass by (more than) 30% compared to the experiment with similar initial conditions (A1).

7.3 Discussion and conclusion

After parameterisation with data from only three historical cultivations, the model's predictive capabilities were sufficient to improve the process performance in an ideally mixed bioreactor system. The biomass density was increased by 30% compared to a previous cultivation (A1) with similar initial conditions. Furthermore, this improvement was achieved with just four new experiments (B1 – B4), instead of the potential 27 cultivations in a full-factorial design or 15 in a central composite design (CCD), leading to a reduction of the required number of experiments by at least 75%.

A total of 120 parameter identifications were performed using the Monte-Carlo uncertainty quantification, efficiently utilising the limited data available from only three historical

cultivations. The desirability function was computed with the aim of maximising the biomass density while simultaneously minimising the variability. This leads to an approximation of the simulated robustness of the process to changes in the model parameters (obtained by varying experimental data and process control of experiment A1 – A3) and to the process control.

A novel approach was chosen for the mDoE, simulating a large number of experimental settings in the design space. A three-dimensional design space for three factors (F_{Glc}, F_N and pH) with 2000 points was created using the *k-means* clustering algorithm maximising the distance between points. These 2000 different settings were simulated with 20 simulations each (as described in chapter 5.3). The expected process performance was evaluated based on the mean result and variability. The result is a detailed representation of the desirability score depending on the three factors (F_{Glc}, F_N, pH value).

The substantial number of different simulated experimental conditions reduces the impact of the design space boundaries on the mDoE result (see chapter 7.2.1). In previous versions of the mDoE and other combinations of DoEs with model-assisted methods, classic design of experiments were utilised for the simulation of the design space. The small number of experimental settings (factor combinations) in classic designs enabled a faster calculation and a direct comparison with experimentally executable design spaces. However, the resolution of the resulting design spaces are lower, selected boundary conditions have a huge influence (see chapter 7.2.1) and an evaluation via response surfaces would be necessary. (Mandenius and Brundin 2008; Möller et al. 2015; Möller et al. 2019; Bayer et al. 2020; Gassenmeier et al. 2022).

Experiment B3 yielded the highest final biomass density. The settings of B3 were located in the centre of the optimum area (see Figure 44) with a slightly lower desirability score of 0.89 compared to experiment B1, which had a slightly higher desirability score (0.92) and about 50% higher volumetric feeding rates. This experiment B3 exhibits a high yield coefficient $Y_{X/S}$ of 0.52 which is close to the most common value found in literature of approximately 0.5 and only slightly lower than experiment A3 used for parameter identification (see Figure 47) (Sonnleitner and Käppeli 1986; Babel et al. 1983; van Dijken et al. 2000).

Overall, the application of the mathematical model within the novel approach of the model-based design of experiments was validated. The optimisation of a *S. cerevisiae* process was achieved, while simultaneously identifying the model parameters. Moreover, these improvements were made in the relevant region of expected desirable process performances.

With this result, the foundations are in place to use the mathematical model for processes under heterogeneous conditions. The next experiments in the physical scale-down model were performed with heterogeneous dissolved oxygen and substrate concentrations. Afterwards the model parameters were identified. Finally, the mathematical model was validated with experimental data and the process was optimised with the application of the mDoE (chapter 8).

8 Scale-down model characterisation

The developed physical scale-down model (SDM) is consisting of two connected stirred tank reactors. Two different heterogeneous conditions were identified for the *S. cerevisiae* cultivations in the SDM. The first one are heterogeneous dissolved oxygen concentrations in production reactors and the second one heterogeneous carbon substrate concentrations. The research objective of the experiments was the characterisation of the influence of these heterogeneities on *S. cerevisiae* cultivations in a controlled non-ideally mixed environment (chapters 8.1 and 8.2). Experimental data on the simultaneous influence of heterogeneous dissolved oxygen and substrate concentrations are not published. The new information and data gained were used for the identification of the parameters of the new mathematical model.

Five batch experiments were performed to quantify the influence of controlled heterogeneous dissolved oxygen conditions (chapter 8.1, Table 13). Batch experiments were carried out to isolate the effect of this heterogeneity without overlaying heterogeneous substrate concentrations. Three experiments were performed in the SDM, two cultivations where both reactors were either aerobic (C1) or anaerobic (C2) as well as one experiment where one reactor of the two-compartment SDM was aerobic and the other anaerobic (C3). These cultivations were compared to two aerobic cultivations in a single reactor system with similar initial conditions (C4 & C5). The experimental data were used to quantify the influence of heterogenous dissolved oxygen concentrations in the SDM.

Afterwards, five fed-batch experiments were conducted and analysed (chapter 8.2, Table 15). These fed-batch experiments introduced a second heterogeneity to the scale-down system (heterogenous substrate conditions). The first two cultivations (D1 & D2) were carried out to determine whether the location of the feeding inlet, either positioned in the aerobic or the anaerobic reactor, had an influence on the process performance. This was not performed and investigated so far for fed-batch cultivations of *S. cerevisiae*. Please note, that the feeding inlet is always located in the first reactor (R1) of the SDM independent from the location of the aeration. The fed-batch cultivations D3 and D4 were performed in a single bioreactor under fully aerobic (D3) and anaerobic conditions (D4). Experiment (D5) in the SDM consisting of two reactors and featuring a batch phase, followed by two feeds with rising volumetric feed rates.

Afterwards, the available data (experimental series A - D) of 16 experiments were utilised for the identification of parameters of the mathematical model. The mathematical model with identified model parameters was validated against three experiments: (1) the single reactor cultivation A3 (in chapter 8.3.2.1); (2) a new experiment with feeding rates

alternating between both reactors (E1, chapter 8.3.2.2); (3) a new experiment with conditions determined with mDoE as will be presented in chapter 8.3.2.3.

For the experiments in the SDM (C1, C3, C5, D1, D2 and D5) the initial volumes of the reactors were identical. This 1:1 ratio is similar to other working groups and roughly represents large-scale stirred tank reactor conditions (Lara and Leal et al. 2006; Oosterhuis 1984; Amanullah et al. 2004). Oosterhuis et al. (Oosterhuis et al. 1985; Oosterhuis 1984) used a 1.6 L reactor for the anaerobic compartment and a 0.6 L as the aerobic compartment. Lara et al. and Sandoval-Basurto et al. utilised a 2:1 ratio between the anaerobic and aerobic compartment (Lara and Leal et al. 2006; Sandoval-Basurto et al. 2005). The mixing times were set to 120 s for the batch experiments and to 4 min for the fed-batch experiments to increase the influence of the heterogenous dissolved oxygen (and substrate) concentrations (see chapter 4.3.3.1). These mixing times are in the range of large production scale reactors as published by Vrábel et. al (Vrábel et al. 2000). Sweere et. al 1988 were able to show that increased mixing times lead to greater effects of heterogeneous dissolved oxygen concentrations in non-ideally mixed reactor systems (Sweere and Matla et al. 1988).

In all experiments, the medium temperature was controlled at 28 °C through an external water bath (MDX bioreactors and SDM) and internal temperature control in the Biostat B. The pH was not controlled.

8.1 SDM - batch experiments (C1 – C5)

Three scale-down model experiments were carried out in batch operation mode, along with two additional experiments in a single bioreactor (MDX bioreactor). Batch experiments were chosen for cultivations C1 – C5 due to their uncomplicated process control and ability to provide valuable data on uptake rates and yield coefficients. The initial volume of each reactor was 1 L. The mixing time was set to approximately 120 s by adjusting the volumetric flow rate of the circulation pump (method described in chapter 4.3.3.1). The conditions in the single bioreactor were assumed to be ideally mixed at stirrer speeds at or above 800 rpm. The initial nitrogen source was consisting of 15 g L^{-1} yeast extract and 15 g L^{-1} peptone in all cultivations. The initial glucose concentration was set to 100 g L^{-1} for the first two experiments. Both cultivations were inoculated to an initial biomass density of 6.0 g L^{-1}. Table 13 lists the initial volumes, mixing times and the initial concentrations of biomass, glucose as well as the nitrogen source (peptone and yeast extract).

Table 13: Experimental conditions of five batch cultivations, three in the scale-down model (C1, C3, C5) and two in a single bioreactor (assumed to be ideally mixed).

Experiment		V_{R1}	V_{R2}	t_{M90}	c_{DCW}	c_{Glc}	c_{YE}	$c_{Pep.}$
#	Combi.	L	L	s	g L^{-1}	g L^{-1}	g L^{-1}	g L^{-1}
C1	Ae-Ae	1.0	1.0	~120	6.0	100	15.0	15.0
C2	Ae	1.0	-		6.0	100	15.0	15.0
C3	An-Ae	1.0	1.0	~120	10.0	50.0	15.0	15.0
C4	Ae	1.0	-		10.0	50.0	15.0	15.0
C5	An-An	1.0	1.0	~120	10.0	100	15.0	15.0

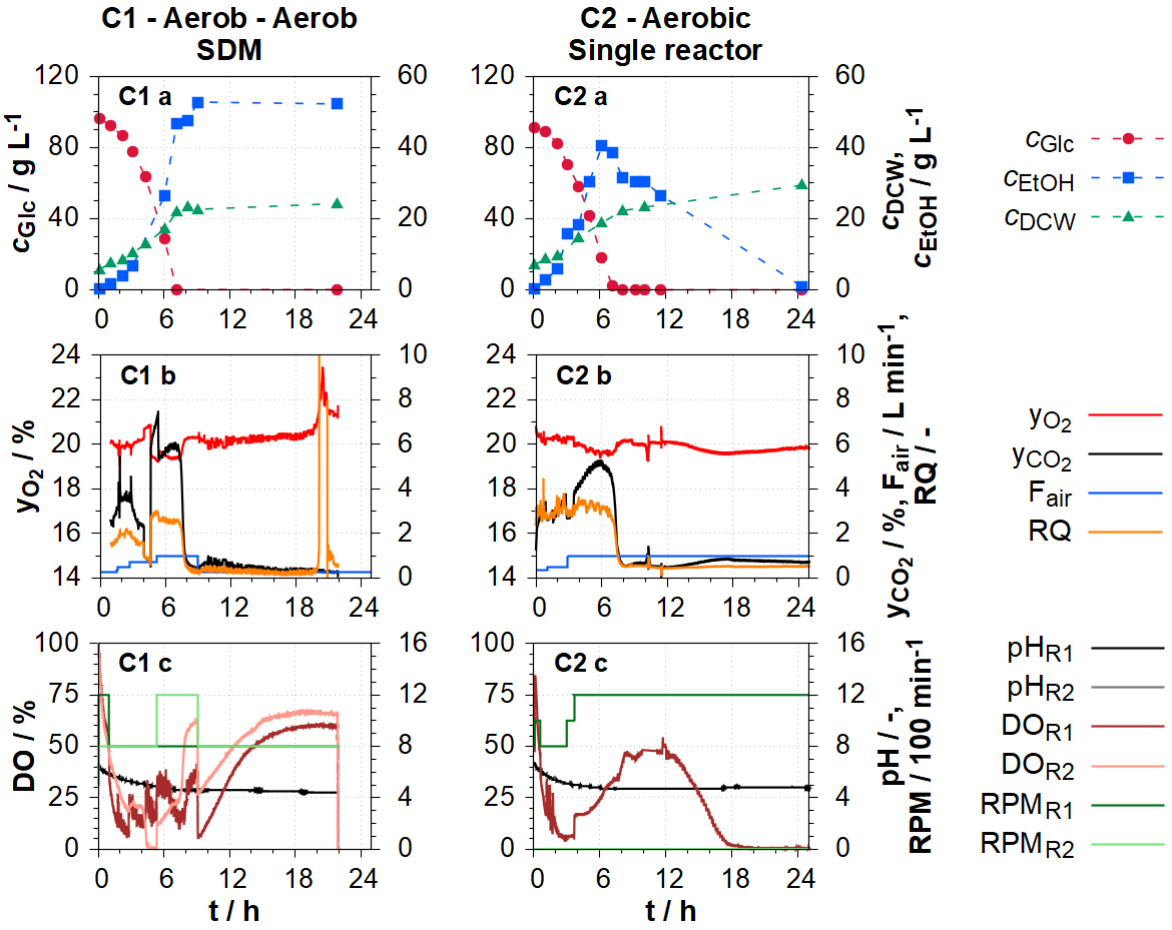


Figure 48: Comparison between a batch cultivation in the scale-down system (R1: reactor 1, R2: reactor 2) under aerobic conditions in both reactors (1st column) with an aerobic experiment in a single bioreactor (2nd column). 1st row: Offline data for glucose and ethanol concentration as well as dry cell weight density. 2nd row: Online data for the exhaust gas composition (y_{O2} and y_{CO2}), RQ and the aeration rates. 3rd row: Online data for the dissolved oxygen concentration, pH value and stirrer speed of the reactors.

In cultivation C1 (C1 Ae-Ae), both reactors were aerated, resulting in aerobic conditions in both reactors. This experiment was compared to experiment C2 in a single bioreactor (1 L MDX bioreactor) under aerobic conditions (DO > 10%) (C2 Ae) to check if the result of the SDM could be reproduced in a single bioreactor (Figure 48).

In C1, glucose was depleted in the first seven hours, leading to an increase of the biomass density from initially 6 g L^{-1} to 22 g L^{-1}. Biomass production stopped almost completely after 7 h and a final biomass density of 23 g L^{-1} was determined at the end of the cultivation C1. In C2, glucose was completely depleted after 7 h as well and a biomass density of 22 g L^{-1} was reached at 12 h which is similar to experiment C1.

Ethanol was produced up to a concentration of approximately 50 g L^{-1} in C1 and 40 g L^{-1} in C2 in the first seven hours. Afterwards, the ethanol concentration was constant at 50 g L^{-1} in C1. Ethanol was therefore not consumed in the SDM under fully aerobic conditions. This difference between the maximum ethanol concentrations in C1 and C2 might be caused by different initial glucose concentrations (96 g L^{-1} in C1 and 90 g L^{-1} in C2) and the ethanol consumption in C2.

However, in C2, ethanol was consumed starting at 7 h for the following 11 h, which caused an increase in biomass density from 22 g L^{-1} to almost 30 g L^{-1}. At 17 h ethanol was fully consumed in C2, which is indicated by the exhaust gas composition. At 17 h, the mole fraction of oxygen in the exhaust gas (y_{O_2}) was at the lowest while the of carbon dioxide (y_{CO_2}) was at its highest point during the ethanol consumption phase (t > 6 h). Afterwards, y_{O_2} was rising and y_{CO_2} falling. The pH was also slightly increasing at that time-point again after remaining at a constant value of 4.7 for 10 h.

The DO was above 10% except for a brief time period from 4 h to 5.5 h in R2 (cultivation C1), indicating aerobic conditions in the SDM-reactors. After glucose was fully consumed, the DO in both reactors was increasing and reached final values of above 60%. The pH was decreasing from initially 6.8 to 4.4.

The third experiment (C3 An-Ae) in the scale-down model was conducted in a way where one reactor was aerated and the other not, resulting in different DO concentrations in the two reactors. For this experiment the initial glucose concentration was set to 50 g L^{-1} to reduce the maximum ethanol concentration and the risk of ethanol inhibition. This experiment was compared to an aerobic experiment (C4 Ae) in a single bioreactor with the same initial conditions, to assess the influence of heterogeneous DO concentration on the cultivation. Both experiments (C3 & C4) were inoculated to a biomass density of 10.0 g L^{-1} (Figure 49).

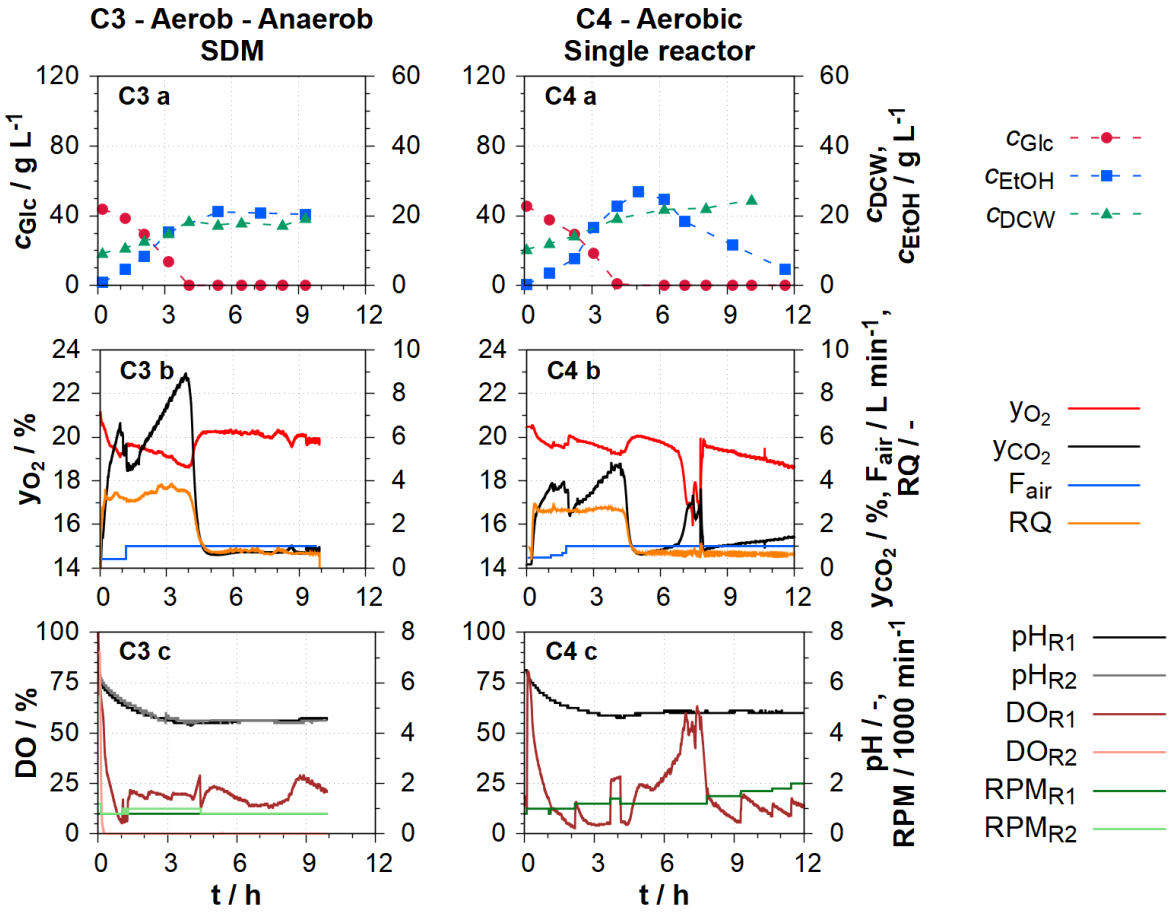


Figure 49: Comparison between batch cultivation C3 in the scale-down system (R1: reactor 1, R2: reactor 2) under heterogenous DO (R1 aerobic, R2 anaerobic) (1st column) with process C4 in a single bioreactor under aerobic conditions (2nd column). 1st row: Offline data for glucose and ethanol concentration as well as dry cell weight density. 2nd row: Online data for the exhaust gas composition (y_{O2} and y_{CO2}), RQ and the aeration rates. 3rd row: Online data for the dissolved oxygen concentration, pH value and stirrer speed of the reactors.

The initial glucose (c_{Glc} = 50 g L^{-1}) was completely depleted within the first 4 hours in both cultivations. In C3 biomass density was rising from 10 g L^{-1} to 18 g L^{-1} during the first 4 hours and remained at that density until the end of the experiment. The ethanol concentration reached about 21 g L^{-1} and remained almost constant with a final concentration of 20 g L^{-1}. The RQ decreased to values below 1 after glucose was fully consumed, the pH was dropping to values of about 4.5. The DO in the first reactor of the SDM (R1) is above 10% throughout the cultivation whereas in the second reactor (R2) the DO is constantly at 0%.

In contrast to C3, during experiment C4 biomass grew throughout the cultivation until reaching a final biomass density of 25 g L^{-1}. Ethanol concentration reached a peak of

25 g L^{-1} at 4 h and was fully consumed until approximately 13 hours cultivation time. The pH dropped to 4.7 until the ethanol peak at 5 h and increased slightly until ethanol was fully consumed. The transition from glucose to ethanol uptake is reflected in the RQ. RQ declined from values above 2 to values below 1 between 3 and 4 h cultivation time when glucose was fully depleted indicating ethanol consumption in both experiments.

The last experiment (C5 An-An) was performed in the SDM with both reactors under anaerobic conditions. This experiment was performed with an initial glucose concentration of 100 g L^{-1} and was inoculated to an initial biomass density of 10 g L^{-1} (Figure 50).

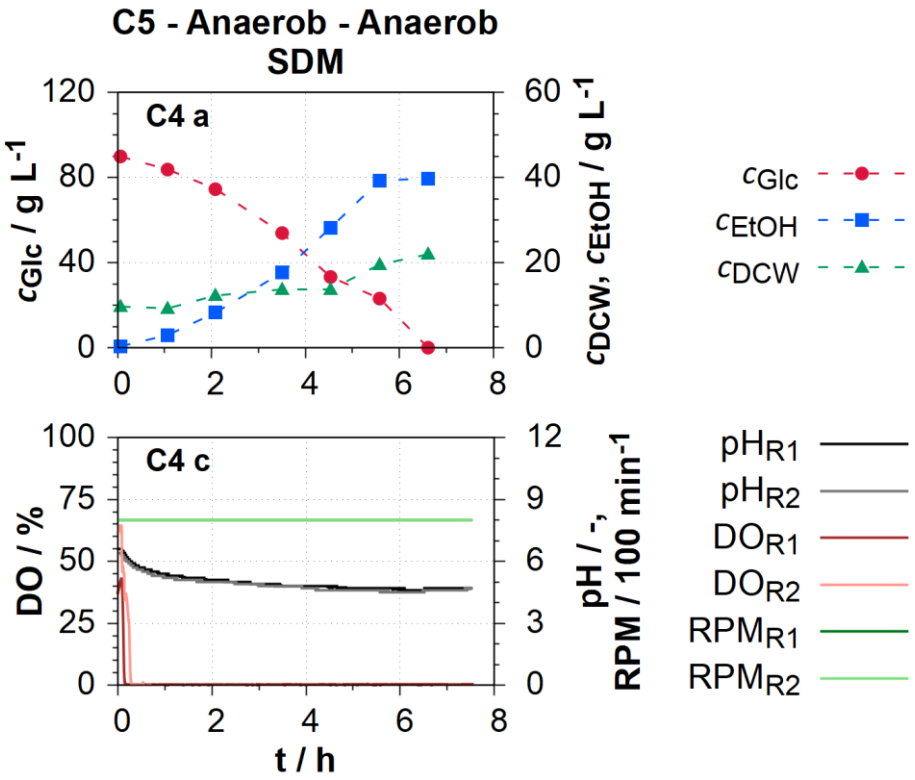


Figure 50: Batch cultivation C5 in the scale-down system (R1: reactor 1, R2: reactor 2) under anaerobic conditions. 1st row: Offline data for glucose and ethanol concentration as well as dry cell weight density. 2nd row: Online data for the dissolved oxygen concentration, pH value and stirrer speed of both reactors.

The initial glucose was depleted after about 6.5 hours. The biomass density was increasing from initially 10 g L^{-1} to a final biomass density of 22 g L^{-1}. The ethanol concentration reached approximately 40 g L^{-1} at 6.5 h. Ethanol was not consumed since the both reactors of the physical scale-down model were under anaerobic conditions. The pH started at about 6.7 and decreased to 4.7 over the course of the cultivation. The DO decreased in both reactors to 0% in the first 15 minutes of the cultivation C5. The composition of the exhaust gas was not determined since both reactors were unaerated.

8.1.1 Comparison of the batch cultivations C1 – C5

Table 14 displays the yield coefficients for the biomass formed per glucose consumed ($Y_{X/SC}$), as well as averaged glucose uptake rates (r_{Glc}) and ethanol uptake rates (r_{EtOH}) for the batch cultivations C1 – C5 (see chapter 4.3.4).

Table 14: Yield coefficients, glucose uptake, growth rate and ethanol production rate for the batch experiments in the physical two-compartment scale-down system and single bioreactors.[1]

Experiment		$Y_{X/SC}$	r_{Glc}	r_{EtOH}
#	Combi.	g_{DCW}/g_{Glc}	$g\ {g_{DCW}}^{-1}\ s^{-1}$	$g\ {g_{DCW}}^{-1}\ s^{-1}$
C1	Ae-Ae	0.19 ± 0.02	3.1 ± 0.4·10^{-4}	-
C2	Ae	0.22 ± 0.02	2.7 ± 0.3·10^{-4}	4.3 ± 0.4·10^{-5}
C3	An-Ae	0.21 ± 0.02	2.4 ± 0.3·10^{-4}	1.5 ± 0.1·10^{-5}
C4	Ae	0.28 ± 0.03	2.2 ± 0.2·10^{-4}	3.9 ± 0.2·10^{-5}
C5	An-An	0.13 ± 0.02	2.7 ± 0.3·10^{-4}	-

Cultivation C3 under mixed conditions shows the highest yield coefficient $Y_{X/SC}$ of 0.21 of the three cultivations performed in the SDM (C1, C3, C5), which is slightly higher than $Y_{X/SC}$ of the fully aerobic experiment C1 ($Y_{X/SC}$ = 0.19) and 50% higher compared to C2 ($Y_{X/SC}$ = 0.13). However, experiment C4 in the single bioreactor exhibits the highest yield coefficient ($Y_{X/SC}$ = 0.28). Without the biomass growth from ethanol uptake this would be reduced to 0.23 (± 0.02) and be in the same range as the other experiments (except C5). The growth based on ethanol uptake in C1, C3 and C5 is considerably smaller (reducing $Y_{X/SC}$ by less than 0.02).

For the cultivations with similar initial glucose concentrations (C1, C2, C5 as well as C3, C4), the glucose uptake rates are similar ranging from 2.2·10^{-4} g g_{DCW} s^{-1} (C4) to 3.1·10^{-4} g g_{DCW} s^{-1} (C1). The glucose uptake rates are slightly higher in the fully aerobic experiments C1 and C2 where a lower initial biomass concentration was chosen and in the fully anaerobic experiment in the SDM (Table 14).

Figure 51 presents the offline data determinations for dry cell weight density, glucose concentration and ethanol concentration for the batch cultivations to support a direct comparison of concentration profiles.

[1] Deviations were estimated for the case of a 5% deviation for the initial glucose or highest ethanol concentration and 5% deviation for the average biomass density. The highest calculated deviation for those cases is shown in the table.

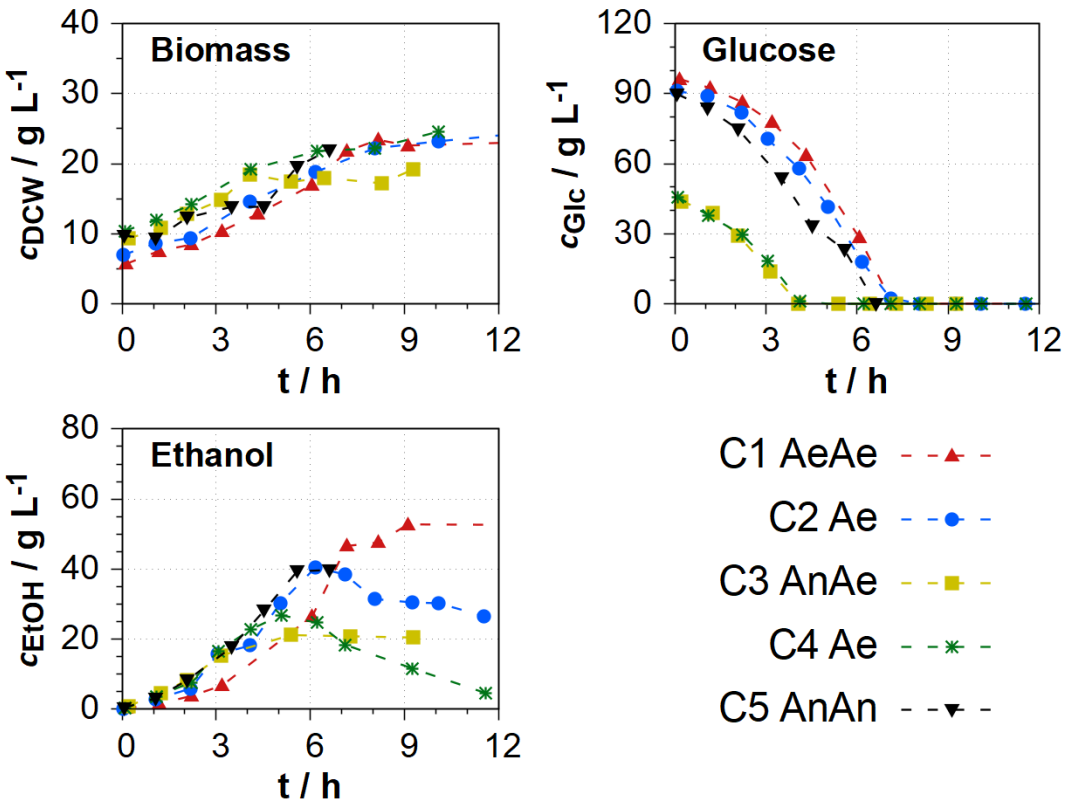


Figure 51: Comparison of dry cell weight density, glucose concentration and ethanol concentration for experiments C1 – C5

All five experiments support the concept of overflow metabolism (Crabtree-effect and anaerobic metabolism) which resulted in ethanol production. The highest ethanol concentrations correspond to about half the initial glucose concentrations which is close to the theoretically possible ethanol produced from glucose for batch processes (see chapter 4.3.4). The biggest difference between the single stirred tank reactor cultivations and the heterogeneous SDM was that ethanol is consumed in the single-bioreactor cultivations at a higher rate compared to the experiments in the SDM (C1 and C3). In STR experiment C2 the ethanol uptake rate was $4.3 \cdot 10^{-5}$ s^{-1} and in STR experiment C4 $3.9 \cdot 10^{-5}$ s^{-1}. The rate is calculated for the interval starting at the peak ethanol concentration (for C2 at 6 h, for C4 at 4 h) until the end of the cultivation (12 h). In contrast, no changes in the ethanol concentration were determined in the fully aerobic (C1) and anaerobic (C5) cultivations in the SDM. In the SDM experiment with heterogenous DO concentrations (C3), an ethanol uptake rate of $1.5 \cdot 10^{-5}$ s^{-1} was determined which is over 60% lower compared to the rates of the aerobic STR experiments C2 and C4.

The reduction of the ethanol uptake rate in C3 is most likely mainly caused by one reactor being under anaerobic conditions. The volume of the anaerobic reactor was set to 1 L; sampling reduced the volume in the aerobic reactor by 10% (0.1 L) during the cultivation.

Since ethanol consumption only takes place under aerobic conditions and approx. 50 - 55% of the volume of the system is anaerobic, a reduced ethanol uptake rate by the same percentage (50 – 55%) is expected. This would result in an ethanol uptake rate of the aerobic reactor of approx. $3.0 \cdot 10^{-5}$ s^{-1}. This rate (in experiment C3) is about 25% to 30% lower than the uptake rates in the aerobic single reactor cultivation (C2). This further reduction of the ethanol uptake rate in the SDM might be attributed to the cell alternating between the aerobic and anaerobic compartment. Furthermore, due to the low ethanol uptake rate and the heterogeneous conditions, biomass growth nearly stopped in the SDM cultivation after glucose depletion.

The process performance was clearly affected by the heterogeneity created in the SDM. The experimental results of this study with batch experiments confirm the results of Sweere et. al. (1988) that the production of *S. cerevisiae* is reduced under heterogeneous dissolved oxygen concentrations (Sweere and Matla et al. 1988). The experiments also confirm the observation by George et. al. (1998) that the growth of *S. cerevisiae* is reduced in non-ideally mixed production reactors (George et al. 1998). This behaviour was in general already predicted by the mathematical model during the simulation study (chapter 6). The main difference between simulation and experiment is, that the ethanol uptake rate was much lower in experiments with heterogeneous conditions compared to fully aerobic conditions. Under fully anaerobic conditions, no ethanol consumption was determined (experiment C5) which was expected since oxygen is needed for the oxidisation of ethanol (Battley 1960).

8.2 SDM - fed-batch experiments (D1 – D5)

Five fed-batch cultivations were performed to characterise the scale-down system with heterogeneous DO concentrations and additionally heterogenous conditions regarding the substrate concentration induced by feeding into only one of the two reactors.

Table 15 lists the initial experimental conditions of the fed-batch experiments.

Table 15: Experimental conditions of five fed-batch cultivations, three in the scale-down model (D1, D2, D5) and two in a single stirred bioreactor (MDX bioreactor) (D3 and D4). The compartment with feed is named first in the 2nd column of the table. The conditions in the single bioreactor were assumed to be ideally mixed (stirrer speed at 600 rpm in D3 and at 300 rpm in D4).

Experiment		V_{R1}	V_{R2}	t_{M90}	c_{DCW}	c_{Glc}	c_{YE}	$c_{Pep.}$	$c_{Glc,Feed}$
#	Combi.	L	L	s	g L^{-1}	g L^{-1}	g L^{-1}	g L^{-1}	g L^{-1}
D1	Ae-An	0.5	0.5	260	10.0	-	25.0	25.0	500
D2	An-Ae	0.5	0.5	260	10.0	-	25.0	25.0	500
D3	Ae	1.0	-	-	10.0	-	25.0	25.0	500
D4	An	1.0	-	-	10.0	-	25.0	25.0	500
D5	An-Ae	0.6	0.6	250	6.0	25.0	25.0	25.0	500

The glucose concentration of the feed was 500 g L^{-1} for all cultivations. The first reactor (labelled R1) in the SDM experiments is always the reactor where the feeding inlet is positioned. One reactor of the experiments in the SDM (D1, D2, D5) was aerated causing aerobic conditions, while the other one was not aerated (anaerobic conditions). The feeding inlet is located in the aerobic reactor in cultivation D1 (D1 **Ae**An), in D2 and D5 in the anaerobic reactor of the SDM (D2 **An**Ae, D5 **An**Ae).

The purpose of D1 and D2 was to determine the influence of the feeding inlet positioning on cultivations in the SDM. It is anticipated that glucose concentrations will be higher in the reactor with feed compared to the reactor of the SDM without the feed causing a heterogeneous glucose concentration in the system.

Cultivation D5 utilised two feeding periods, with different volumetric feeding rate, which should lead to two separate ethanol consumption phases with lower ethanol concentrations compared to all other experiments performed in the SDM. The first feeding period is performed with lower volumetric flow rates to obtain data for simultaneous feeding and ethanol uptake. The initial medium volume in each of the reactors was set to 0.5 L, totalling 1 L.

Cultivations (D3 and D4) were performed under either fully aerobic (D3) or fully anaerobic (D4) conditions utilising the identical feeding strategy, initial conditions, and initial volumes as experiments D1 and D2 (see Table 15).

The circulation flow rate was set to 150 mL min^{-1} resulting in experimentally determined mixing times of approximately 260 s for processes D1 and D2. The mixing time was doubled compared to the batch-mode experiments of the previous chapter 8.1 with the aim to amplify the influence of the heterogenous dissolved oxygen concentrations. This higher mixing time

of approx. 4 minutes is still in the range of large scale reactors as published by Vrábel et. al (Vrábel et al. 2000).

The STR cultivations D3 and D4 had initial volumes of 1.0 L and were performed in the 2 L Biostat B bioreactor. The reactor was assumed to be close to ideally mixed, as mixing times below 5 s were estimated by visual observations. D3 and D4 were performed to quantify the influence of the heterogeneous conditions regarding the substrate and dissolved oxygen concentrations against experiments in ideally mixed homogenous systems.

For cultivation D5 in the SDM the volume in each reactor was increased to 0.6 L and the circulation flow to 180 mL min^{-1}. The reactor volumes and the circulation flow rate are both 20% higher compared to D1 and D2 which resulted in a mixing time of 250 s. This mixing time is close to the mixing times of D1 and D2 and only about 4% (10 s) lower.

The initial biomass density (c_{DCW}) was set to 10 g L^{-1} to decrease the process duration in D1 – D4 to 24 h with sufficient information content. This biomass density of 10 g L^{-1} can approximately be expected after 12 h in a process with an initial biomass density of about 2 g L^{-1}. In experiments D1 – D4 glucose was exclusively provided by the feed medium (no glucose in the initial medium); nitrogen source was provided with the initial medium (peptone and yeast extract concentration set to 25 g L^{-1}). Feeding took place during the initial six hours of processes D1 – D4 at a feeding rate of 0.6 mL min^{-1} and a glucose concentration in the feed of 500 g L^{-1}.

Experiment D5 was performed in the SDM with a different feeding strategy with two separate feeding periods, an initial batch phase and a longer process duration of 48 h. The volumetric feeding rate of the first feed was set to 0.05 mL min^{-1} to acquire data for a feed that causes a lower ethanol production. The feeding rate of the second feed was set higher at 0.17 mL min^{-1} to acquire additional data for ethanol production. In addition, the behaviour of the system for a feeding strategy with an interruption and an ethanol consumption phase with lower ethanol concentrations compared to the other processes shall be investigated.

8.2.1 Influence of the feeding inlet position on simulations

Three preliminary simulations were carried out, with the feed positioned in the aerobic, the anaerobic, and lastly with a combination of both with half the feed in each reactor. The utilised parameter set was identified utilising data from the experiments presented in chapter 7.1.2. The initial conditions, the glucose concentration of the feed, and mixing times are identical to the settings of processes D1 and D2 (chapter 8.2.2). The simulations utilised six connected STR-models, two of which are operated as the reactors in the SDM, while the other four simulate the piping between the reactors. The circulation flow between the

reactors is set to 125 mL min^{-1}, resulting in a mixing time $t_{M,90}$ of 260 s. Figure 52 shows the simulation results different for positions of the feeding inlet for c_{Glc}, c_{EtOH} and c_{DCW} for cultivations in the SDM. The right side of Figure 52 displays a zoomed-in section of the process between 5h and 7h. The simulated relative concentration difference between the two reactors for each of the three simulations is small (less than 0.1%). This difference is not observable in Figure 52.

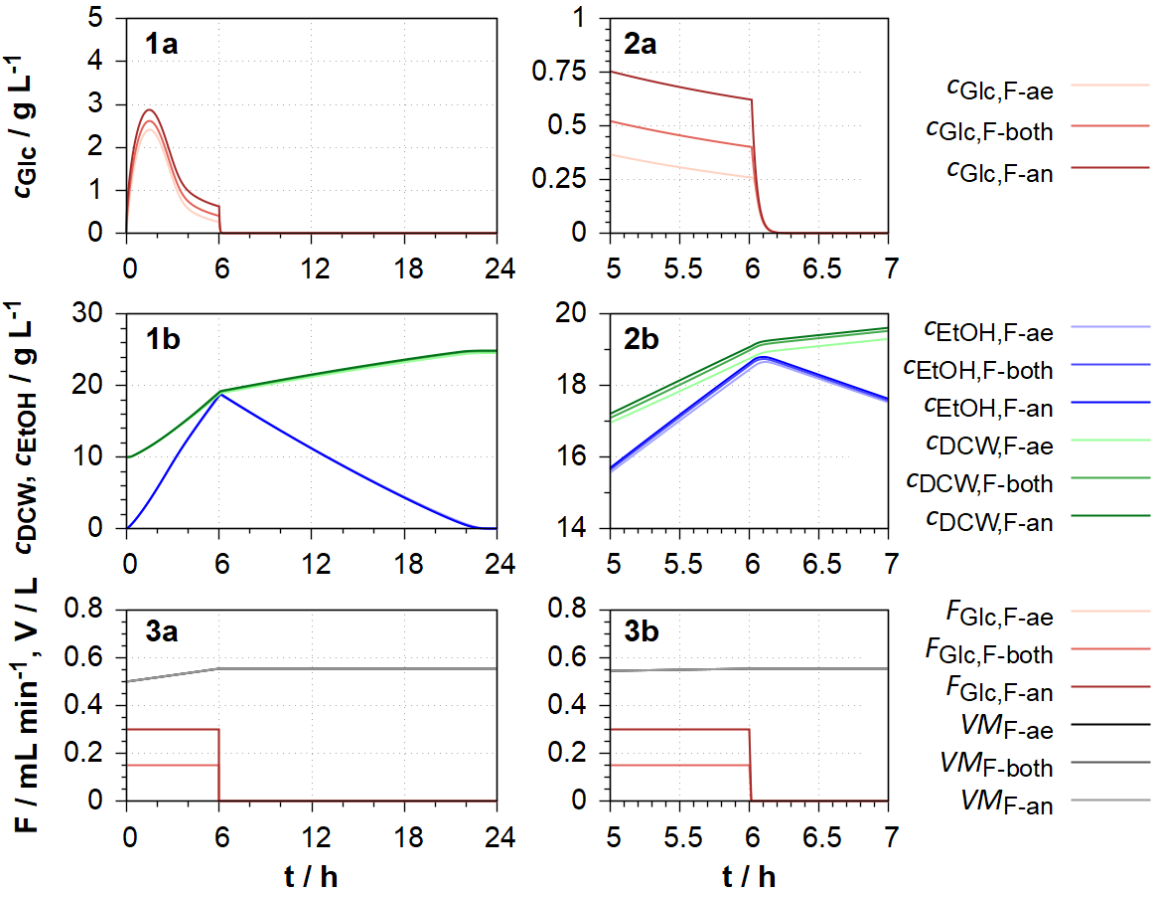


Figure 52: Influence of the positioning of the feeding inlet on simulations of two-reactor scale-down systems. Simulation with feeding in the aerobic reactor, both reactors and anaerobic reactor. Mixing time is set at 260 s for all simulations. Left column shows the entire simulation of the process; right column a zoomed in time period ranging from 5 h to 7 h. 1st row: Glucose concentration, 2nd row: Ethanol and biomass concentration, 3rd row: Feed volume flow and medium volume.

The glucose uptake rate is approximately 20% higher when the feeding inlet is located in the aerobic reactor compared to the anaerobic reactor. The ethanol concentration and the biomass density show slight decreases (approximately 2%), when glucose is fed in the aerobic reactor. In the next step, the influence of the positioning of the feeding inlet was

experimentally examined (D1 and D2), in comparison to fully aerobic (D3) and anaerobic (D4) fed-batch processes in a single bioreactor.

8.2.2 Fed-batch experiments (D1 – D4)

Two experiments corresponding to the simulations (Figure 52) were carried out. In experiment D1, glucose was fed into the aerobic reactor, in D2 into the anaerobic reactor. The volume of the anaerobic reactor was kept constant at 0.5 L, the volume in the aerobic reactor was variable, depending on the feeding rate and the sample volume. For both experiments, the circulation flow was set to 200 mL min^{-1} resulting in an experimentally determined $t_{M,90}$ of approx. 260 s (method described in chapter 4.3.3.1). The feeding (0.6 mL min^{-1}) was turned on for the first six hours with a glucose concentration of 500 g L^{-1} matching the flow rate in the simulations. The stirrer speed was set to 800 min^{-1} throughout the experiment.

The glucose and ethanol concentration, biomass density, the exhaust gas composition, and the pH were measured, and the aeration rate, feeding rate as well as the calculated reactor volumes have been recorded (Figure 53).

The biomass densities, metabolite concentrations and exhaust data of cultivations D1 and D2 are remarkably similar. In both cultivations c_{DCW} is increasing up to about 18 g L^{-1} (D1: 18.4 g L^{-1}; D2: 17.6 g L^{-1}) in the first eight hours of the experiment. In D1 the biomass density was slightly lower (16.2 g L^{-1}) at the end of the process, whereas in D2 c_{DCW} was increasing to 18.4 g L^{-1}. During D2, the glucose concentration reached a higher peak concentration of 9 g L^{-1}, which is one-third higher than in D1 where glucose was fed into the aerobic reactor ($c_{Glc,peak}$ = 6 g L^{-1}). Ethanol was produced up to a maximum of 22 g L^{-1} in both processes and was consumed afterwards. After 18 h of ethanol reduction (after feeding stopped), c_{EtOH} decreased to 15 g L^{-1}. The experimental ethanol uptake rate is about 70% slower compared to the uptake rate in the simulation (chapter 8.2.1).

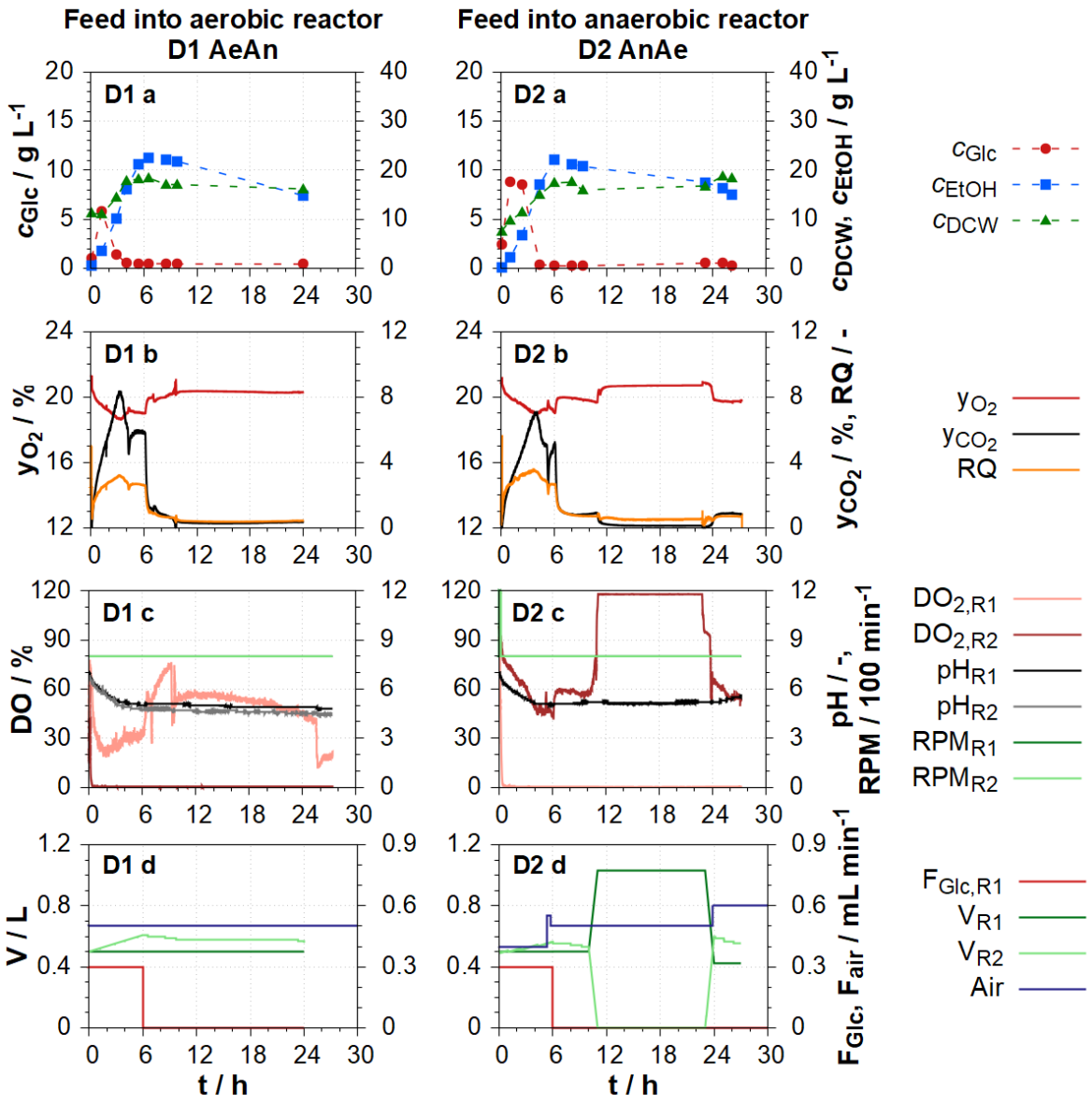


Figure 53: Fed-batch cultivations performed in the SDM with the feeding inlet positioned in the aerobic reactor (D1) and anaerobic reactor (D2). 1st row: Offline data for c_{DCW}, c_{Glc} and c_{EtOH}. 2nd row: Online data for the exhaust gas composition (y_{O2} and y_{CO2}) and the RQ. 3rd row: Online data for the dissolved oxygen concentration, pH value and stirrer speed of both reactors. 4th row: Feeding and aeration rate as well as the calculated volumes in both reactors.

The molar fraction of CO_2 in the exhaust gas increased until it reached a maximum (D1: 8.4%; D2: 7.0%) and subsequentially decreased after glucose was consumed. The RQ was above 1 for the initial 6 hours during feeding, indicating ethanol production in both reactors. After feeding stopped, RQ dropped below 1, indicating ethanol consumption. The exhaust gas composition remained stationary (y_{O_2} = 20.3%, y_{CO_2} = 0.3%) in D1 until the end of the process.

The DO in the aerobic reactor of cultivation D1 stayed above 20%. In the anaerobic reactors of both cultivations the DO never increased above 1% after the initial process phase. In D2 the DO in the aerobic reactor (R2) was staying above 20% until 10 h, afterwards the DO in R1 increased to 120%[2]. The DO in the reactor with the feeding inlet stayed close to 0% for the entire process D2 after the initial phase. The pH started at 7 and decreased to values at about 5.

Cultivations D3 and D4 were conducted in a single bioreactor in parallel to D1 and D2, with identical initial conditions, as well as feeding rates and concentrations. Cultivation D3 was performed aerobically while D4 was conducted anaerobically in an unaerated bioreactor. The volumes in D3 and D4 matched the total initial volume of the SDM (V_M = 1 L). Figure 54 presents the resulting cultivations D3 and D4.

[2] In experiment D2 the membrane filter for the exhaust gas was blocked from 10 h until 24 h. This caused the aerated reactor to empty through an increase in pressure. This is also evident in the exhaust gas data, as the measured exhaust gas composition matched the composition of the aeration flow (ambient air). The DO simultaneously reached approx. 120%, because the DO probe was in contact with air rather than the reactor medium, which was used for calibration of the probe. Furthermore, during this period, the circulation pump delivered a small unknown quantity of air into the anaerobic reactor, which was visually confirmed but insufficient to increase the DO of the anaerobic reactor. The blocked filter was replaced at t = 24 h. The experiment was extended to 26 h to ensure that enough data for the ethanol consumption phase were collected. The emptying of R2 was considered in simulations of experiment D2. The data from the experiment was not discarded, as this deviation from the planned process control can be considered in the mathematical model. Using mathematical models, the data of such "unsuccessful" experiments can be utilised and the need of experimental resources can be reduced.

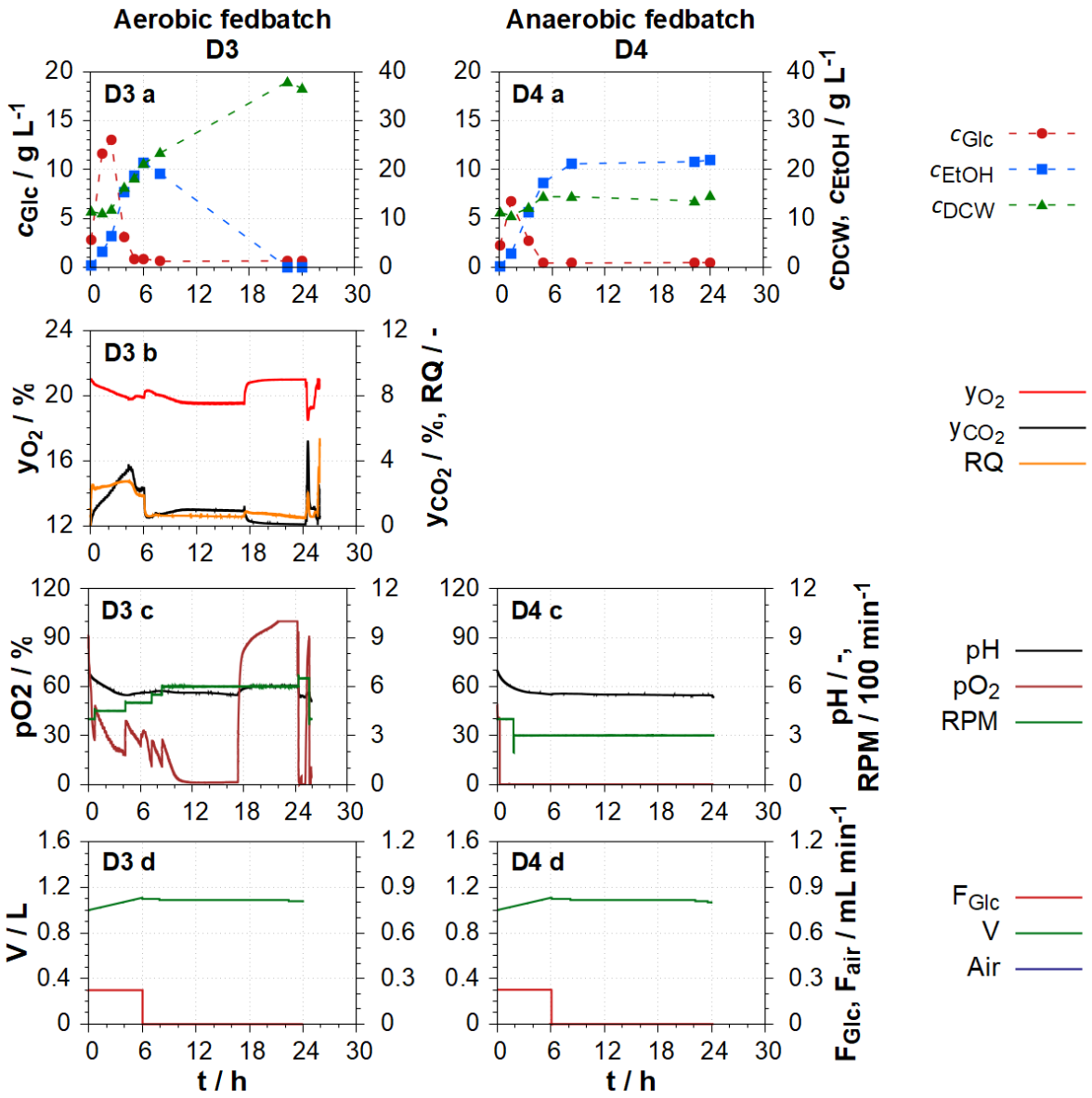


Figure 54: Fed-batch cultivations performed in the single bioreactor (aerobic conditions D3 and anaerobic D4). 1st row: Offline data c_{DCW}, c_{Glc} and c_{EtOH}. 2nd row: Online data for the exhaust gas composition (y_{O2} and y_{CO2}) and the RQ. 3rd row: Online data for the dissolved oxygen concentration, pH value and stirrer speed. 4th row: Feeding and aeration rate, stirrer speed as well as the calculated volumes.

The peak glucose concentration in the aerobic experiment D3 was the highest among the experiments, reaching 13 g L^{-1}. In D4 a peak glucose concentration of 6.7 g L^{-1} was reached. Although c_{EtOH} increased to 21 g L^{-1} in both experiments, it was only consumed in D3 under aerobic conditions and at a higher rate compared to experiments D1 and D2 with a heterogeneous DO concentration (chapter 0). After 17 h ethanol was completely depleted, indicated by the online data of DO, pH and in the exhaust gas composition, which all exhibit

abrupt changes at this time point. In D4 ethanol concentrations did not decrease, as was expected due to the anaerobic conditions throughout this cultivation.

At the end of D3, c_{DCW} reached 36 g L^{-1} (t = 24 h). In the anaerobic experiment D4, the biomass density remained constant after reaching approximately 14 g L^{-1}, which is similar to experiment D1 were the biomass density remained constant at about 16 g L^{-1}.

The fraction of CO_2 in the exhaust gas reached 4% and oxygen decreased below 20%. The RQ was above 1 during ethanol production and below 1 during ethanol consumption. During D3, the pH decreased to a value close to 5 but increased after ethanol depletion.

Comparison of the fed-batch experiments D1 – D4

Figure 55 illustrates a comparison of the offline data between the four performed fed-batch experiments D1 – D4 to quantify the influence of non-ideally mixed conditions on the cultivations, as well as the influence of the feeding inlet positioning. The feeding inlet was positioned in the aerobic reactor in experiment D1 and in the anaerobic reactor in experiment D2.

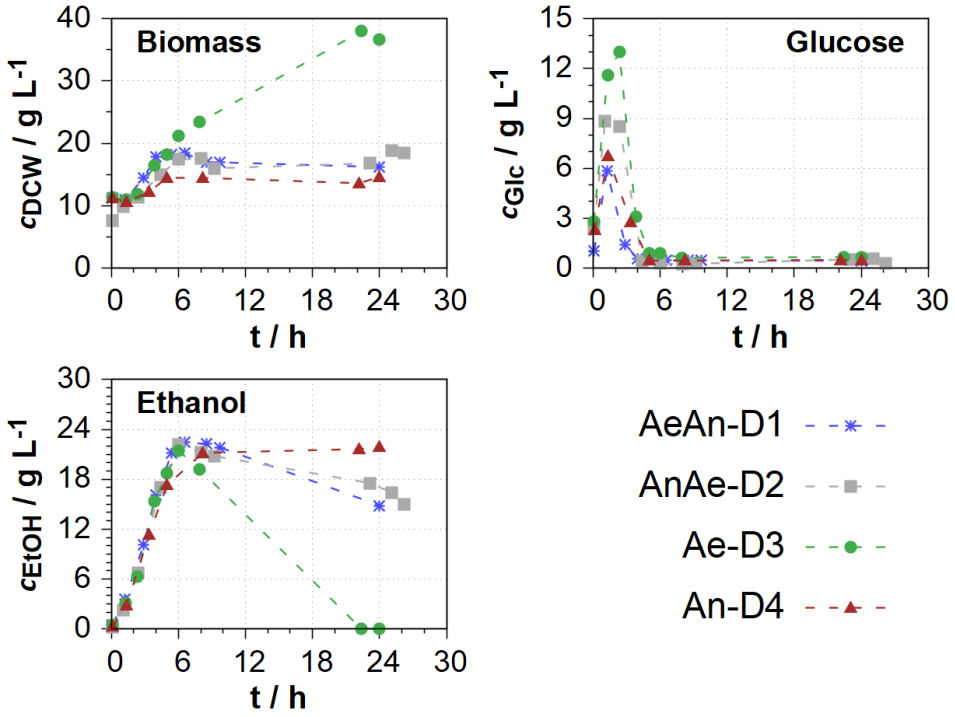


Figure 55: Comparison of offline data (c_{DCW}, c_{Glc}, c_{EtOH}) for experiments D1, D2, D3 and D4.

The final biomass density was twice as high in the aerobic single bioreactor experiment D3 as compared to the SDM experiments D1 and D2. The biomass density remained constant after the feeding stopped in the SDM experiments D1 and D2 and in the anaerobic experiment D4. In D3, biomass was produced during the ethanol consumption phase to a final density of 36 g L^{-1}.

There was no impact of the feeding inlet location on biomass growth, but there was an impact on the glucose concentrations. The peak glucose concentration was higher when the feeding inlet was positioned in the aerobic reactor. The aerobic experiment D3 shows the highest peak glucose concentration of about 13 g L^{-1}, while the lowest peaks were observed in the anaerobic cultivation (D4) with 6 g L^{-1} and in experiment D2 with 7 g L^{-1} (feeding into the anaerobic reactor of the SDM). The glucose uptake rate is therefore higher under anaerobic conditions compared to aerobic conditions, which confirms the results of batch cultivations published by Hollander et al. (1986). This may be caused by the so-called Pasteur effect in which an increased uptake rate of glucose is observed in the absence of oxygen. The influence of this effect on *S. cerevisiae* cultivations has been previously reported by other research groups (Deken 1966; Hollander et al. 1986).
Table 16 lists various yield coefficients (calculation method in chapter 4.3.4) for specific time periods as well as the ethanol uptake rates for the fed-batch experiments D1 – D4.

Table 16: Yield coefficients and uptake rates of ethanol for cultivations D1 - D4. Deviations were estimated for the case of a 5% deviation for the initial glucose or highest ethanol concentration and 5% deviation for the average biomass density. The highest calculated deviation for those cases is shown in the table.

		r_{EtOH}	$Y_{X/SC}$	$Y_{X/SC}$	$Y_{P/SC}$	$Y_{X/PC}$
Time period		During EtOH uptake	0 - 6 h	0 - 24 h	0 - 6 h	6 - 24 h
#	Combi.	$g_{EtOH}\ g_{DCW}\ s^{-1}$	$g_{DCW}\ g_{Glc}^{-1}$	$g_{DCW}\ g_{Glc}^{-1}$	$g_{EtOH}\ g_{Glc}^{-1}$	$g_{DCW}\ g_{EtOH}^{-1}$
D1	Ae-An	$8.2 \pm 0.5 \cdot 10^{-6}$	0.17 ± 0.02	0.13 ± 0.02	0.45 ± 0.04	-
D2	An-Ae	$1.0 \pm 0.3 \cdot 10^{-5}$	0.17 ± 0.02	0.20 ± 0.02	0.43 ± 0.05	0.25 ± 0.03
D3	Ae	$2.4 \pm 0.3 \cdot 10^{-5}$	0.22 ± 0.03	0.51 ± 0.03	0.42 ± 0.05	0.72 ± 0.08
D4	An	-	0.15 ± 0.02	0.09 ± 0.01	0.42 ± 0.04	-

In every cultivation about the same maximum ethanol concentration (22 g L^{-1}) was reached after 6 h. In both cultivations performed in the SDM (D1 and D2) ethanol concentration decreased and was completely depleted in the aerobic single reactor cultivation (D3).
The ethanol uptake rate r_{EtOH} of the fully aerobic cultivation D3 ($2.4 \cdot 10^{-5}$) was approximately doubled compared to the SDM experiments D1 ($8.2 \cdot 10^{-6}$) and D2 ($1.0 \cdot 10^{-5}$). A possible explanation for this difference might be the proportion of aerobic volume in the total medium volume. D3 was performed under fully aerobic conditions, whereas in cultivations D1 and D2 only half the medium volume was under aerobic conditions since only one reactor of the SDM was aerated and ethanol uptake requires aerobic conditions.
Furthermore, the ethanol uptake rate in D1 is lower compared to D2. This might be again caused by the ratio of aerated to unaerated medium volume. The aerobic reactor in D1 had

a medium volume of 0.5 L, which was about 15% lower than the volume in the anaerobic reactor (0.58 L). In D2 the aerobic reactor (R2) contained a volume of 0.62 L, which is 20% larger compared to the volume in the anaerobic reactor (R1 with 0.5 L). This was caused by the fixed medium volume of 0.5 L in reactor 1 of the SDM during the experiments (reactor with feeding). With the assumption that ethanol was only consumed in the aerobic reactor, the ethanol rates for the aerobic reactor change to $1.7 \cdot 10^{-6}$ (D1) and $1.8 \cdot 10^{-6}$ (D2). This is still about 35% lower compared to the fully aerobic experiment D3. This reduction is close to the calculated performance loss in chapter 8.1.1 for the performed batch experiments. This confirms the assumption that performance loss might be attributed to the cells constantly alternating between aerobic and anaerobic zones in the SDM. To find a possible explanation for these differences, a brief model-based investigation will be presented in chapter 0.

The oxygen fraction in the exhaust gas was lower and carbon dioxide fraction higher in D3 during ethanol uptake compared to the other processes, despite a higher aeration rate. Therefore, the oxygen uptake rate and carbon dioxide production rates were higher which was caused by the higher ethanol uptake rate.

Although the absolute ethanol uptake was lower in D2, the specific uptake rate was almost equal to D1, even though only the last four hours of the cultivation were used for the calculation[3]. In D1, r_{EtOH} was calculated with data obtained from a larger period between 10 h and 24 h. For the calculation of r_{EtOH} of experiment D3 it was assumed that ethanol was fully consumed at 17 h.

The yield coefficients $Y_{P/SC}$ were approximately equal in all fed-batch cultivations and about 10 – 15% lower than the theoretical maximum value. Thus, most of the glucose was metabolised into ethanol via overflow metabolism and anaerobic fermentation. The ethanol formed was consumed during the aerobic experiments and the SDM experiments.

The yield coefficient for biomass from glucose ($Y_{X/SC}$) over the total process time was highest in the aerobic single reactor experiment D3 with 0.51, which is close to values found in literature (Babel et al. 1983). In the SDM experiments this yield was reduced to 0.13 in D1 and 0.20 in D2. The anaerobic cultivation showed the lowest $Y_{X/SC}$ of 0.09.

Two cultivations with different feed locations, in the aerobic (D1) and anaerobic (D2) reactor of the physical scale-down model, and otherwise identical operation have not yet been

[3] The blocked nozzle caused one of the reactors in the physical scale-down system to empty from 10 h until 24 h. As a result, the data could only be used after the desired volume ratio had been restored. This made it possible to use exclusively the last four hours to calculate the ethanol uptake rate.

performed for *S. cerevisiae*. The experimental results of these cultivations indicate that the positioning of the feeding inlet did not have an influence on the process performance for the executed feeding rates. The deviations between the experimental results of D1 and D2 for c_{DCW} and c_{EtOH} and for most relevant process performance indicators was lower than 5% for most of the time, which was approximately the expected measurement inaccuracy. This is consistent with the minor differences found in the previously performed simulations (chapter 8.2.1). Since the experimental data for D1 and D2 are that similar, it is safe to assume that reproduceable experiments can be performed using the SDM.

8.2.3 Fed-batch experiment with two feeding periods (D5)

Experiment D5 was performed in the SDM with a feeding strategy comprising two separate glucose feeding periods, separated by a period without feed. This experiment was performed to investigate the behaviour of the system with a feeding interruption. Another aim of the experiment was to achieve simultaneous metabolisation of glucose and ethanol with a low volumetric feeding.

The first feeding period started at 8 h and ended at 22 h with a lower feeding rate of 0.05 mL min^{-1}; the second feeding period was shorter starting at 25 h and ended at 29 h with a higher feeding rate of 0.17 mL min^{-1}. This should lead to two ethanol consumption phases with lower ethanol concentrations compared to all other experiments performed in the SDM. Experiment D5 started with a total volume of 1.2 L with both reactors of the SDM containing the same initial medium volume. The initial medium contained of 25 g L^{-1} glucose, 25 g L^{-1} peptone, and 25 g L^{-1} yeast extract and was inoculated to a biomass density of 6 g L^{-1}. Experiment D5 had a duration of 48 h.

The feeding inlet in D5 was positioned in the anaerobic reactor (R1) since the feed in industrial-scale reactors is usually located at the top. At the top of industrial reactors, an aerobic zone is expected which is created by surface aeration with an anaerobic zone underneath (Oosterhuis and Kossen 1984). In both zones, ethanol production is expected either by overflow metabolism or by anaerobic fermentation of glucose to ethanol. Both cases are covered by the positioning of the feed inlet in the anaerobic reactor. Figure 56 visualises the experimental results of cultivation D5 with two separate glucose feeding periods.

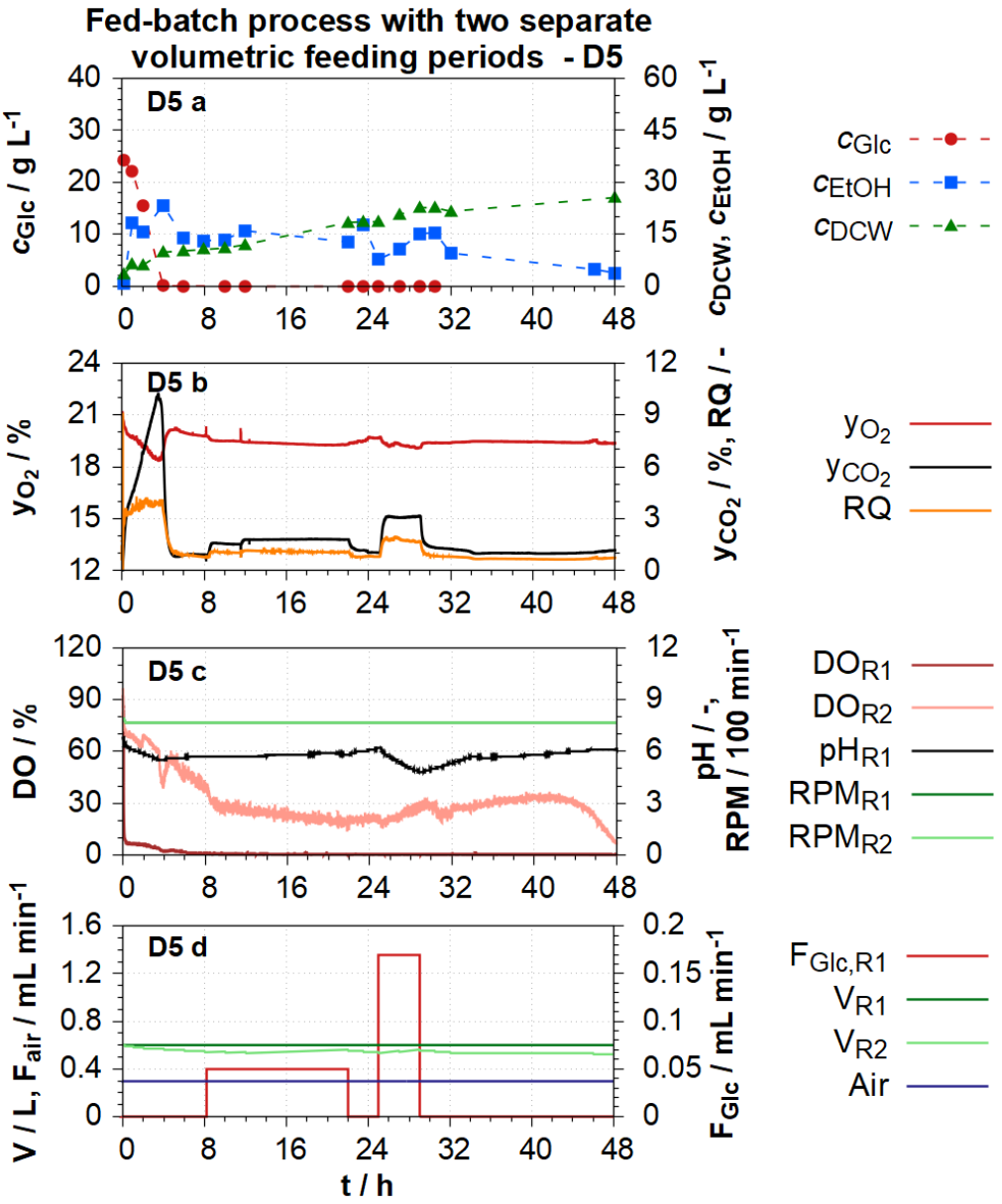


Figure 56: Experimental data of the repeated fed-batch cultivation with two separate volumetric feeding periods D5 in the physical scale-down model (the feeding inlet is located in the anaerobic reactor R1). 1st row: Offline data for c_{DCW}, c_{Glc} and c_{EtOH}. Glucose was assumed to be depleted after the second feeding period. 2nd row: Online data for the exhaust gas composition (y_{O2} and y_{CO2}) and the RQ. 3rd row: Online data for the dissolved oxygen concentration, pH value and stirrer speed of both reactors. 4th row: Feeding and aeration rate as well as the calculated volumes in both reactors.

Glucose was consumed within the first four hours of the cultivation. Afterwards, the determined glucose concentration never exceeded 0.02 g L^{-1}. During the initial batch phase, ethanol increased up to a peak concentration of 23 g L^{-1}, and afterwards decreased continuously, except during the second feeding period (from 25 h to 29 h), where the

ethanol concentration increased from 5 g L^{-1} to 10 g L^{-1}. At the end of the cultivation, an ethanol concentration of 4 g L^{-1} remained in the SDM. These changes of the ethanol concentration were also reflected in the measurements of the exhaust gas composition and the calculated RQ.

The RQ was below one after the initial batch phase, indicating ethanol uptake by the cells, except during the second feeding period when ethanol was produced. The ethanol uptake rate is lower compared to fully aerobic experiments, since it can only be consumed in the aerobic reactor of the SDM (chapter 8.1.1 and 0). Biomass density increased up to a dry cell weight density of 26 g L^{-1} until the end of the cultivation. The time course of the pH correlates to the inverse of c_{EtOH}. When ethanol was being produced, the pH decreased and vice versa.

DO was remaining between 20% and 30% in the aerobic reactor (R2) and decreased within 8 h to 0% in the anaerobic reactor (R1). During the second feeding period the DO was increasing in R2. This might have been caused by the overflow metabolism of glucose to ethanol, which requires less oxygen as compared to the ethanol oxidation during the first feed. The resulting yield coefficient $Y_{X/SC}$ was determined to be 0.38.

The experiment D5 has shown that it is possible to design a feeding strategy which leads to simultaneous glucose and ethanol uptake (first feeding period) in a system with heterogenous DO concentrations. In both previous fed-batch experiments in the physical SDM (D1 and D2), the feeding rates were higher in relative terms. These higher feeding rates resulted in ethanol production rather than ethanol uptake.

8.3 Model validation with fixed feeding profile cultivations

The first step towards model validation involves identifying the model parameters and initial values. Furthermore, the precision of the model is checked visually and with the coefficient of determination R^2 (chapter 8.3.1). This is followed by the check of the model validity using experiments which were not used in the identification process. These experiments should be performed under different experimental conditions than the experiments whose data are used for parameter identification. The experimental data from 16 of the 17 previously discussed experiments in chapters 7 and 8 were utilised for the parameter identification. All three selected experiments for model validation are fed-batch cultivations with predetermined fixed feeding profiles.

Experiment A3 with the highest final c_{DCW} of all experiments was used for model validation rather than parameter identification. Additionally, an experiment in the SDM was carried out with alternating feeding inlet position (E1). Finally, a model-based design of experiments

was performed. Experimental conditions obtained from this mDoE were utilised for fed-batch experiment E2 in the SDM (E2).

8.3.1 Model parameter identification

Since a total of 16 experiments were utilised for the parameter identification, a short computation time of the model was crucial. Consequently, a simple and thus fast model containing the base functions was created with the model generation tool. The physico-chemical submodel, which performs the calculations of DO, pH, and temperature as well as submodels for the biocatalysis, hydrolysis, foam formation and antifoam addition were not included into the process model. The DO in aerobic reactors was set to 20% and in anaerobic reactors to 0% for the simulations during parameter identification. The pH was adjusted to pH 5 since an influence of the pH on the process was not expected. The medium temperature in the simulation was set to a constant value of 28°C.

A total of 27 parameters were identified during the parameterisation. All identified model parameters, their initial and final values, as well as a short description are listed in Table 23 in appendix (Appendix chapter 12.2.1.2). In addition to the selected model parameters, the initial values of glucose concentration (c_{Glc}), ethanol concentration (c_{EtOH}) and biomass density (c_{DCW}) were identified. The identification boundaries for the initial values were set to ±25% of the analytically determined values.

The evaluation of the parameter identification can be done visually (qualitatively), mathematically (quantitatively) or both combined. Figure 57 and Figure 58 present the offline data and control variables for the 16 experiments used for the parameter identification and the resulting R^2 after identification. The graphs were used for the qualitative evaluation of the parameter identification result.

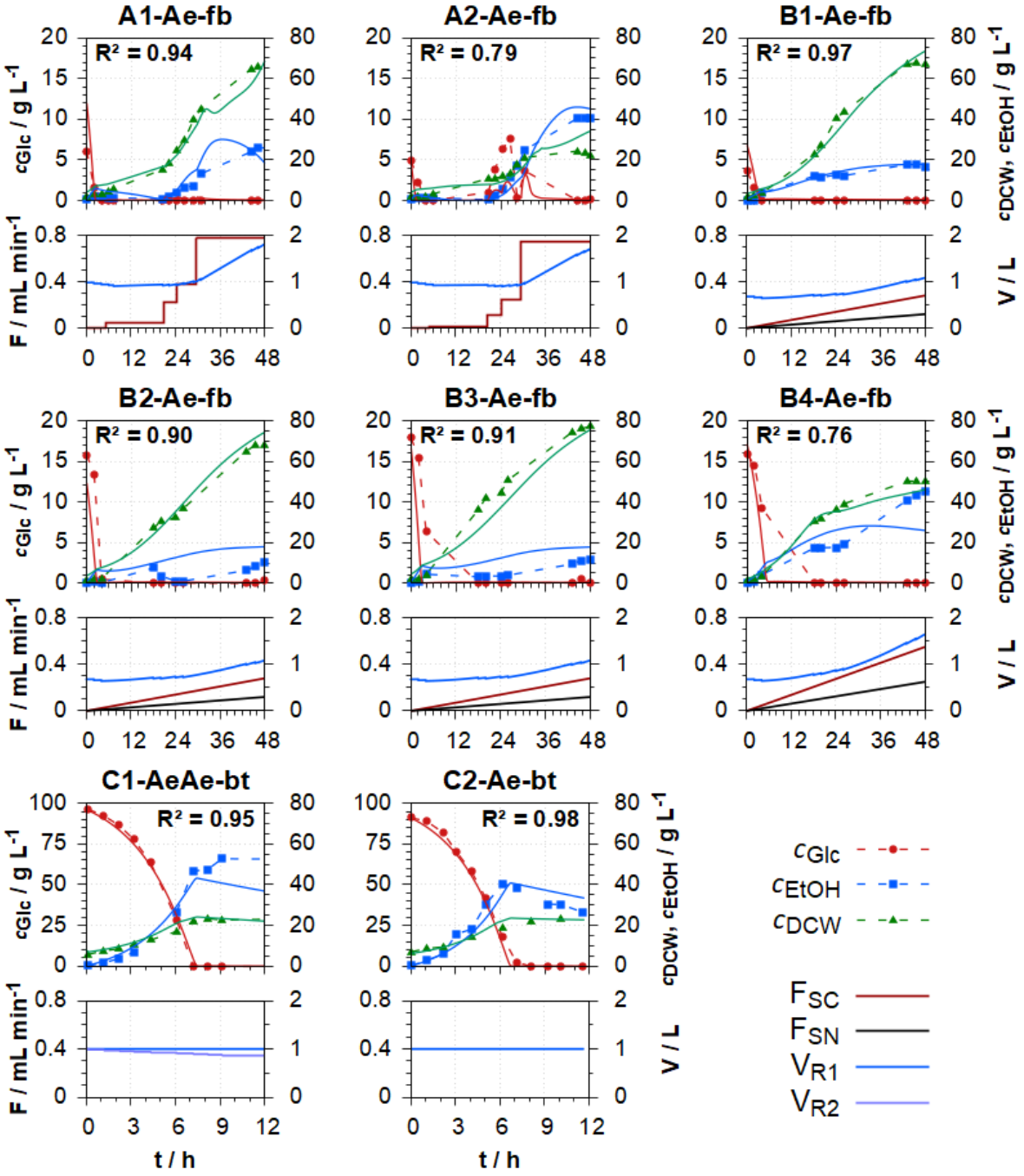


Figure 57: Simulated data in comparison with experimental data for experiments A1, A2, B1 – B4, C1, and C2 after the parameter identification. 1st rows: Experimental and simulated data for c_{DCW}, c_{Glc} and c_{EtOH}. 2nd rows: Volumetric feeding rates for glucose and nitrogen-source feed as well as reactor volumes

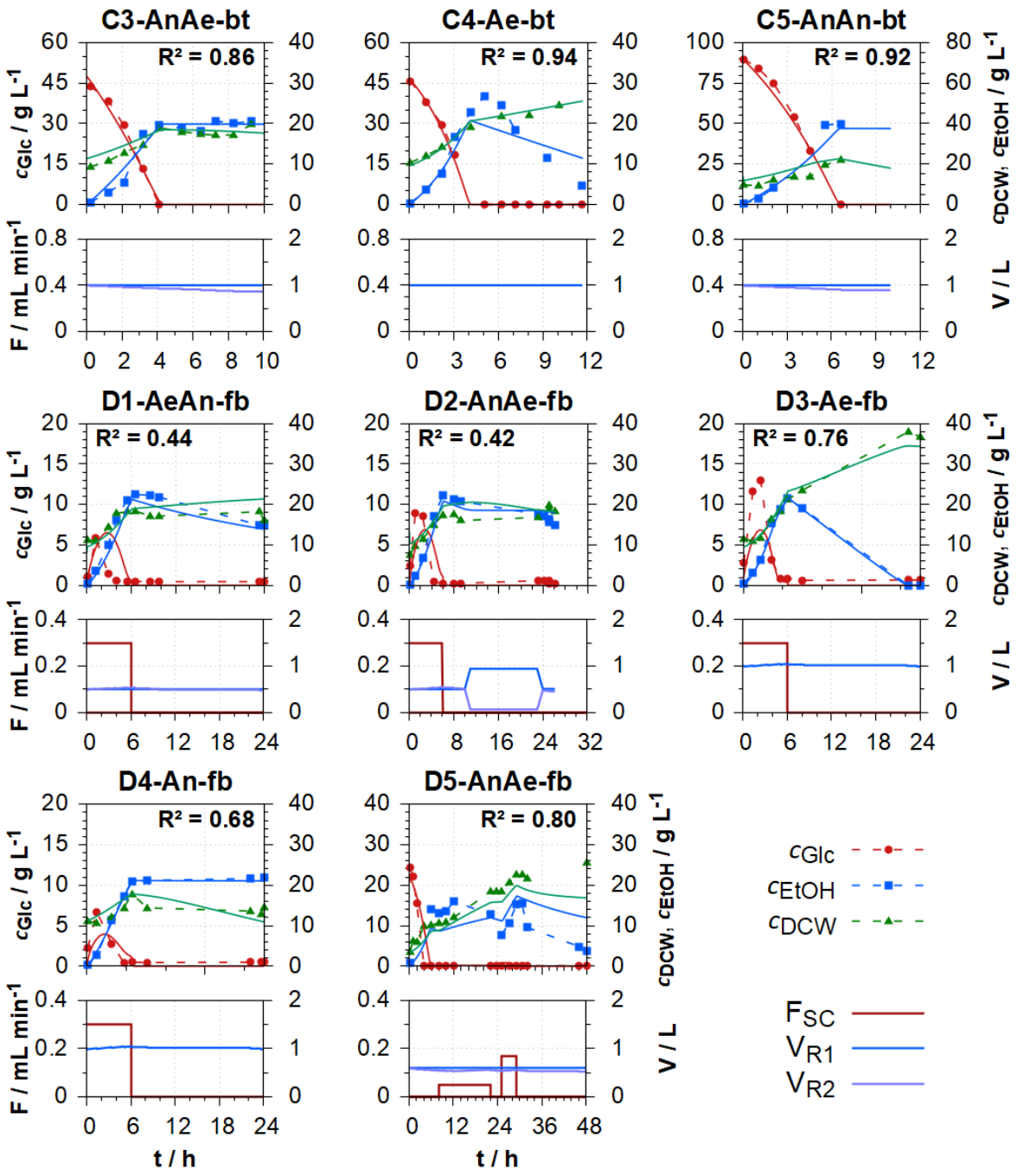


Figure 58: Simulated data in comparison with experimental data for experiments C3 – C5 and D1 – D5 after the parameter identification. 1st rows: Experimental and simulated data for c_{DCW}, c_{Glc} and c_{EtOH}. 2nd rows: Volumetric feeding rates for glucose and nitrogen-source feed as well as reactor volumes

13 of the 16 (or over 80%) of experiments are qualitatively parameterised well with a single set of model parameters. A comparison of the batch cultivations (and batch phases in the initial phase of fed-batch cultivations) reveals a near-perfect correlation between the experimentally determined and simulated glucose concentrations and biomass densities.

For this reason, the parameters related to the glucose uptake rate as well as to the biomass and ethanol production rates are well identified. Overall, this results in a remarkably high **average R^2 value of 0.93 for the batch processes** (C1 – C5) which were performed in the physical SDM and single bioreactors.

Only three fed-batch experiments (A2, B4 and D5) exhibit clear deviations between the simulated and experimental data. A2 and B4 featured glucose feeds, which eventually led to ethanol inhibitions. D5, however, featured lower volumetric glucose feed rates.

For experiments (A2, B4 and D5), the simulations are well parameterised for the batch phases. The deviations in cultivations B4 and D5 between simulated and experimental are only occurring during fed-batch phases and mainly affect the ethanol concentration while glucose concentrations are low. During these feeding phases, the sensitivity of the model to small variations of the simulated glucose concentration is high. Simulated low glucose concentrations are in a range, where the simulation might switch between ethanol production (overflow metabolism) as well as ethanol uptake (Diauxic growth). This increased sensitivity at low glucose concentrations is also expected in physical experiments. Because of this increased sensitivity, the behaviour during these cultivations is difficult to simulate. This new insight will be used to design experiments in future with the specific aim of improving the parameterisation of this behaviour at low glucose concentrations.

In A2 the simulated glucose concentration during feeding is lower than in the experimental data. The simulated biomass density is increasing until the end of the simulation, while there is no increase in the experimental data of A2. However, the relative differences between the simulated and experimental ethanol concentrations are low at around 5% throughout experiment A2.

The fed-batch cultivations determined by the mDoE application (B1 – B4) are well represented by the simulations with the identified parameter set. For cultivation B4, the mathematical model is only not able to represent the experimentally determined ethanol production during the final process phase (chapter 8.3.1). This ethanol production is a unique behaviour for this cultivation. Nonetheless, the first 24 hours of cultivation B4 are well identified.

Similarly, for the simulated process D5 the model only inaccurately simulated the experimental ethanol concentrations, initially underestimating, and later overestimating the concentrations. In the experiment the glucose concentrations during the first feeding period were low enough to enable ethanol uptake, whereas in the simulation ethanol was produced. This is indicated by the experimentally determined steep decline directly after the second feed, followed by a much lower ethanol concentration gradient. However, this lower

ethanol gradient (correlating with the ethanol uptake rate) is similar in the simulation and the experiment after feeding stopped. This result is not conclusive and may be attributed to analytical errors for the determination of the ethanol concentration immediately after feeding.

Despite these qualitative differences, the R^2 of these three experiments are good with resulting values above 0.76 (A2: R^2 of 0.79, B4: R^2 of 0.76, D5: R^2 of 0.8) which are close to 0.8. An R^2 of 0.8 was defined as the goal for the parameter identification.

Figure 59 presents the R^2 for the quantitative comparison of all 16 experiments[4].

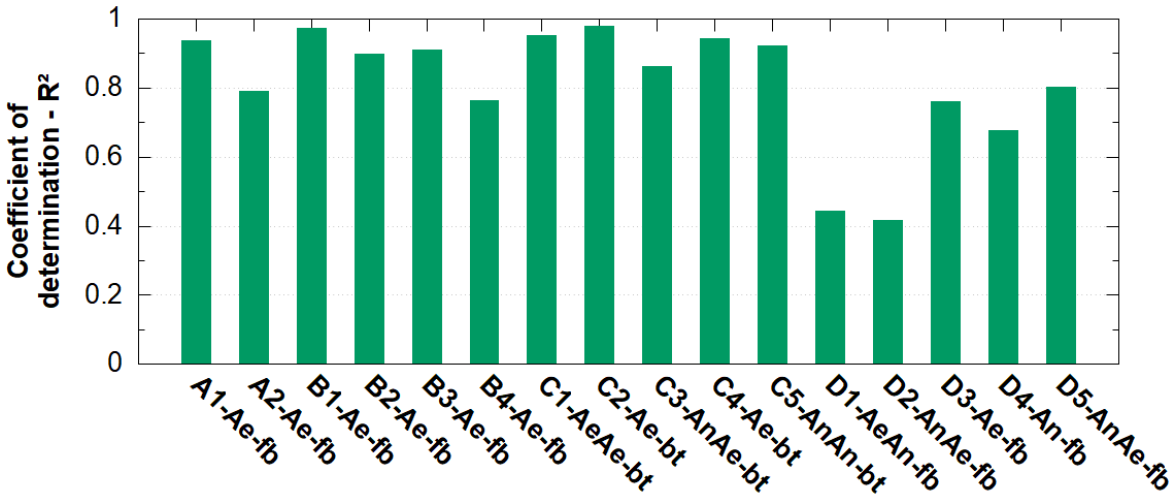


Figure 59: Resulting coefficients of determination (R^2) for the 16 experiments used for parameter identification.

Out of the 16 experiments, 10 show an R^2 above 0.8. The fed-batch experiments D1 and D2 in the SDM exhibit the lowest R^2 of 0.44 and 0.41. While these experiments are qualitatively well identified (in Figure 57), many experimental values are clustered around the mean value of the experimental data, leading to a lower R^2 without large discrepancies between the experimental and simulated data. The higher R^2 was also caused since the WMSD (chapter 4.6.1) was minimised during the parameterisation. The WMSD only considers the (squared) difference between simulated and experimental data, without taken the mean value into account (unlike R^2). For this reason, a qualitative comparison should always be carried out simultaneously.

The lower R^2 (D1 – D4) are also caused by differences between experiment and simulation regarding the glucose concentrations during the initial fed-batch phase of the cultivations. In D1 the experimental and simulated data produce similar peak concentrations of glucose,

[4] For the calculation of the coefficient of determination R^2, all glucose concentrations determined by HPLC that are below 0.5 g L^{-1} were set to 0 g L^{-1}.

but the peak occurs approximately one hour later in the simulation. This cannot be factored into the purely quantitative calculation of R^2 and is only qualitatively asserted. D2, D3 and D4 exhibit lower calculated peak glucose concentrations compared to the experimental data.

Overall, the model parameters were identified successfully. The 16 experiments are performed under aerobic, anaerobic, and non-ideally mixed conditions while employing various control strategies (batch and fedbatch). The mathematical model is therefore capable of describing experiments under homogeneous (single bioreactors) and heterogeneous (SDM model) conditions regarding heterogenous dissolved oxygen and substrate concentrations with only one set of parameters.

Model-based investigation of the reduced ethanol uptake rate

With the mathematical model utilising the identified parameter set, model-based investigations of effects, observed in the physical scale-down model, are now possible. The effect investigated in this chapter is the reduced ethanol uptake rate in the experiments under the influence of heterogeneous dissolved oxygen concentrations.

For this purpose, simulations of the two fed-batch cultivations D1 and D3 were carried out and the results are compared with the experimental results. D1 was carried out in the SDM with different dissolved oxygen concentrations in the reactors of the SDM (see chapter 8.2.1). The feeding inlet was positioned in the aerobic reactor in D1.

Experiment D3 was carried out in a single, ideally mixed bioreactor under fully aerobic conditions. The ethanol uptake rate in the aerobic reactor of experiment D1 was about 35% lower compared to cultivation D3 (performed in a single reactor).

Figure 60 displays a comparison between the SDM experiment D1 and the single bioreactor cultivation D3. The model calculates the activation (r_{Xa}) and inactivation rates (r_{Xi}) of the biomass in each reactor of both experiments which are presented in the second row of Figure 60.

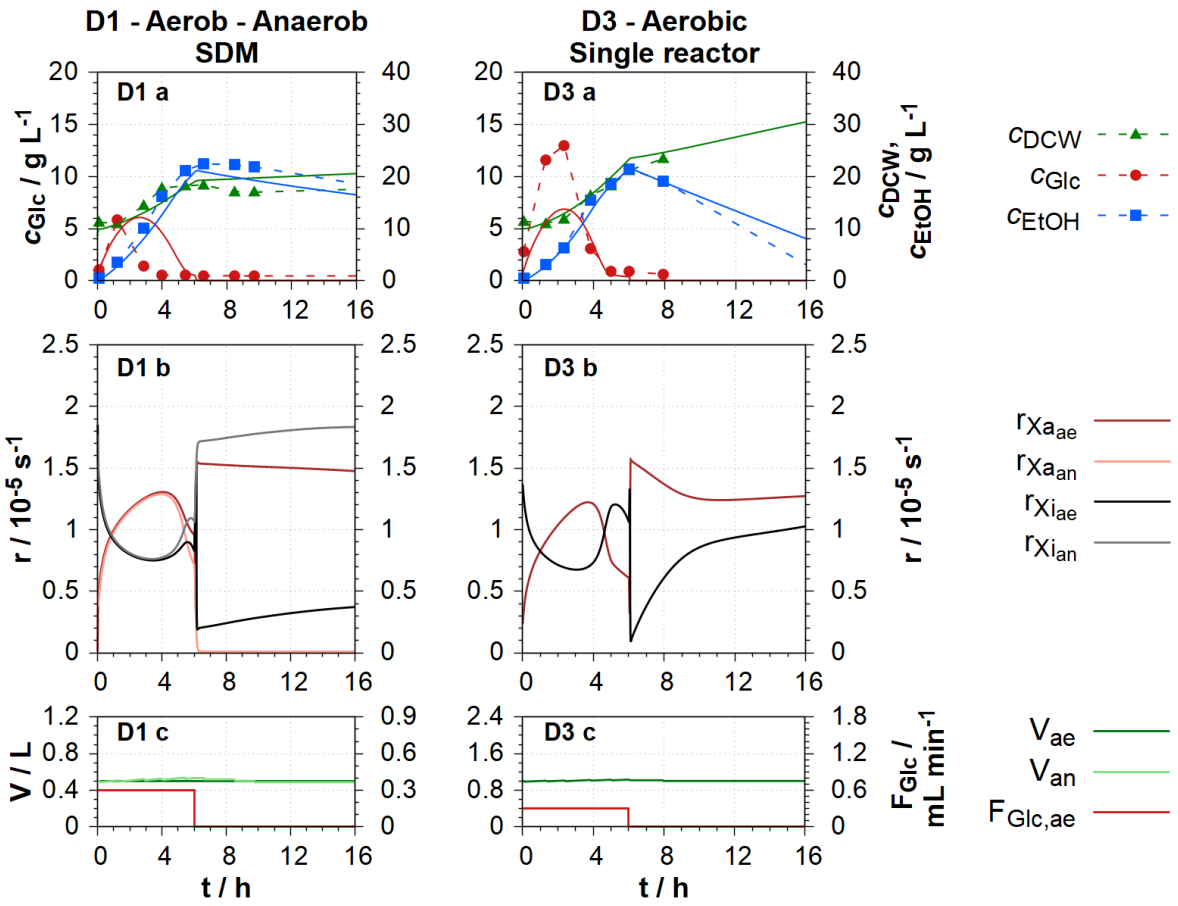


Figure 60: Experimental and simulated data for fed-batch cultivations D1 with the feeding inlet positioned in the aerobic reactor and aerobic single bioreactor cultivation D3. 1st row: Offline data and simulated cultivation courses for c_{DCW}, c_{Glc} and c_{EtOH}. 2nd row: Cumulated activation and inactivation rates of the biomass. 3rd row: Feeding rate as well as the calculated volumes.

When glucose is available through glucose feeding (first 6 h of the simulation), the biomass inactivation and activation rates are more similar between both simulations. These rates are also almost identical in both reactors of D1 compared to the second process phase after feeding stopped.

The main differences between the two simulated processes occur during the ethanol uptake phase (after feeding is stopped at 6 h). Right at beginning of this phase, the average inactivation rate in D1 is initially ten times higher compared to D3 (D1: $r_{Xi,average} \approx 1.0$, D3: $r_{Xi} \approx 0.1$). Towards a process time of 16 h, the inactivation rates converge and D1 is only 10% higher than D3 (D1: $r_{Xi,average} \approx 1.1$, D3: $r_{Xi} \approx 1.0$).

The activation rate of the biomass seems to be higher in D1 than in D3. However, this only applies to the aerobic reactors; no reactivation occurs in the simulation of the anaerobic reactor. The biomass in D1 is continuously inactivated in the anaerobic reactor of the SDM

and reactivated in the aerobic reactor. This continuous cycle reduces the ethanol uptake rate compared to the simulation of the fully aerobic, homogeneous system D3.
Furthermore, the overall activation rate in D1 is approximately 40% lower compared to D3 (D1: $r_{Xa,average} \approx 0.7$, D3: $r_{Xa} \approx 1.3$). Therefore, a greater proportion of the biomass is inactivated in D1 and no longer metabolically active. This also results in a reduction of the ethanol uptake rate in D1, as less active biomass is available. This phenomenon is also plausible for the performed physical experiments. The biomass in the anaerobic part of the physical SDM was unable to undergo metabolism, as no carbon substrate was available that could have been utilised under anaerobic conditions by *S. cerevisiae*. This would have resulted in an increasingly inactive biomass which in turn caused the lower ethanol uptake rates that were experimentally determined during the experiments in the heterogenous scale-down system compared to the experiments in the homogenous single bioreactors.

8.3.2 Experimental investigations for model validation

The mathematical model with the identified parameter values was validated using data from three experiments that were performed under different process conditions. The experimental data were not used for parameter identification. Like all experiments, any new experiments for validation should increase knowledge and provide new insights into the process, in addition to being used for model validation.
The first experiment used for model validation is cultivation A3. This cultivation had the highest final biomass density of all experiments (more than 100 g L^{-1}). The second highest biomass density is about 20% lower which gives A3 a distinctive and unique feature. This experiment was selected for the model validation because a successful validation would demonstrate the mathematical model's capability to successfully predict cultivations with higher densities than those in the data used for the parameter identification.
The second validation experiment is a new experiment in the SDM with four repeating glucose feeds with increasing volumetric rates and alternating feeding locations (E1). The feeding rates of E1 were set to lower values than in previous experiments to reduce the overflow metabolism of glucose. With this experiment the precision of the model prediction could be tested for four phases of ethanol production and uptake in the scale-down system.
Finally, a modified model-based design of experiments was performed for the optimisation of a fed-batch experiment in the SDM (E2). Three factors were optimised: (1) The volumetric feed rate of a continuous glucose feed, (2) the initial nitrogen source concentration, and (3) the initial glucose concentration. Experimental settings for one experiment were selected based on the desirability score and carried out. The aim of experiment E2 was to

demonstrate the effectiveness of the mathematical model within a model-based process design method for the design of the process operation of an experiment in a scale-down system with heterogeneous conditions (regarding DO and substrate concentration) while predicting experimental data.

8.3.2.1 Experiment A3 with highest biomass concentration

Experiment A3 was the experiment with the highest final biomass concentration of all experiments. This cultivation was not utilised for parameter identification. Figure 61 displays a comparison between the concentrations of glucose and ethanol, as well as the dry cell weight density of the experimental and simulated data.

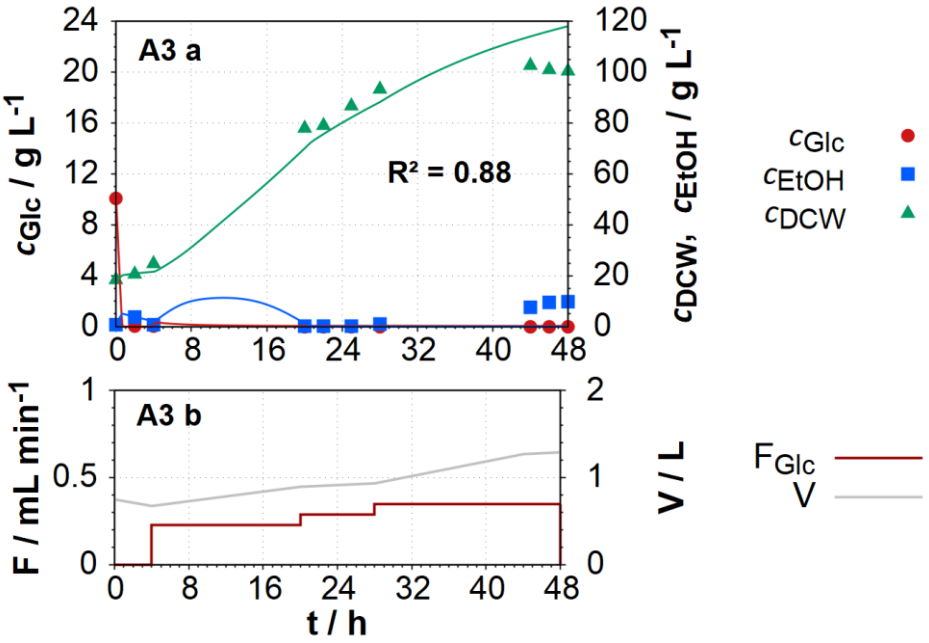


Figure 61: Comparison of simulated and experimental data of process A3. 1st row: Simulated and experimental data for c_{DCW}, c_{Glc} and c_{EtOH}. 2nd row: Volumetric feed rate and reactor volume.

Overall, a satisfactory agreement between the real experimental data of A3 and the simulation is reached with the identified parameter set. The first 32 h of the process are simulated with a high accuracy. Nonetheless, discrepancies between the simulation and experiment occur towards the end of the cultivation. The biomass growth stagnated during the experiment, whereas in the simulation a slightly decelerated increase of the biomass density can be observed. The biomass growth deceleration in the experiment caused an increasing ethanol concentration in the experiment, which cannot be observed in the simulation results. Nonetheless, the R^2 for this experiment is 0.88, demonstrating a good quantitative precision. Consequently, this part of the model validation was successful.

8.3.2.2 SDM experiment with alternating feed location (E1)

For the second model validation scenario, a new experiment was designed. Experiment E1 utilised a repeated feeding approach with increasing rates. The feed was turned on for one hour and then switched off for the same time-period. This cycle was repeated four times. The first two feeding rates were set to 0.05 mL min^{-1}, and the subsequent two to 0.1 mL min^{-1}. The glucose concentration in the feed was 500 g L^{-1}. The initial nitrogen-source concentration was set to 50 g L^{-1} (25 g L^{-1} peptone, 25 g L^{-1} yeast extract). The feeding inlet positioning alternated between the aerobic and anaerobic reactors for each cycle starting in the aerobic reactor. The temperature was controlled at 30°C, the stirrer speed was fixed at 800 rpm throughout the cultivation.

The research goal of this experiment (E1), in addition to model validation, was to determine whether there is a measurable effect on the cultivation course resulting from the cyclic interruption of the feeding. At the same time, the influence of the alternating feeding inlet positioning on a single experiment was investigated. To avoid overflow metabolism the feeding rates were set at a lower rate in comparison to the other fed-batch cultivations D1 to D4 in the SDM.

The ethanol concentration was determined in both reactors using HPLC analysis. The experimental results presented in Figure 62 display the average ethanol concentration determined in both reactors. The mixing time $t_{M,90}$ was set to 250 s which is approximately the mixing time of the other fed-batch experiments in the SDM (D1 and D2). Figure 62 presents the comparison of the experimental and simulated data for the model validation experiment E1. The model accurately predicts the cultivation with both quantitative and qualitative precision. The R^2 value for the entire experiment is 0.84, which is reflected in the high qualitative agreement depicted in Figure 62.

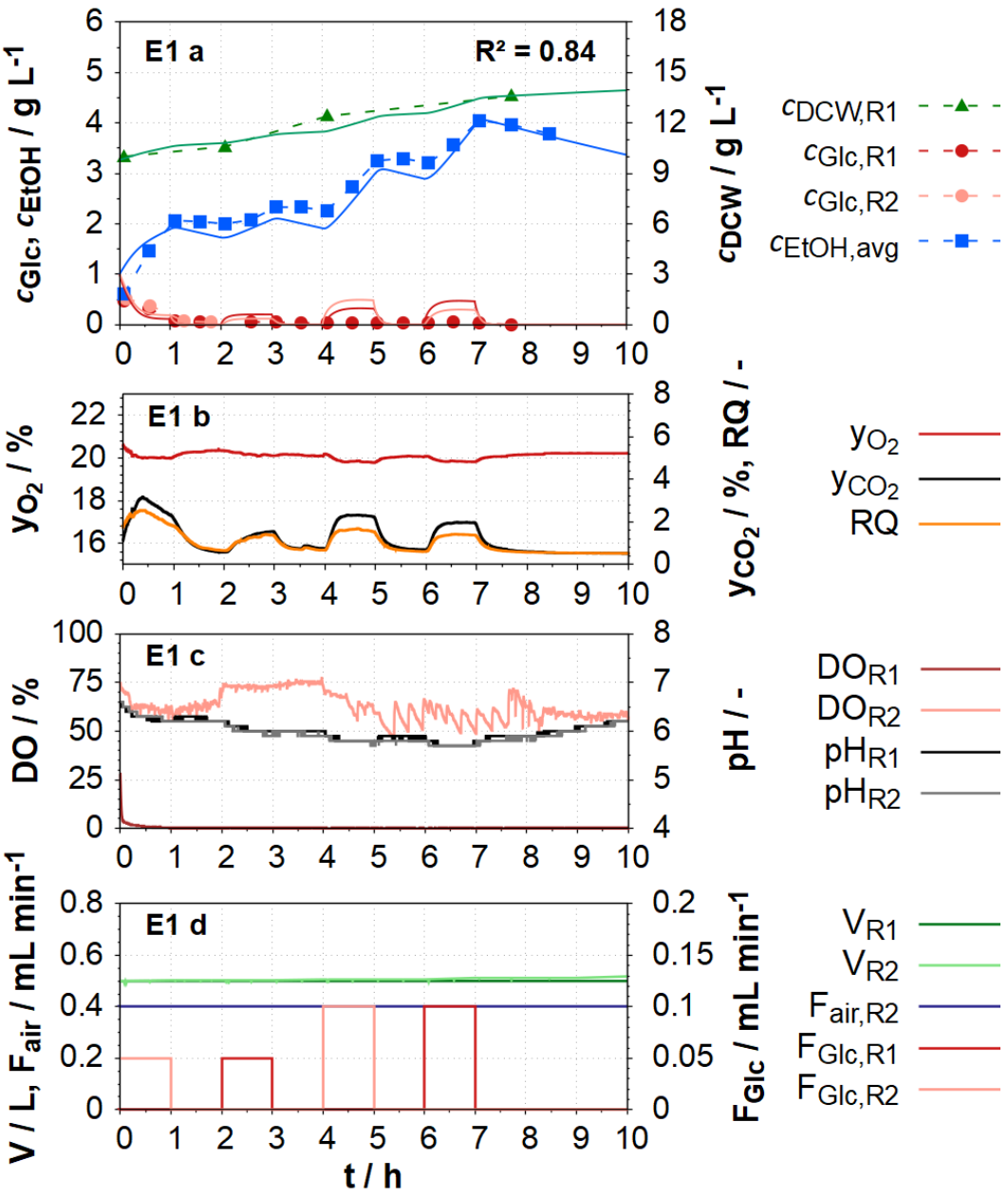


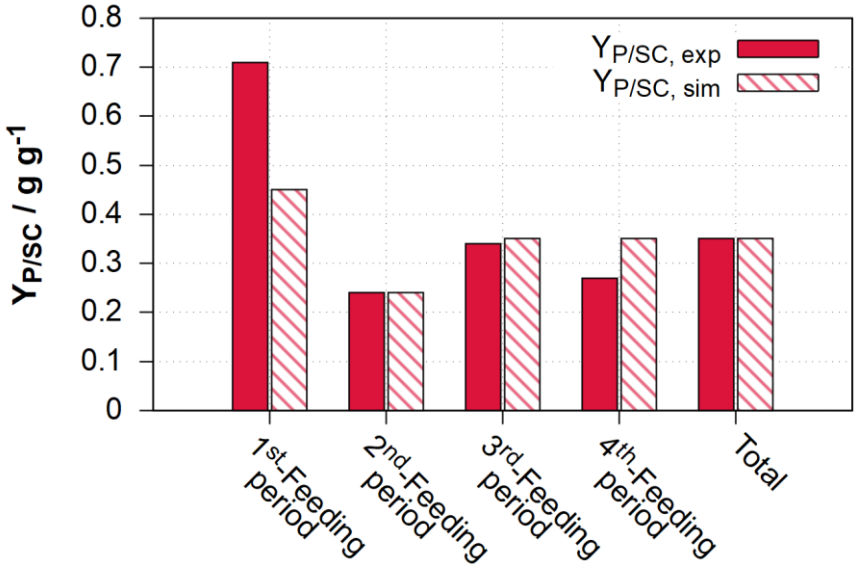
Figure 62: Comparison of simulated and experimental data in a repeated fed-batch cultivation with alternating feeding location in the physical scale-down Model (for the model validation experiment E1. 1st row: Experimental and simulated offline data for c_{DCW}, c_{Glc} and c_{EtOH}. 2nd row: Online data for the exhaust gas composition (y_{O2} and y_{CO2}) and the RQ. 3rd row: Online data for the dissolved oxygen concentration and pH value of both reactors. 4th row: Feeding and aeration rate as well as the calculated volumes in both reactors.

The initial glucose concentration (1 g L^{-1}) is fully depleted within the first hour during the initial feeding cycle, in both the experiment and the simulation. Glucose was initially determined in both reactors, however, since the difference between both reactors was negligible, glucose concentrations were subsequently determined only in the aerobic

reactor (R1) after two hours. Thereafter, the determined c_{Glc} in the experiment remained almost constant and close to 0 g L^{-1} during batch-phases without feeding and at 0.06 g L^{-1} when feeding was activated. During feeding, calculated c_{Glc} increased to 0.4 g L^{-1} which was more than six times higher compared to the experimental data. Additionally, there is an observable difference between the reactors in the simulations. The simulated reactor with feeding exhibits up to 30% higher glucose concentrations compared to the reactor without the feeding inlet. These small absolute differences between the reactors are difficult to determine analytically and could therefore potentially match the simulation results.

The ethanol concentration was slowly increasing throughout the experiment, ultimately reaching 4 g L^{-1}. The increase only occurred during feeding and the ethanol concentration decreased once feeding was stopped. The dry cell weight density was increasing from 10 g L^{-1} up to 14 g L^{-1} until the end of the cultivation. The accuracy of the model for the ethanol concentration (R^2 = 0.92) and dry cell weight density (R^2 = 0.80) is high.

Figure 63 presents a comparison between values for the yield coefficient $Y_{P/SC}$ determined from simulated and experimental data.



	Inlet positioning	$Y_{P/SC,exp}$	$Y_{P/SC,sim}$
1st Feeding period	STR2, aerobic	0.71	0.45
2nd Feeding period	STR1, anaerobic	0.24	0.24
3rd Feeding period	STR2, aerobic	0.34	0.35
4th Feeding period	STR1, anaerobic	0.27	0.35
Total	-	0.35	0.35

Figure 63: Experimental and simulated yield coefficient $Y_{P/SC}$ of ethanol from glucose for each individual feed of cultivation E1.

Overall, the simulated and experimental yield coefficients for ethanol from glucose are in the same range. However, the yield coefficient during the 1st feeding phase is higher, which might indicate that ethanol is produced from overflow metabolism and anaerobic metabolism. $Y_{P/SC}$ calculated from experimental data was also higher than the theoretically possible yield coefficient of 0.511 (calculated in chapter 4.3.4).

This might be explainable by the potential existence of small quantities of carbon sources in either in the nitrogen source (or the dry yeast used for inoculation), which are not considered in the mathematical model. If only 2% of the initial nitrogen source would consist of carbon-based substrates (e.g. sugars), corresponding to an additional 1 g L^{-1}, the yield coefficient $Y_{P/SC}$ would be reduced to 0.48. This is close to the simulated value of 0.45.

These small quantities of additional carbon-based substrates within the nitrogen source have therefore a potentially strong impact on the yield coefficient. This strong influence is caused by the low glucose input through feeding and by the short time-period used for the calculation of $Y_{P/SC}$. The influence on the yield coefficients of these carbon-based substrates within the initial nitrogen source would be significantly lower all other experiments, as considerably more glucose is considered for the calculations.

It could be demonstrated that the experimental yield coefficient $Y_{P/SC}$ are dependent on the feeding location. $Y_{P/SC}$ was lower when the feeding inlet was positioned in the anaerobic reactor (except first feeding period). This result may seem counterintuitive, considering that glucose is metabolised into ethanol under anaerobic conditions. However, it is likely, that the ethanol produced in the anaerobic reactor was instantly consumed in the aerobic reactor when the feeding inlet is positioned in the anaerobic reactor. When the feeding inlet was situated in the aerobic reactor, ethanol was produced in the aerobic reactor since the glucose concentration was above the threshold of overflow metabolism or in both reactors. However, there is no ethanol uptake possible under anaerobic conditions, resulting in higher $Y_{P/SC}$ (ethanol from glucose). In conclusion, simultaneous ethanol production and uptake takes place in the SDM when the feeding inlet is positioned in the anaerobic reactor.

This hypothesis can be reinforced if the ethanol consumed in the aerobic reactor is added to the yield calculations (of the 2nd and 4th feeding period). To estimate the quantity of the ethanol consumed during the feeding periods the average ethanol uptake rate during the phases without feeding is calculated ($r_{EtOH,avg}$ = 3.0·10^{-6} s^{-1}). With this rate the ethanol consumed in the aerobic reactor during feeding can be estimated. Considering this consumed ethanol and recalculating the yield coefficients, $Y_{P/SC}$ would be increased to 0.32 for the 2nd and to 0.31 for the 4th feeding period. These changes are almost negating the difference compared to the 3rd feeding period.

The slightly lower values can be explained because only one reactor is available in the recalculation of the coefficients, in comparison to the yield coefficient of the 3rd feeding period, where ethanol is produced in both reactors.

The increased experimental $Y_{P/SC}$ observed during the third feeding period is also reflected in the exhaust gas concentrations and respiratory quotient (RQ). More ethanol is produced; this caused an increased CO_2 production and subsequentially a higher RQ.

However, this is not reflected in the simulation, where $Y_{P/SC}$ is equal for the 3rd and 4th feeding period, and therefore independent from the feeding location. This is caused by the higher simulated glucose concentrations during feeding compared to the experiments. The concentrations in the aerobic reactor during the third feed were above the threshold for the overflow metabolism. This effect, occurring in the experiments, would be observable in the simulations if the simulated glucose concentration in the aerobic reactor would be lower (below Crabtree limit) or the threshold for the overflow metabolism higher (above glucose concentration).

Figure 64 presents the result of a simulation where the glucose feed concentrations were reduced to 100 g L^{-1} (1/5 of original concentration) leading to lower simulated glucose concentrations in the reactors of the SDM.

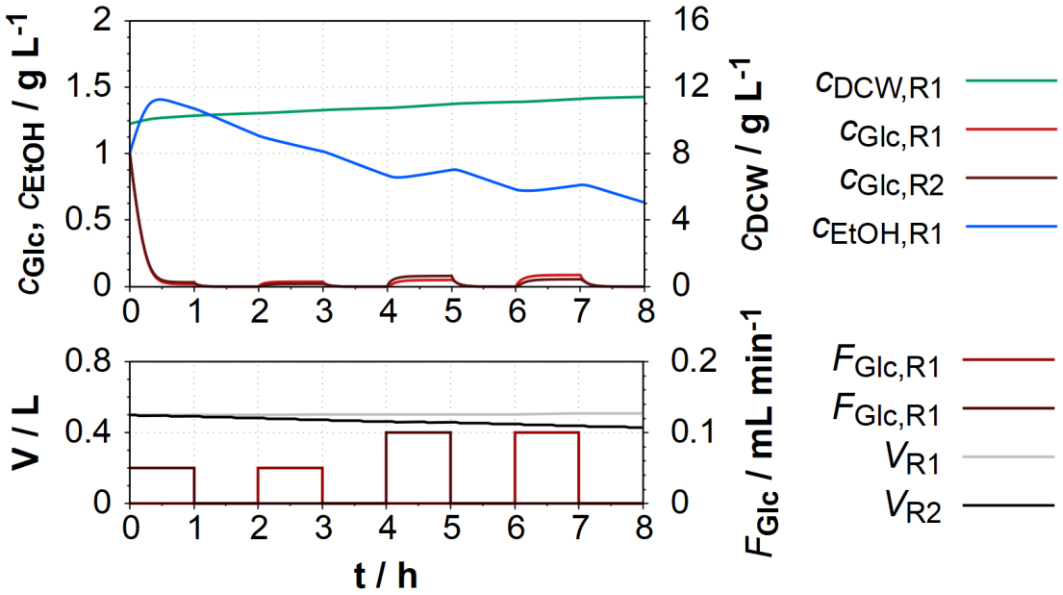


Figure 64: Simulation of cultivation E1 with decreased feed concentrations from originally 500 g L^{-1} to 100 g L^{-1}. This causes the simultaneous production and uptake of ethanol during the 4th (and 2nd) feed resulting in a lower increase of the ethanol concentration. 1st row: Simulated biomass density in R1, as well as glucose concentrations of both reactor and ethanol concentration in R1. 2nd row: Volumes of the reactors R1 and R2 and feeding rates.

During the third feeding period the yield coefficient $Y_{P/SC}$ is about 30% higher compared to the fourth feeding period (3rd feed: $Y_{P/SC}$ = 0.09, 4th feed: $Y_{P/SC}$ = 0.07). In the experimental data of E1, $Y_{P/SC}$ during the third feeding period was about 26% higher compared to the $Y_{P/SC}$ of the fourth feeding period. This relative difference of the experiment is close to the relative difference of the simulation (Figure 64). This demonstrates that the mechanistic, mathematical model is in general capable of simulating this effect caused by heterogeneous substrate and DO concentrations in the SDM. The result highlights the sensitivity of low glucose concentrations in cultivations.

In summary, an experiment was performed with four feeding periods, separated by batch periods without feeding. Such an experiment was not part of the data used for parameter identification. The model is able to accurately describe the experimental results of a fed-batch cultivation in the scale-down model with a high degree of agreement (R^2 = 0.84). The biomass growth and ethanol production were simulated with high precision. However, the experimentally determined concentrations almost remained at zero (below the limit of determination), whereas the simulated glucose concentrations were increasing during the feeding periods. Furthermore, the model is producing realistic yield coefficients.

In contrast to cultivations D1 and D2, the feeding location has an increased influence on the experiment as seen in yield coefficient $Y_{P/SC}$. This difference is most likely caused by the lower glucose concentrations in the medium during feeding in experiment E1 compared to D1 and D2. This could lead to simultaneous ethanol uptake (in aerobic reactor) and production (in anaerobic reactor) when the feed is located in the anaerobic reactor. It might therefore be beneficial, to feed into (well-mixed) anaerobic regions of the large-scale bioreactor, in order to reduce the ethanol concentration through simultaneous ethanol uptake in the aerobic reactor zones and production in the anaerobic regions. The mathematical model is able to simulate this effect when the glucose concentrations are below the limit necessary for ethanol uptake.

Furthermore, it is probable, that carbon-based substrates exist in the nitrogen-source (yeast extract and peptone) which have a measurable influence on the yield coefficients. The yield coefficients are primarily influenced when the amount of glucose considered in the calculation is low (in comparison to the available amount of the nitrogen source).

Furthermore, the experiment E1 demonstrated that simultaneous uptake of glucose and ethanol in reactors and reactor systems with increased mixing times (in this case, four minutes) is plausible and now experimentally proven for fed-batch cultivations of *S. cerevisiae*.

8.3.2.3 Model validation utilising model-based Design of Experiments

The primary goal of the third validation experiment was to demonstrate that a biotechnological experiment under controlled heterogeneous dissolved oxygen and substrate concentrations can be systematically designed with model-based design of experiments (research question no. 2). This cultivation was carried out in the physical SDM with one non-aerated reactor resulting in heterogeneous conditions regarding the dissolved oxygen concentration. The experiment was inoculated with 10 g L^{-1} of dry yeast. The initial volume in both reactors is 0.5 L, the aeration is located in STR 1 at a rate of 0.5 L min^{-1}.

The developed mathematical model was utilised with the parameter set identified in chapter 8.3.1. The feeding trajectories were designed using the modified mDoE method introduced in chapter 5.3. 2000 different experimental settings were simulated in a three-factorial design space. A Monte-Carlo based uncertainty quantification was not performed during parameter identification resulting in one parameter set since enough data from 16 experiments was available (chapter 8.3.1). Therefore, each point in the design was simulated only once and the desirability score was replaced by selected values of process variables (e.g. maximum biomass density or ethanol concentration). The experimental settings were chosen to maximise the final dry biomass density $c_{DCW,end}$. Three factors were selected for (process) optimisation / maximisation of the final dry biomass density:

1) The initial glucose concentration $c_{Glc,ini}$.
2) The combined initial nitrogen source concentration $c_{N,ini}$ consisting of yeast extract and peptone in equal proportions. This approximate ratio was also used for the experiments utilised for the parameter identification.
3) The volumetric feeding rate of a glucose solution with a glucose concentration of 500 g L^{-1}. Feeding was performed between 4 h until 8 h. The feeding inlet was positioned in the aerobic reactor.

Table 17 displays the boundaries for these factors used for the calculation of the mDoE.

Table 17: Boundaries of the selected factors for the mDoE. The factors are the volumetric feed rate of a glucose feed, as well as the initial glucose and nitrogen source concentration. The glucose feed has a concentration of 500 g L^{-1}. The nitrogen source is consisting of peptone and yeast extract with equal fractions.

	Unit	Lower Boundary	Upper boundary
$c_{Glc,ini}$	g L^{-1}	0.0	40.0
$c_{N,ini}$	g L^{-1}	0.0	40.0
F_{Glc}	mL min^{-1}	0.0	1.0

Result of the model-based Design of Experiments

2000 different experimental settings within the design space were simulated with the (modified) mDoE utilising the same parameter set as the previous two validations. The results of the mDoE may also be used for a sensitivity analysis regarding the chosen factors without further additional computing effort, which is a great advantage of this method.

Figure 65 shows the final simulated ethanol concentrations and biomass densities of each setting in the design space. The settings selected for experimental implementation are marked with “E2”. Four further settings for illustrative purposes are marked with “A” to “D”.

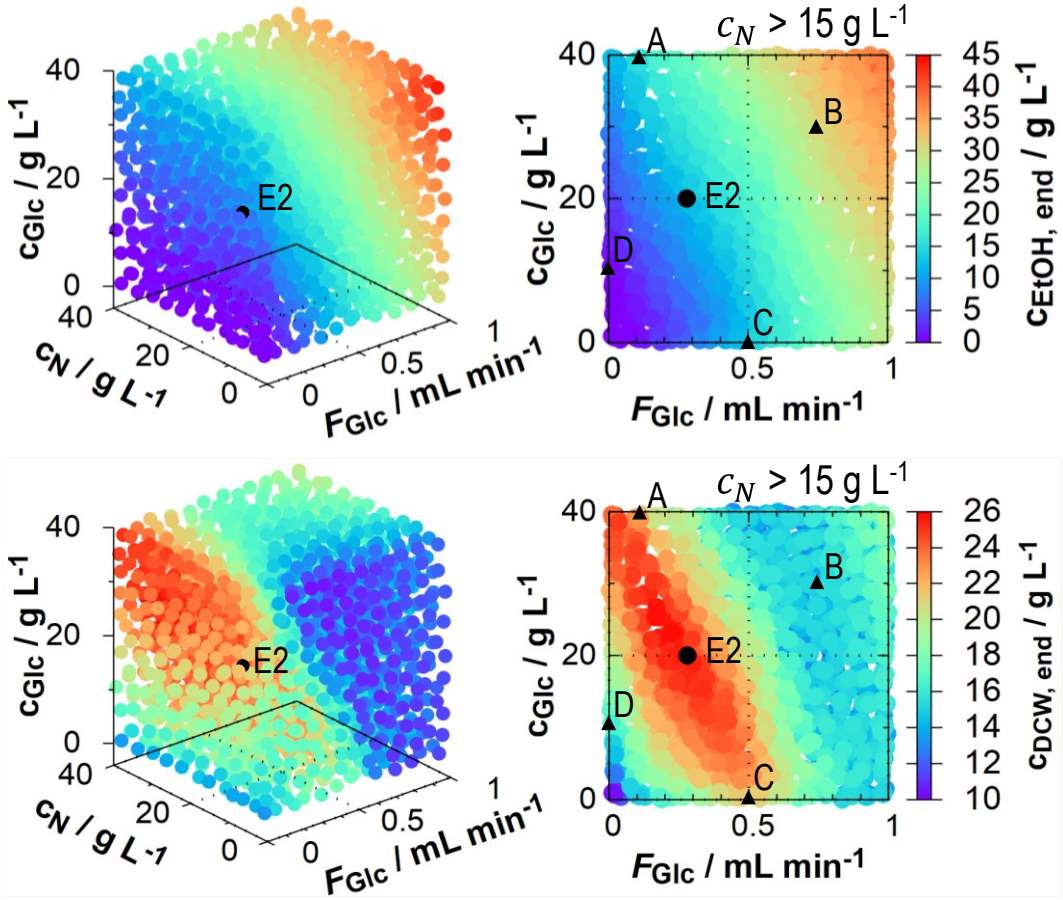


Figure 65: Final ethanol concentrations (1st row) and dry cell weight densities $c_{DCW,end}$ (2nd row) of each point in the mDoE design. The selected experimental setting (E2) and four further illustrative settings (A – D) are symbolised for the 2D plot. The individual simulation results of these points are illustrated in Figure 66. Left side: 3D plots of all 2000 points in the design space. Right side: 2D plots for c_{Glc} and F_{Glc}; all points are for a nitrogen-source concentration above 15 g L^{-1}.

The higher the initial glucose concentrations or the glucose feeds, the higher the final ethanol concentrations, which is both expected and plausible (top row of Figure 65).

The initial nitrogen concentration has a negligible effect on the simulation results. This is seen in the gradients parallel to the c_N axis. These gradients are low (indicated by a weak change in colour, parallel to the c_N axis)., provided enough of the nitrogen-source is available to support biomass production throughout the entire process. The final biomass densities are the highest when the initial nitrogen-source concentration is at 40 g L^{-1}. Therefore, all selected points have the same initial concentration of the nitrogen source (20 g L^{-1} of both peptone and yeast extract), which should support growth throughout the cultivation.

The experimental settings of E2 are in the centre of the optimal area with the highest simulated final biomass densities. In this area, small deviations from the settings only have a small influence on the simulation results.

The illustrative points ("A" – "D") are chosen in different regions of the design that could be of interest. Figure 66 presents the simulation results for the chosen experimental conditions and the four illustrative points selected in the mDoE.

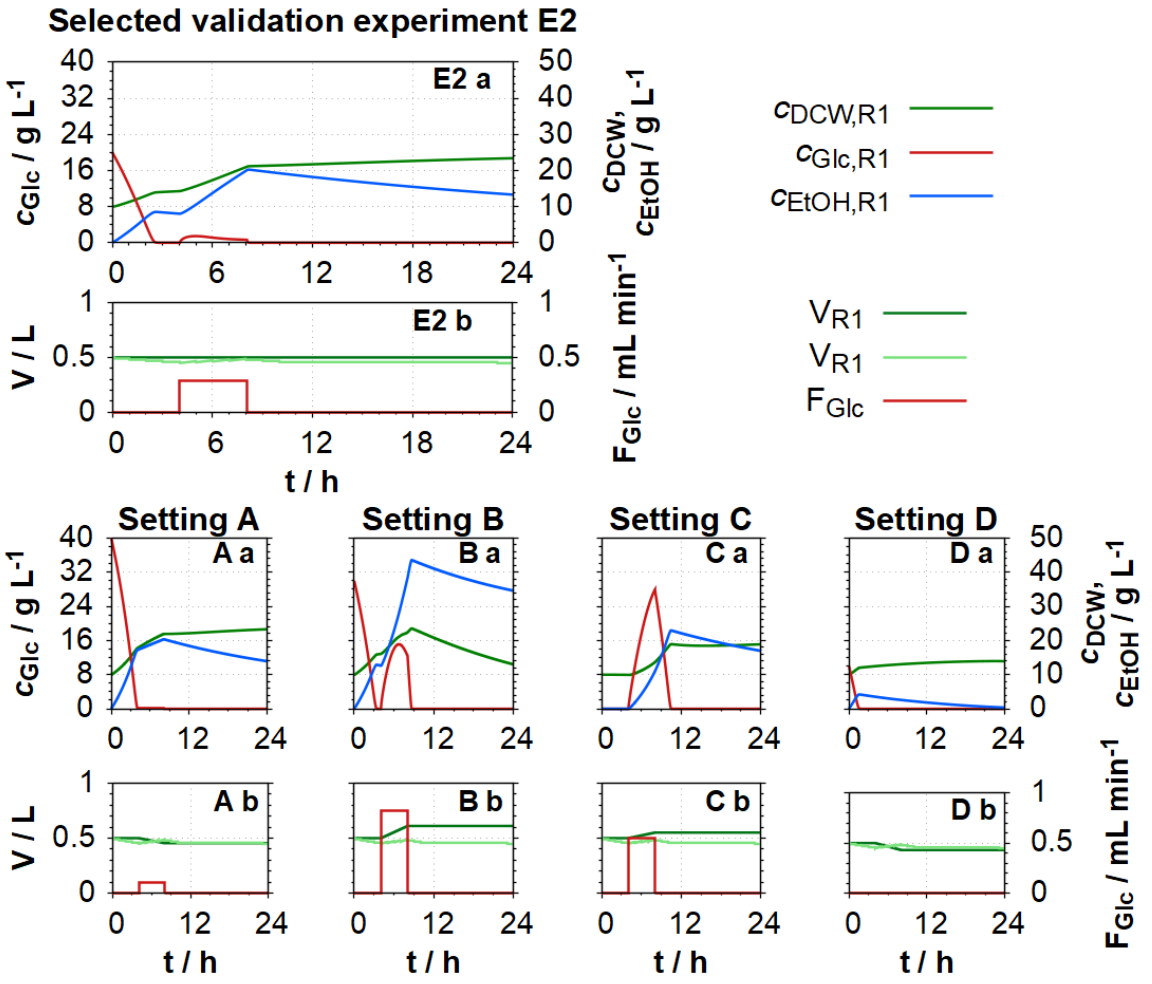


Figure 66: Simulations of selected experimental settings maximising dry cell weight density determined by mDoE for the model validation experiment E2 and simulations of four other settings in the design space (A – D).

Setting A is in a region with a high initial glucose concentration and low feed rate. This simulation almost reached the final biomass density of simulation E2 with 22.5 g L^{-1}.
Setting B has a high initial glucose concentration and simultaneously an increased feed rate, which resulted in an increased production of ethanol (overflow metabolism). A maximum ethanol concentration of 43 g L^{-1} is produced which led to ethanol inhibition and to a decreasing biomass density.
The simulation of the illustrative **setting C** is missing an initial batch phase; all glucose enters the system through the glucose feed. The process was initially glucose limited without biomass production due to the lack of a carbon source. The high feeding rate led to an ethanol production of 26 g L^{-1} which is slightly above the identified ethanol inhibition limit of approximately 25 g L^{-1}.
The process in **setting D** was glucose limited and only has a batch phase with a low initial glucose concentration leading to a reduced biomass production compared to the other simulations reaching only 15 g L^{-1}. Due to the low glucose feed, ethanol is only produced to a concentration of 4 g L^{-1} and is fully consumed until the end of the process.
The initial glucose concentration of the **selected experimental setting E2** was determined to be 20 g L^{-1} and the initial nitrogen-source concentration at the upper border of the design space at 40 g L^{-1}. The feeding rate was determined to be 0.28 mL min^{-1}. At the end of the simulation a final dry cell weight density of 23.5 g L^{-1} and an ethanol concentration of 13.5 g L^{-1} were reached. This experiment was carried out based on the experimental settings determined by the mDoE and the initial conditions. The simulation results of process E2 as well as the simulations of the illustrative settings are all plausible.

Experimental execution (E2)

The experimental settings in the centre of the optimal area with the highest simulated final biomass densities were carried out. The predicted simulation results of these conditions were compared with the experimental data. This comparison was used for the validation of the mathematical model.
Figure 67 presents the experimental results in comparison to the simulation of the mDoE. The experimentally set initial glucose concentration and the initial dry cell weight density were about 15% lower than initially planned, the depicted simulation was adapted accordingly.

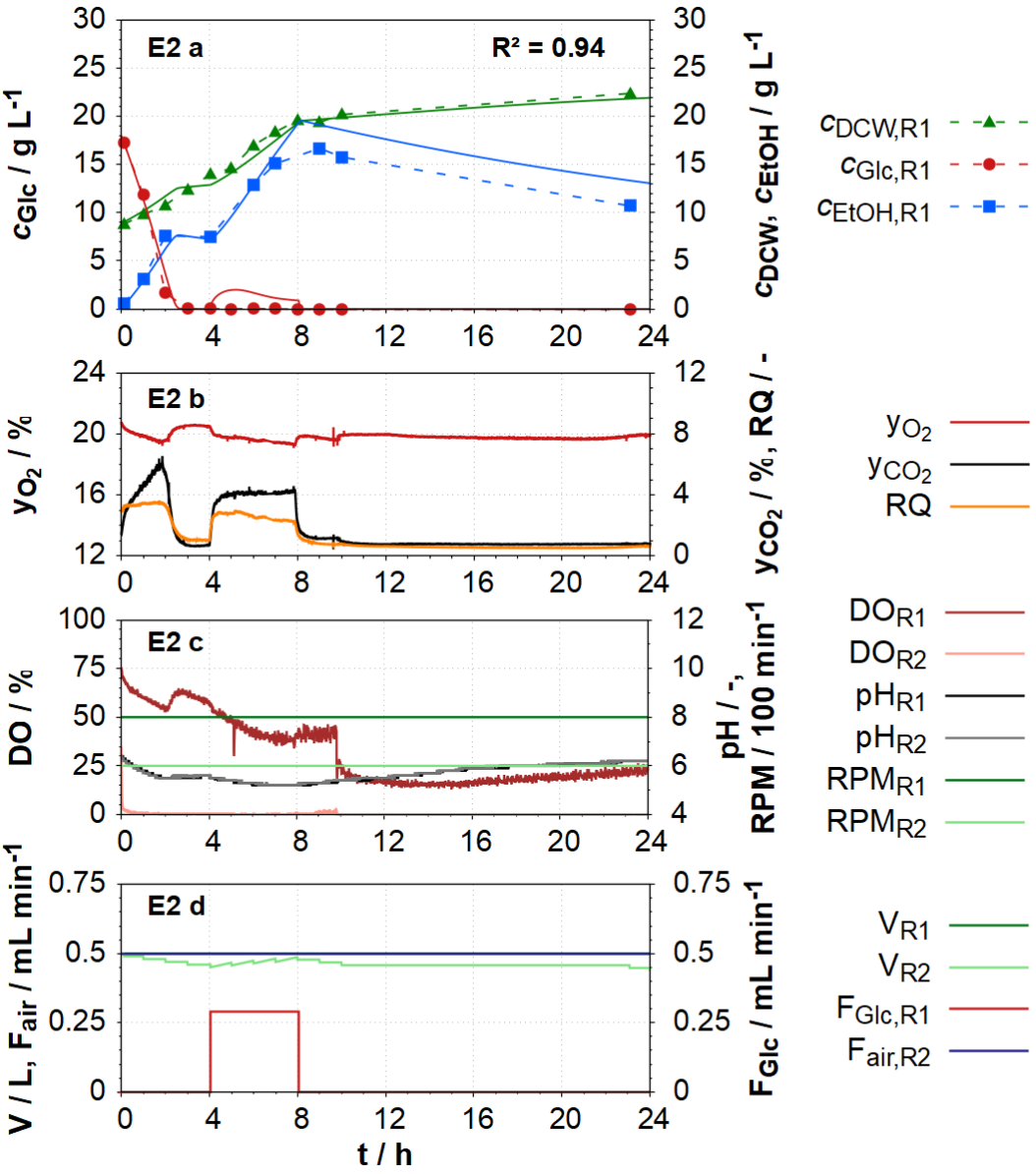


Figure 67: Experimental data of E2 in comparison to simulated data. 1st row: Experimental and simulated offline data for c_{DCW}, c_{Glc} and c_{EtOH}. 2nd row: Online data for the exhaust gas composition (y_{O2} and y_{CO2}) and the RQ. 3rd row: Online data for the dissolved oxygen concentration and pH value of both reactors. 4th row: Feeding and aeration rate as well as the calculated volumes in both reactors.

The initial glucose was fully depleted after approximately two hours. Biomass was produced up to a density of 22 g L^{-1}, ethanol was produced to a maximum concentration of 17 g L^{-1} at the end of the feeding period and was reduced to 10 g L^{-1} until the end of the process. During the initial batch phase and the subsequent feeding phase, the RQ reached values larger than 1, indicating ethanol production, which is reflected by the offline data. During phases where ethanol was consumed the RQ decreased to values below 1. The DO remained around or above 20% in the aerated, aerobic reactor and around 0% in the

anaerobic reactor. The measured pH values decreased during ethanol production and increased when ethanol was consumed. At the end of the experiment the pH value reached 6.2. The pH course of E2 is in well agreement with all other experiments.
The simulation predicted the experimental data accurately with a calculated R^2 = 0.94. The initial glucose uptake phase (batch phase) and therefore the glucose uptake rate was predicted well. During feeding the glucose concentration in the simulation was at 2 g L^{-1}, whereas in the experiment glucose concentrations were below detectable levels.
The dry cell weight density was almost perfectly predicted, reaching an experimental and simulated final density of about 22 g L^{-1}. The maximum difference between the simulated and the experimental biomass density is about 1 g L^{-1}.
Slightly lower ethanol concentrations were consistently experimentally determined after the feeding stopped at 8 h compared to the simulated values. In the experiment ethanol concentration reached a peak of about 17 g L^{-1} which is only 2 g L^{-1} less than in the simulation. After the feed was stopped, ethanol was consumed, and an ethanol concentration was consumed until 11 g L^{-1} were left at the end of the cultivation. In the simulation the final ethanol concentration was 13 g L^{-1} which is still 2 g L^{-1} more compared to the experiment. Therefore, the ethanol uptake rate was predicted accurately since the ethanol concentration decreased similarly in experiment and simulation (and the biomass density is almost identical in experiment and simulation).
In summary, the mechanistic model is capable of predicting a process planned with the modified mDoE in a physical scale-down model imitating heterogenous dissolved oxygen concentrations that might occur in principle also in large scale production systems. The experimental settings were selected to simultaneously validate the model and optimise the process. An optimisation of the cultivation conditions was achieved since the experimental settings were selected in the area of the highest expected biomass production. The resulting R^2 of 0.94 for the experiment shows an exceptionally good precision for the prediction.

8.4 Discussion and conclusion

In this study the combined influence of heterogeneous substrate and DO concentrations in a scale-down reactor system on *S. cerevisiae* cultivations was investigated for the first time. For this investigation, a total of 19 cultivation experiments were performed.
The data from 16 experiments were used to identify one single parameter set that is able to represent processes with different operational strategies (batch and fed-batch), in both single (homogeneous) bioreactors and in the scale-down model (heterogeneous). The parameter identification was successful, and the simulation demonstrated high

precision with the identified parameter set in most experiments (chapter 8.3.1). Simulations with the identified parameter set have shown that the lower ethanol uptake rates in cultivations in non-ideally mixed systems can be explained by higher biomass inactivation and lower biomass activation rates resulting in a proportionally more inactive biomass. Furthermore, model validation was successfully performed with the data of three experiments with predetermined fixed feeding profiles. The model and the parameterisation thus reflect the information and knowledge of all 19 cultivation experiments.

Two of the validation processes were performed in the SDM with controlled heterogeneous conditions regarding the dissolved oxygen and glucose concentrations (E1 and E2) and one experiment in a single bioreactor (A3). These experiments were not utilised for parameter identification and had different goals than those of the identification experiments.

Table 18 presents the coefficients of determination (R^2) for the three validation experiments, as well as those for the concentration of glucose, ethanol, and biomass individually.

Table 18: Resulting R^2 for the three validation experiments A3, E1 and E2. The R^2 for the individual offline variables are also calculated and shown.

Experiment		R^2			
		c_{DCW}	c_{Glc}	c_{EtOH}	total
A3	Ae	0.90	0.99	-0.51	0.88
E1	AnAe	0.80	-0.47	0.92	0.84
E2	AeAn	0.96	0.96	0.90	0.94

All experiments have high total R^2 of above 0.84 and even as high as 0.94 for the third experiment E2. Almost each R^2, even for the individual concentrations, is above 0.8, with only two exceptions for ethanol in A3 and glucose in E1, both of which are below zero. This is mainly caused by the method for the R^2 calculation, which is using the mean value of the experimental data. Therefore, R^2 is often low or even negative when there are no significant changes in the values of either the simulated or experimental variables.

The R^2 for ethanol in **experiment A3** is at -0.51 since the simulated concentrations remain consistently at about 0 g L^{-1}. However, the experimentally determined ethanol concentration was increasing at the end of the cultivation. This small deviation has a huge influence on the resulting R^2 for the ethanol concentration but less influence on the R^2 for the full set of the determined offline variables, which has a value of 0.88.

Experiment E1 was carried out with an alternating positioning of the feeding inlet between both reactors of the SDM. This resulted in heterogeneous glucose concentration and

dissolved oxygen concentration. The experimental data indicated that phases of simultaneous ethanol production and consumption in the SDM may occur when glucose is fed into the aerobic reactor. This could generally be confirmed with the simulations of the mathematical model when the glucose feed concentration was significantly reduced. Since those differences are quite small, the R^2 for E1 is still high with 0.84.

The settings of **experiment E2** were determined with the application of model-based design of experiments. This experiment has an initial batch phase followed by a fed-batch phase with ethanol production and an ethanol consumption phase. Therefore, this experiment featured three relevant experimental phases of large-scale cultivations. The experiment also demonstrated that the simultaneous production of ethanol and consumption is possible in a non-ideally mixed reactor system. Only three experiments used in the parameter identification have a resulting R^2 higher than experiment E2 (R^2 = 0.94). This demonstrates that the mathematical model is able to be used for the design of process with fixed feeding profiles in systems with heterogeneous environments (in the SDM).

Overall, the total R^2 for the experiments used for model validation are on the same level (median R^2 at 0.88, average R^2 at 0.89) as the R^2 for the experimental data used for parameter identification (median R^2 at 0.88, average R^2 at 0.81). The qualitative visually performed comparisons in the previous chapters are supporting these high R^2 values.

In conclusion, the processes in the physical model were utilising different operational strategies, mixing times as well as heterogenous dissolved oxygen and substrate concentrations (chapters 2.4 and 3). The novel developed mechanistic, compartmentalised model structure is capable of describing 16 cultivations and was able to predict 3 cultivations under homogeneous and heterogeneous conditions with a single set of parameters with both high quantitative and high qualitative accuracy. The mathematical model and its structure are therefore successfully validated.

In addition, the approach of using a physical scale-down model in combination with a mathematical process model used in model-based design of experiments was validated together for the first time for biotechnological fed-batch cultivations (*S. cerevisiae*) under non-ideal mixing conditions.

9 Application case study

There are performance differences for the industrial yeast production between large scale production and laboratory scale cultivations. These differences could be caused by non-ideal mixing conditions and resulting concentration differences within the large-scale reactors. The validated mathematical model shall be utilised to support explanations of the performance differences.

For this study, data from a pilot reactor cultivation were provided by an industrial partner. The data from this process are described in chapter 9.1. First, experiments on laboratory scale with downscaled process conditions were performed to quantify the influence of potential non-ideally mixed conditions of the pilot scale. These experiments were performed in single bioreactors with very short mixing times (considered as ideally mixed) to determine whether the longer mixing times of the pilot reactor influence the process performance (chapter 9.2.1). Furthermore, experiments with heterogeneous substrate concentrations under fully aerobic conditions were performed. The mixing time was selected to be in the expected order of magnitude of the pilot reactor were carried out in the SDM (chapter 9.2.2). The data obtained from these experiments were used to adapt the mathematical models to the new process. The parameter set from chapter 8.3.1 is not directly applicable as (industrial) media were used, which are different from the media used in experiments described in the previous chapters.

A process performed with the application of an NMPC is described and was carried out with the objective of obtaining data with a low ethanol production for parameter identification. Further objectives of this experiment were the improvement of the process performance and the reduction of the difference between pilot and laboratory scale (chapter 9.3).

Furthermore, the resulting parameter set of the NMPC application was used for a simulation study with the goal to explain the differences between the scales which might be caused by non-ideally mixed conditions (chapter 9.4).

9.1 Pilot reactor process

The data from the pilot reactor cultivation was provided by an industrial partner. The initial reactor medium in the pilot reactor cultivation (and the other cultivations) was only containing the nitrogen source, which was an industrially used yeast extract medium. The concentration of the yeast extract was high enough to support biomass growth throughout the complete cultivation. The applied feed had a sucrose concentration of 216 g L^{-1}.

The pilot process reactor has a maximum working volume of 1.6 m^3. The data provided from the industrial partner were combined from two separate processes with identical conditions

and feeding rates. In the first process the offline data and the DO were determined, while the online data for the exhaust gas concentrations were measured in a second process with identical initial conditions and process control. The pilot reactor recorded the volume when it reached certain levels. The volumetric feeding rate presented has been calculated from the mean feeding rate of the two pilot-scale processes, adjusted to intervals of 30 minutes with constant rates. Subsequently, a scaled down version of this feeding rate was utilised for the experiments on the laboratory scale as part of this case study.

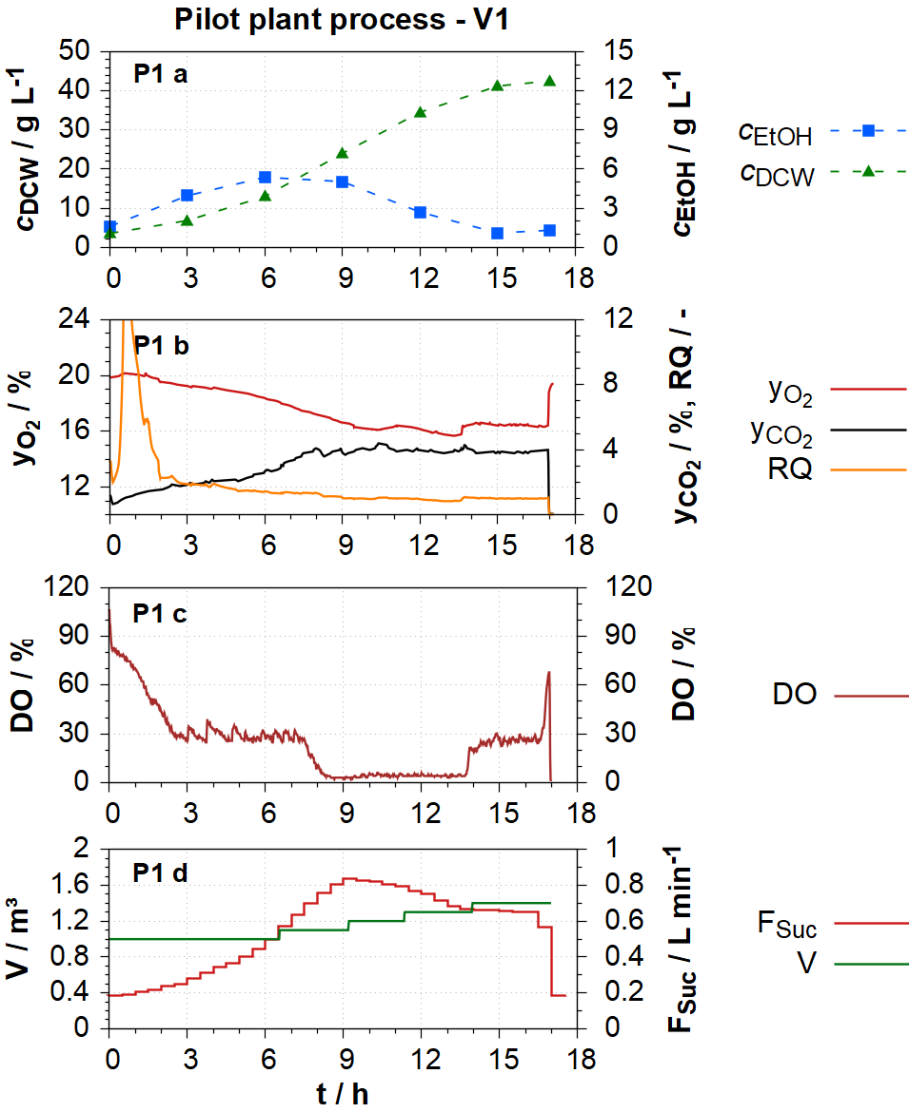


Figure 68: Experimental data of two experiments in a pilot reactor. 1st row: Experimental offline data for c_{DCW} and c_{EtOH}. 2nd row: Online data for the exhaust gas composition (y_{O2} and y_{CO2}) and the RQ. 3rd row: Online data for the dissolved oxygen concentration. 4th row: Feeding rate and the determined volume.

In the pilot reactor process a dry cell weight density of approximately 42 g L^{-1} and a peak ethanol concentration of 5.3 g L^{-1} were reached. The ethanol concentration decreased to

1 g L^{-1} until the end of the process. The increase and decrease of the ethanol concentration is consistent with the RQ value, which is above one until 8 h and decreases below one afterwards. Furthermore, during ethanol uptake, the DO value decreased to values around 5%. The pilot process is assumed to be under fully aerobic conditions.
The feeding rate of this process showed a quasi-exponential increase in the initial 9 hours reaching 0.85 L min^{-1}, afterwards the feeding rate was gradually decreased to 0.6 L min^{-1}. The maximum volume recorded in the pilot reactor at the end of the process was approx. 1.35 m^3.

9.2 Laboratory-scale experiments

The overall aim of the laboratory-scale experiments was to improve the understanding of effects caused by large-scale conditions and of their influence on the cultivation process. The laboratory experiments were performed with the same industrial yeast extract medium as the pilot reactor process. The necessary nitrogen was provided by an industrial yeast extract medium and the carbon source with a sucrose solution. On both scales, the same feed to volume ratio was applied. An equal scale-down approach with a constant feed to volume ratio was also used by Bylund et. al (Bylund et al. 1998).
The sample volume on the laboratory scale was 10 mL with 1L initial volume; the sample volume in pilot scale was 100 mL with 1 m^3 initial volume. This different sample to reactor volume ratio between scales was considered for the calculation of the feeding rates in the small-scale experiments (except for experiment F1).
Five experiments were conducted in single bioreactors (three experiments in a 2 L Biostat B and two experiments in a 1 L MDX reactor) and two further experiments in the physical scale-down model (SDM) consisting of two connected stirred tank reactors.
The feeding strategy in the pilot reactor process was designed with a low sucrose feed rate to prevent ethanol production and inhibition. When the feeding rates are low, process relevant substrate gradients might occur in the pilot reactor that might influence the performance. For this reason, heterogeneous substrate concentrations were researched in the SDM.
The five experiments in ideally mixed reactors were performed to determine whether the process performance in the pilot reactor was influenced by heterogeneous conditions. For the first experiment in the Biostat B bioreactor system (F1) the different sample to volume ratio between the scales was not considered. This was corrected for the second (F2) and all subsequent experiments. The third experiment (F3) featured a decrement step of the glucose feed during the final process phase. Since the DO concentrations were consistently low during these experiments, two further experiments were conducted in the

MDX bioreactor (F4 & F5). This bioreactor system has a different air dispenser which creates smaller air bubbles resulting in a higher surface to volume ratio, allowing for a better oxygen transfer into the liquid phase. This should lead to higher DO concentrations.
The mixing time in the pilot reactor process was calculated according to the method described in chapter 4.3.3.2. The resulting theoretical mixing time had to be calculated mathematically using dimensionless numbers and was approximated to be 71 s (detailed calculation in appendix, chapter 12.2.3). This calculated mixing time is realistic for a 2 m³ bioreactor (Delvigne 2006). The mixing time in the SDM experiments was adapted to resemble the calculated mixing time and was set to about 70 s for the next experiments.
The volume ratio (V_{R1} to V_{R2}) in the two SDM experiments was kept at 1:1 at the start of the process. In these experiments (F6 & F7) the primarily researched heterogeneity were heterogeneous substrate concentrations. Additional data for cultivations with heterogenous conditions were acquired for the identification of the model parameters. These cultivations were also performed to check if heterogenous conditions might reduce ethanol production in the SDM. During the final phase of experiment F7 performed in the SDM, the RQ was measured in both reactors of the SDM to experimentally determine if ethanol is consumed and produced simultaneously in the cultivation.
These preliminary experiments with the new media were also necessary to practise the handling on the small scale (for the NMPC experiment) and for the process transfer to a different laboratory.

9.2.1 Single bioreactor experiments (F1 – F5)

Five experiments were conducted in single lab-scale bioreactor systems. The first three cultivations (F1 - F3) were performed in the Biostat B bioreactor with an initial volume of 1 L (1.1 L in F1).
F1 started with a feeding rate set to 0.30 mL min^{-1} and had a peak volumetric flow rate of 0.90 mL min^{-1} at 9 h. After 9 h the feeding rate was set to a constant value of 0.79 mL min^{-1}. After 1.75 h, the feeding rate was adapted to lower values until 3.5 h, to compensate for too high rates in the initial phase.
The feeding rates of processes F2 and F3 were starting at 0.18 mL min^{-1}. F2 was reaching 0.77 mL min^{-1} at 9 h, afterwards the feed rate was adjusted to a constant value of 0.70 mL min^{-1}. The rate of F3 reached 0.80 mL min^{-1} at 9 h, afterwards the rate of F3 was adjusted to 0.77 mL min^{-1} and further reduced to 0.64 mL min^{-1} at 12.5 h.
The feeding and process duration was extended to 21 h in F1 and F2. Nonetheless, identical total volumes were fed during processes F2 and F3 (530 mL) during the initial 17 h, which corresponds to the duration of the pilot reactor process. In F1 the total feed volume

was higher at about 600 mL in the first 17 h. The resulting experimental data are presented in Figure 69.

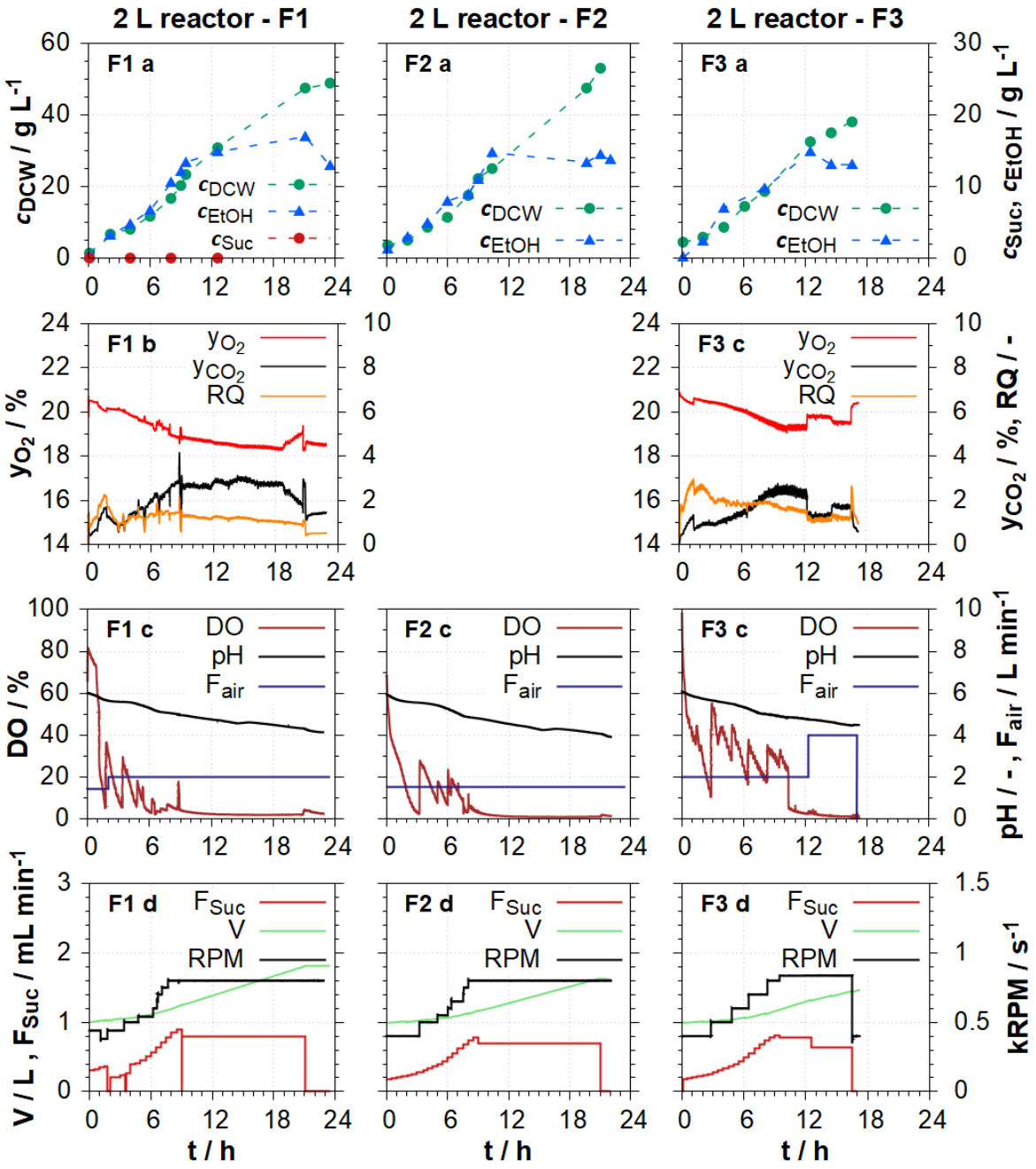


Figure 69: Experimental data of case study cultivations F1, F2 and F3 in Biostat B bioreactors 1st row: Experimental offline data for c_{Suc}, c_{DCW} and c_{EtOH}. 2nd row: Online data for exhaust gas composition (y_{O2} and y_{CO2}) and the RQ (no data for F2 available). 3rd row: Online data for the dissolved oxygen concentration, the pH value of both reactors as well as the aeration rate. 4th row: Feeding rate, stirrer speeds, as well as the calculated volumes.

Determinations of the sucrose concentration were carried out in the first cultivation F1; the sucrose concentration was consistently determined at approximately 0 g L^{-1} (below the level of determination). Therefore, it was assumed that the sucrose concentration was always 0 g L^{-1} in cultivations F1 – F5.

In the three cultivations (F1 – F3) displayed in Figure 69 biomass growth was quite similar, reaching a dry cell weight density of about 40 g L^{-1} after 18 h and above 50 g L^{-1} towards 24 h. The ethanol concentration increased to 15 g L^{-1} in F1 – F3 instead of the 6 g L^{-1} in the pilot scale process. After reaching the highest feeding rate at 9 h, the ethanol concentration was slightly increasing in F1. In F2 and F3 the ethanol concentration was slightly decreasing possibly due to the lower feeding rates.

The RQ in F1 and F3 was above 1 for nearly the whole cultivation. Only at the end of the cultivations at 18 h the RQs reached values close to one. The online data for the exhaust gas composition of cultivation F2 were not determined. The pH was steadily decreasing to pH 4 starting at pH 6 in all three cultivations.

The stirrer speed was adjusted multiple times in attempts to increase the DO. The DO still decreased to values closely above 0% at 10 h. This could potentially have influenced the process performance and might explain the higher ethanol concentrations compared to the pilot reactor process V1 where c_{EtOH} reached a maximum of 6 g L^{-1}. For this reason, two simulations were performed for process F2 with the parameter set identified in chapter 8.3.1. One simulation was carried out under fully aerobic conditions (DO = 30%) throughout the process, the other one with anaerobic conditions (DO = 0%) after 8 hours. The resulting simulated data are presented in Figure 70.

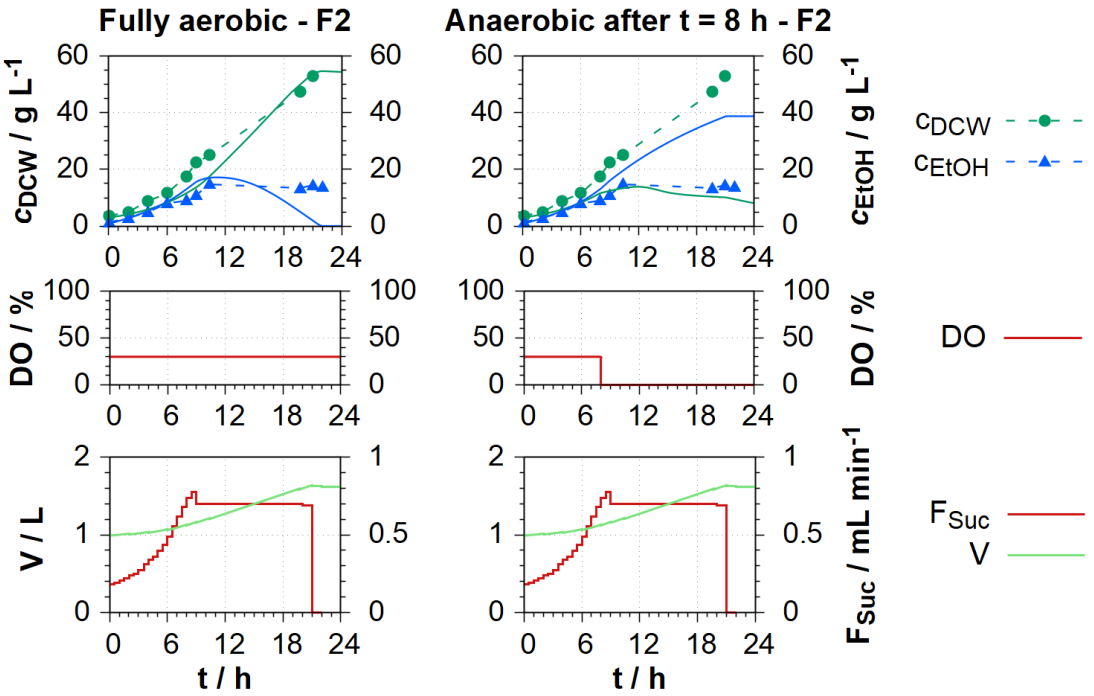


Figure 70: Comparison between two simulations of process F2. First simulation is under fully aerobic conditions (1st column); second simulation is anaerobic after 8 hours (2nd column).

The simulation of a fully aerobic process F2 is more similar to process F2 compared to the simulation of a process which is anaerobic after 8 h. The model simulates the biomass growth quite well; ethanol was decreasing after 10 h in the simulation but is constant in the process data. Based on the simulation, it can be concluded, that even though the DO measurement is close to 0%, the experiment is under aerobic conditions, however most likely under some oxygen limitation (microaerobic conditions).

Two experiments (F4 and F5) were performed in the MDX bioreactor with an initial volume of 0.5 L. This bioreactor system utilises a frit for the gas dispersion, which might be able to support a higher DO during cultivation. In F4 the same feeding strategy as F3 was applied. The feeding strategy of F5 with stepwise decreasing feeding rates as performed in the pilot reactor process. Figure 71 presents the experimental data for processes F4 and F5.

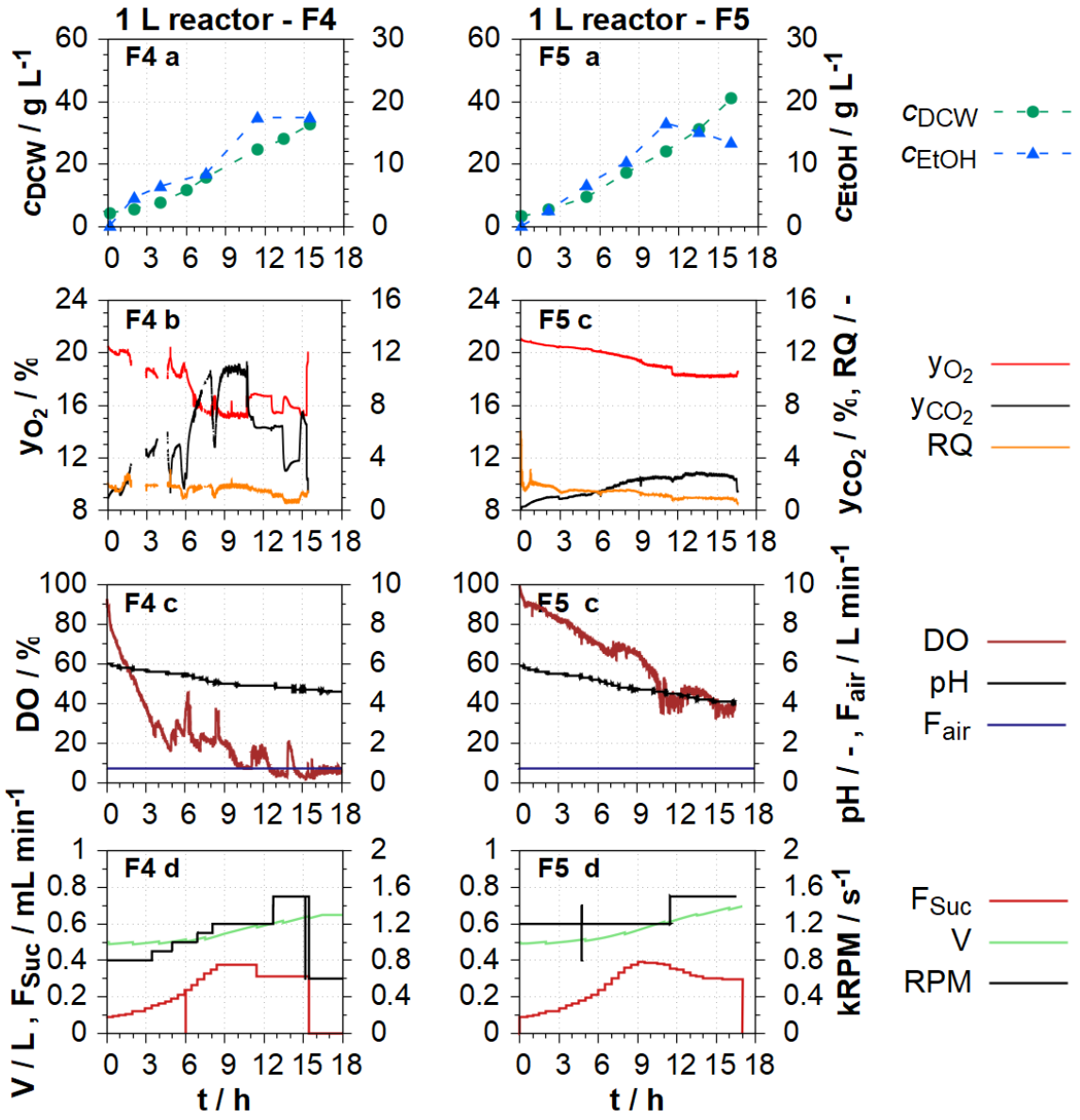


Figure 71: Experimental data of case study cultivations F4 and F5 in MDX bioreactors. 1st row: Experimental offline data for c_{DCW} and c_{EtOH}. 2nd row: Online data for the exhaust gas composition (y_{O2} and y_{CO2}) and the RQ. 3rd row: Online data for the dissolved oxygen concentration, the pH value of both reactors as well as the aeration rate. 4th row: Feeding rate, stirrer speeds, as well as the calculated volumes.

Cultivation F4 reached a final dry cell weight density of about 30 g L^{-1} which is approx. 10 g L^{-1} less than the final density of F5 (40 g L^{-1} at 16 h). Ethanol was produced to a maximum concentration of 18 g L^{-1} in F4, which is the highest concentration of the five processes in single bioreactors (F1 – F5). The ethanol concentration was steady in the samples after 11 h. The RQ was above 1 indicating an increase of the ethanol concentration until 13.5 h when the RQ decreased to values below 1 indicating ethanol consumption.

The aeration rate was set to be constant at 0.7 L min^{-1} in both cultivations. The DO in cultivation F4 was decreasing to 5% and in F5 the lowest measured DO was around 40%. The pH was decreasing to pH 4 in both cultivations similar to F1 – F3.
Even with the DO always above 30% in F5, ethanol was produced to a concentration of 16 g L^{-1} indicating the overflow metabolism being active. Ethanol was reduced in F5 after 11 h as indicated by the RQ being below 1. The exhaust gas data in process F4 exhibits gaps because the airflow required for the gas analyser was not always reached. This low air flow through the analyser is also reflected in the high y_{CO_2} and low y_{O_2} in the exhaust gas compared to the other four processes (F1 – F3, F5). Since the DO in F4 was always above 20%, it can be assumed that the reduced gas flow did not influence the cultivation. The determined offline data are therefore usable for the identification of the model parameters. Nevertheless, the experiment was repeated in the form of experiment F5.

Comparison of single bioreactor experiments (F1 – F5)

The pilot scale process V1 and the five processes F1 – F5 in single bioreactors are compared. This comparison is based on the time courses of the biomass densities and ethanol concentrations (Figure 72). The volumetric feeding rate for experiment F1 was about 10% higher. The sample volume to reactor volume ratio of the pilot scale (100 mL sample volume / 1 m³ medium volume) was assumed to be identical to the laboratory scale experiments for the first experiment (10 mL sample volume / 1 L medium volume) resulting in a relatively higher feeding rate per medium volume on the laboratory scale (Figure 72).

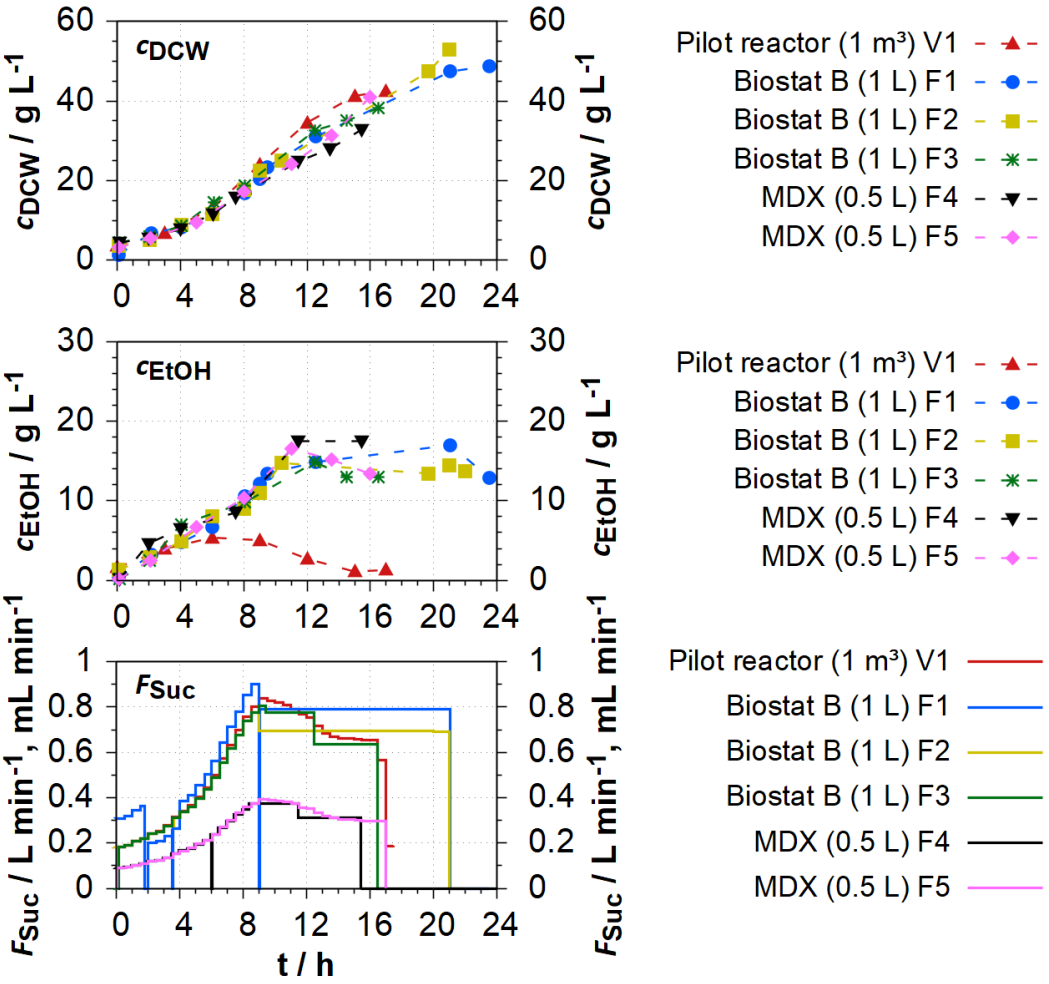


Figure 72: Comparison between the cultivation in the pilot reactor V1 (red), and the experiments in the single bioreactors (F1 – F5) for c_{DCW}, c_{EtOH} and feeding rates.

The cultivations V1, F2 and F5 in single reactor systems exhibited similar biomass production reaching about 40 g L^{-1} after 16 h. Process F4 reached a maximum biomass density of 32 g L^{-1} in 15 h. Between 10 h and 16 h the biomass density was higher in the pilot reactor process as compared to F1 – F5. At 12 h the difference between the pilot reactor and the lab-scale experiment was about 10% to 25%. However, the growth in the laboratory scale reactor was not slower compared to the pilot reactor process V1.

For the first four hours, c_{EtOH} was similar in all processes. Afterwards, the ethanol production rate in the pilot reactor decreased and c_{EtOH} reached its peak after 6 h, which is four hours earlier than in the single bioreactor experiments (F1 – F5). The maximum ethanol concentration is 6 g L^{-1} in the pilot reactor and 15 g L^{-1} to 18 g L^{-1} in the single laboratory scale reactors. The ethanol concentration decreased in the pilot scale reactor after 6 h, whereas it remained almost constant in the laboratory-scale reactors, after reaching their maxima at about 10 - 12 hrs.

The cultivations exhibit comparable biomass productions but diverge clearly for the ethanol production. This difference might be caused by heterogeneous sucrose concentration within the pilot scale reactor. Consequently, experiments were conducted in the physical scale-down model with heterogenous substrate concentrations induced by feeding into one of the reactors (F6 & F7). Both reactors were aerated and thus aerobic. Moreover, experiments were conducted with the objective to obtain experimental data from non-ideally mixed processes for the identification of the model parameters. A model-assisted investigation to identify a potential explanation or cause of these differences between the pilot plant and laboratory bioreactor is presented in chapter 9.6.

9.2.2 Scale-down model experiments (F6 & F7)

Up to this point, experiments have only been carried out under ideally mixed conditions, using the feeding strategy and conditions of the industrial cultivation. In order to carry out model-based investigations, data from cultivations under controlled heterogenous conditions are required for the parameter identification of the model, as the industrial pilot reactor experiment is probably not ideally mixed.

The heterogeneous conditions of the substrate concentration in the pilot reactor might cause simultaneous ethanol production and uptake, which could result in lower ethanol concentrations. In systems with higher mixing times, regions in the reactor may be below the sugar concentrations that enable ethanol uptake, while other regions may exhibit concentrations high enough for ethanol production, as indicated in Figure 32. This was demonstrated during the simulation study in chapter 6.2, where different mixing times were compared and it was also indicated by other research groups (Sweere and Matla et al. 1988). These conditions would also be conceivable for the process in the pilot reactor.

For this reason, two experiments were performed in the scale-down model to test the hypothesis of concurrent ethanol uptake and production. Each reactor had an initial volume of 0.5 L with a 1:1 volume ratio (V_{R1} to V_{R2}) between the reactors. A similar volume ratio was chosen by other research groups. Oosterhuis et al. (Oosterhuis et al. 1985; Oosterhuis 1984) used a 1.6 L reactor for the reactor with the feeding inlet and a 0.6 L reactor without feeding. Lara et al. und Sandoval-Basurto et al. used a ratio of 2:1 between the two reactors (Sandoval-Basurto et al. 2005; Lara et al. 2016).

Both processes utilised the feeding strategy which was used in process F5 with a stepwise decreasing feed after the peak feeding rate. The exhaust gas from both reactors were combined and determined together in the gas analyser.

The mixing time of the pilot reactor was roughly estimated at about 71 s (chapter 12.2.3). The mixing time was set to about 70 s in F6 and to 73 s in F7. Cultivation F6 was extended

to a total cultivation duration of 26 h. Figure 73 presents the experimental data of these two processes.

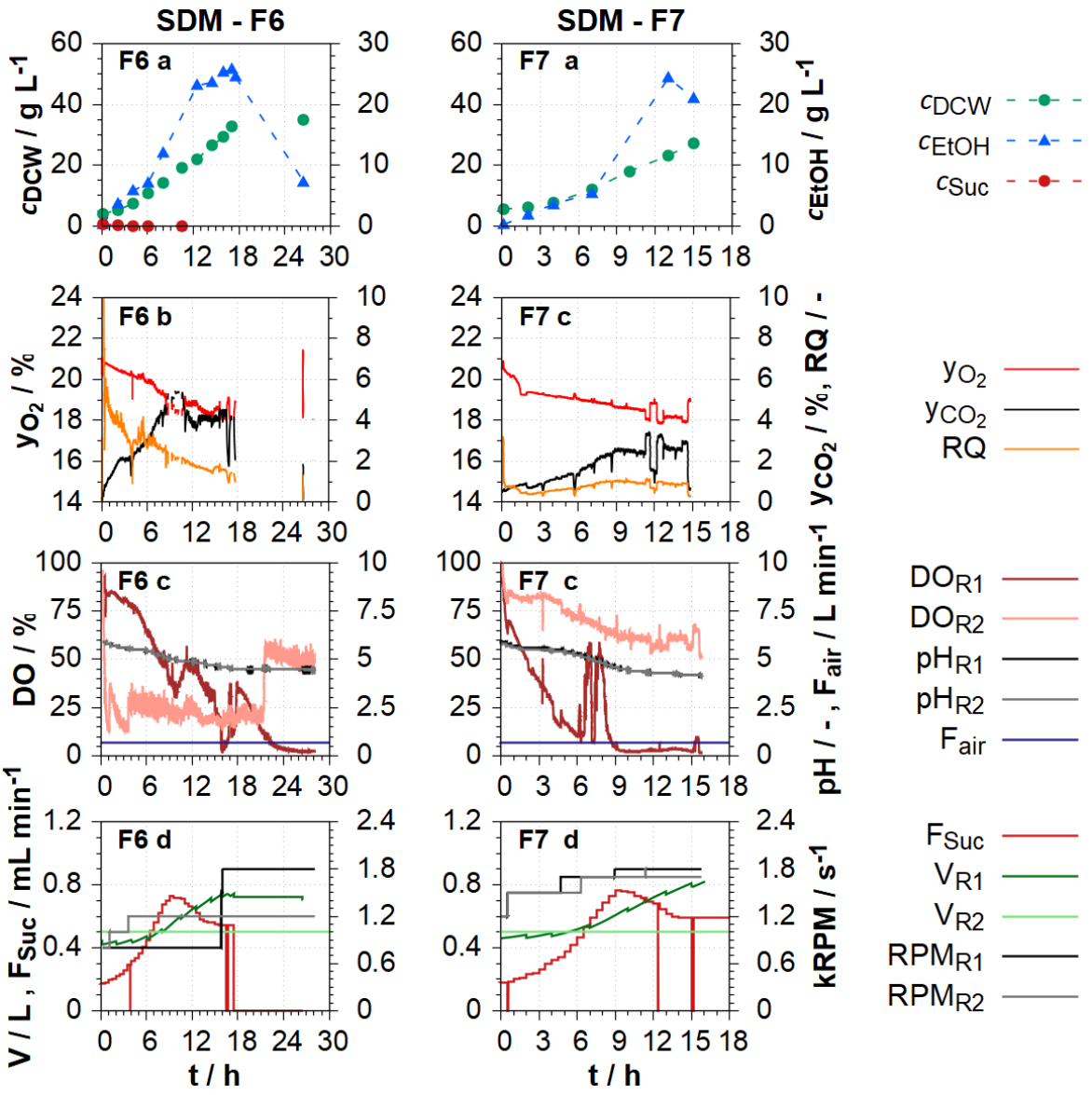


Figure 73: Experimental data of F6 and F7 in the scale-down model. 1st row: Experimental offline data for c_{Suc}, c_{DCW} and c_{EtOH}. 2nd row: Online data for the exhaust gas composition (y_{O2} and y_{CO2}) and RQ. 3rd row: Online data for the dissolved oxygen concentration, the pH value of both reactors as well as the aeration rate. 4th row: Feeding rate, stirrer speeds, as well as the calculated volumes in both reactors.

In both experiments the gas input nozzle of one reactor (R1) was partly blocked and less air was let through into the system. Consequently, feeding was performed into the reactor with the blocked nozzle since overflow metabolism was expected in the reactor with the feeding inlet. The resulting experimental data can be utilised despite the blocked nozzle since ethanol production takes place at high sugar concentrations and under anaerobic

conditions (if not enough oxygen is supplied to R1). However, oxygen is supplied through the blocked nozzle, probably resulting in microaerobic conditions.

In F6, a biomass density of 35 g L^{-1} was reached after 26 h, while in F7 27 g L^{-1} were produced in 15 h. In both processes the peak ethanol concentrations were about 25 g L^{-1}. Ethanol was subsequently consumed in both processes, which was also reflected in the RQ. The RQ decreased to values around 1 in F6 after 18 h and below 1 in F7 at 12 h. At the end of the process, an ethanol concentration of 7 g L^{-1} remained in F6.

In F6 the DO was always above 10% in both reactors except during the final process phase. In F7 the DO was at 60% in reactor 2 and decreased to under 10% in reactor 1 where the feed was located. The DO spikes for reactor 1 of cultivation F7 between 6 h and 9 h were probably caused by gas bubbles in front of the sensor, which were seen by visual inspection. The pH value decreased to pH 4 over the course of the processes.

During feeding, in F7 from 11 h until 14 h, the exhaust gas composition was determined in individual reactors of the SDM. Figure 74 presents the exhaust gas composition and the RQ for both reactors together and for both individually.

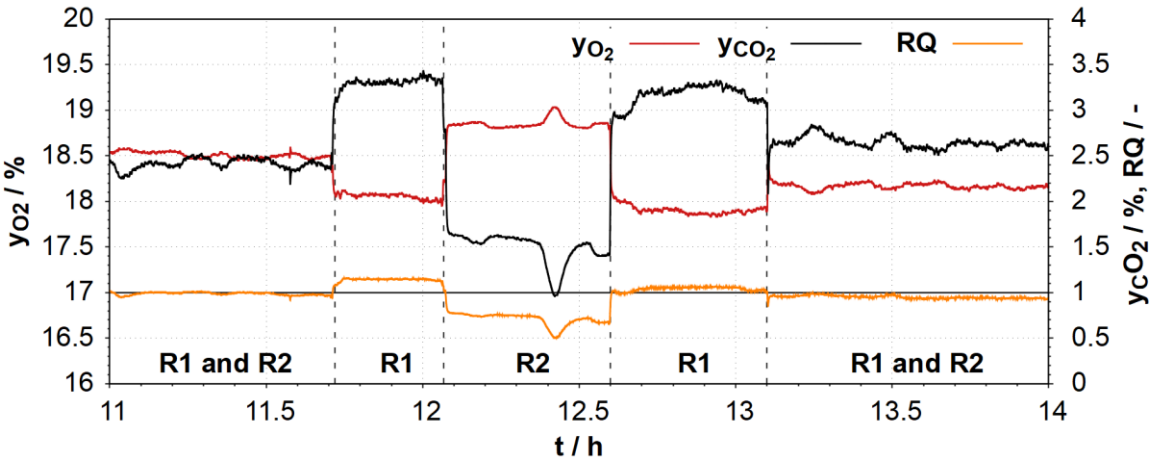


Figure 74: Exhaust gas composition for y_{O2} and y_{CO2} as well as the RQ for experiment F7 between 11 h and 14 h. Measurements are for both reactors together and each reactor (R1 and R2) individually.

The RQ in the reactor where the feed was located (reactor 1, R1) was slightly above 1. This might have been caused by the overflow metabolism of sucrose to ethanol, which mainly takes place in the reactor with the feed. In R2, the RQ is clearly below 1 indicating that ethanol might be consumed instead of being formed. Consequently, it is possible that ethanol production and consumptions can occur simultaneously in the SDM. This was already experimentally shown by Sweere et. al (Sweere and Matla et al. 1988) and experimentally indicated in this work in chapter 8.3.2.2 with different yield coefficients, and now experimentally shown for a *S. cerevisiae* cultivation with conditions and a medium

utilised in industrial cultivations. The physical scale-down system is therefore experimentally capable of simultaneous ethanol production and consumption. It is also possible that this effect could have occurred in the pilot reactor.

The two experiments performed in the SDM have demonstrated that the cultivation conditions in the SDM must be planned (and carried out) precisely and carefully for an experimental reproduction of the pilot scale data. Consequently, model-based (in-silico) experiments were carried out (see chapter 9.6). Given that only lab-scale experiments have been conducted thus far that consistently yielded more ethanol than the experiment on the pilot scale, a further experiment was carried out with the goal of a reduced ethanol production.

9.3 NMPC experiment in the scale-down model (F8)

The available data of cultivations with the industrial media are limited in their variety, as a similar feeding strategy has been used in cultivations F1 – F7. All previous cultivations on the laboratory scale also produced substantially more ethanol than on the pilot scale. Therefore, the aim of the next cultivation was to obtain data with lower ethanol concentrations. These ethanol concentrations should be as close as possible to the concentration of the pilot reactor process. The process will be designed using model based methods.

The first option for the design of the cultivation is the application of the modified mDoE. As mDoE calculates fixed feeding profiles, potential experiments would be simulated in advance. The most promising experiment would be carried out. Since all experiments available for the parameter identification utilised similar feeding strategies, the identified initial parameter set is expected to be contain limited informative value even if the parameters are identified multiple times. For this reason, the possibility of missing the target cannot be ruled out using the mDoE method. This would most likely result in a second run of the mDoE and another cultivation experiment.

NMPC's have been proposed as a potential powerful tool in process development for the model-based design of operational strategies (e.g. feeding strategies) (Witte 1996; Frahm and Hass et al. 2003; Li 2015). In the presented work, an NMPC of the OLFO-type (Witte 1996; Schneider et al. 1993) shall be utilised for the development of a feeding strategy leading to an ethanol concentration profile in the SDM which is as close as possible to the ethanol concentration profile in the pilot scale reactor (see also chapter 2.6.1 for a description of the NMPC).

For the first time, this application of an NMPC in combination with the SDM would also test the capabilities of an NMPC to support the model-assisted scale-down of operational

strategies with the aim of reproducing the larger scale data. Overall, the NMPC application reduces the risk of additional experiments (Witte 1996) and, ideally, only one experiment has to be carried out for these objectives instead of multiple experiments.
Additionally, the model and parameter set from the NMPC application is identified with the experimental data of all experiments (except F5 & F6) (chapter 9.3.2 for detailed information). It was subsequently used for model-assisted investigations of the scale-dependent effects (chapter 9.6).

9.3.1 Definition of the objective

NMPC's require an optimisation criterion which is to be minimised or maximised. Here, the optimisation criterion was designed to reduce ethanol production while encouraging biomass growth since the main difference between the cultivations in different scales is the increased production of ethanol on the lab-scale. The criterion is therefore a combination of the maximisation of the final biomass density and the minimisation of the maximum ethanol concentration.

$$K_{opt} = \begin{cases} 100 - c_{DCW,final} & when \quad c_{EtOH,max} < 5\ g\ L^{-1} \\ 100 - \left(c_{DCW,final} - 5 \cdot (c_{EtOH,max} - 5)\right) & when \quad c_{EtOH,max} \geq 5\ g\ L^{-1} \end{cases} \tag{9.1}$$

The first equation is active when the ethanol concentration is below 5 g L^{-1} which was about the maximum ethanol concentration of the pilot-scale process. The second equation is active when $c_{EtOH,max}$ exceeds 5 g L^{-1} and is acting as a penalty function, which increases the optimisation criterium K_{opt} when active. The optimisation part of the NMPC is designed to minimise K_{opt}. This should limit the ethanol production to a maximum of around 5 g L^{-1} while still encouraging biomass growth.

9.3.2 Parameter identification

The NMPC started with an initial parameter set which was identified before the start of the optimisation experiment. The parameter identification utilised the data of five experiments performed on the laboratory scale (F1 – F4 and F7).
Cultivations F5 and F6 were not considered for parameterisation to save computation time. The data of F5 and F6 were instead used to test the accuracy of simulations and validate the identified parameter set. A fast parameter identification during the NMPC experiment is crucial. Therefore, a mathematical model for the SDM only comprising of two reactor submodels was utilised, neglecting the piping between the reactors (see chapter 6.3.2).
During the NMPC experiment the parameter identification and the control function optimisation were performed five times (chapter 9.3.3). 25 parameters were identified (listed in Table 24 in the Appendix), mainly the same parameters as in chapter 8.3.1 without

parameters describing anaerobic conditions as the experiments of this study are performed under aerobic conditions. The initial parameter set is the resulting parameter set from chapter 8.3.1. The resulting comparisons between the offline and simulated data after parameter identification are presented in Figure 75.

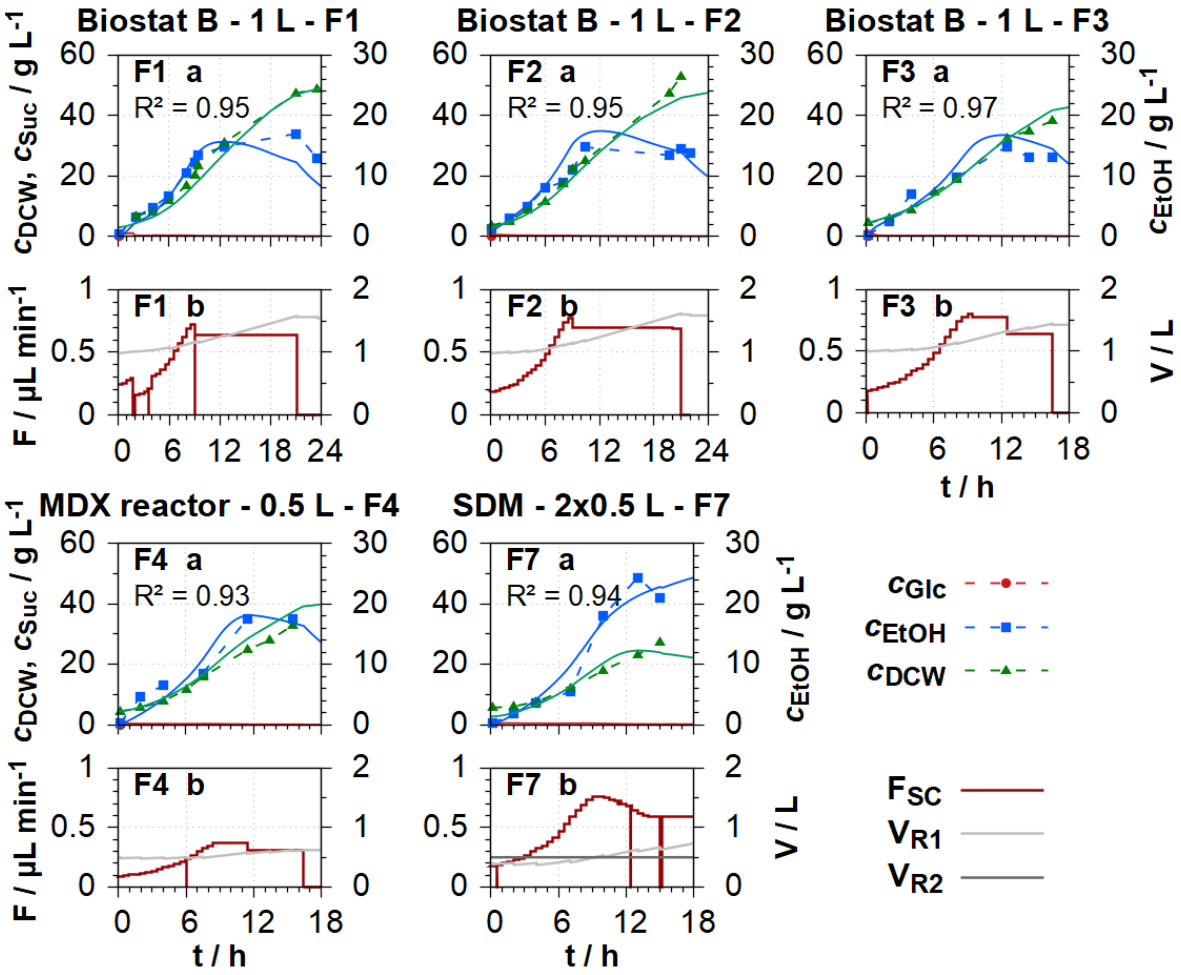


Figure 75: Result of the parameter identification before the application of the NMPC (OLFO-algorithm) for processes F1 – F4 and F7.

In the cultivations F1 and F2 a higher c_{DCW} is reached compared to cultivations F3, F4 and F7, this is also reflected in the simulations of these processes. The SDM cultivation F7 has a higher final ethanol concentration, and less biomass growth compared to the single bioreactor experiments. This is also achieved by the model simulations after the parameter identification. Overall, the simulated data are representing the experimental data also qualitatively well and the model is able to describe differences between the processes.
The resulting parameter set is validated with the experimental data of cultivations F5 and F6 (Figure 76).

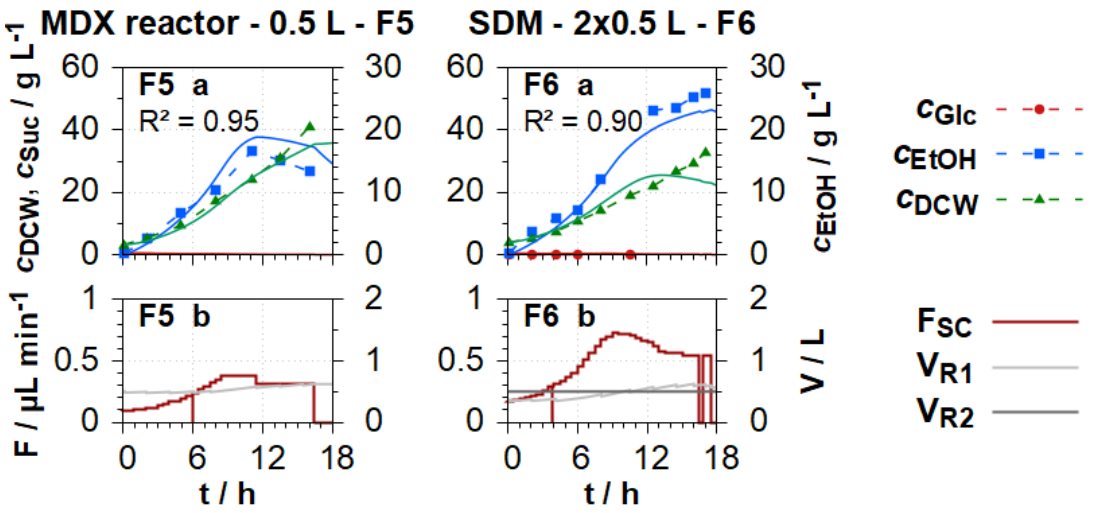
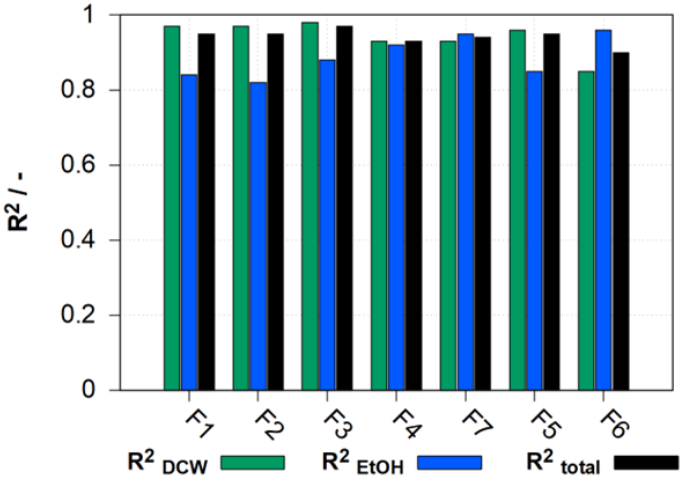


Figure 76: Precision test of simulation with the identified parameter set compared to the experimental data of processes F5 and F6.

In comparison to the experimental data, the final biomass density is slightly lower in both simulations, and the growth rate seems to be decreasing at the end. Ethanol concentration is slightly higher in the simulation of F5 and lower in F6. The experimental data are described well by the model, even for processes that were not utilised for the parameter identification.

Figure 77 presents the resulting R^2 for c_{DCW}, c_{EtOH} and for both concentrations together for the chosen five processes for parameter identification and the two remaining processes used for checking the parameter identification.



Experiment	R^2		
	c_{DCW}	c_{EtOH}	total
F1	0.97	0.84	0.95
F2	0.97	0.82	0.95
F3	0.98	0.88	0.97
F4	0.93	0.92	0.93
F7	0.93	0.95	0.94
F5	0.96	0.85	0.95
F6	0.85	0.96	0.90

Figure 77: Coefficients of determination for the five processes utilised for the parameter identification (F1 – F4, F7) and the remaining two processes (F5 & F6) before the application of the NMPC (OLFO-algorithm).

The resulting R^2 are all above 0.8 for c_{EtOH} and even above 0.9 for c_{DCW} considering the cultivations used for identification. The combined total R^2 is above 0.9 for all seven

processes, even for the validation processes F5 and F6. This indicates that the parameter identification was not only qualitatively but also quantitively successful, and the model parameter set may be applied in simulations for cultivations not used for the parameter identification. The obtained parameter set may be utilised as the initial parameter set for the application of the NMPC in the next chapter.

9.3.3 Results of the NMPC experiment

The parameter set identified in chapter 9.3.2 was utilised as the initial parameter set. The utilised control function was a linear increasing feeding rate. The optimisation factors of this linear function $F(t)$ are the initial (F_{start}) at the beginning of the cultivation (t_{start} = 0) and final volumetric rate (F_{final}) of the feed F at the end of the process (t_{final}).

$$F(t) = \frac{(F_{final} - F_{start})}{t_{final}} \cdot t + F_{start} \tag{9.2}$$

This function with only two optimisation factors increases the performance of the NMPC compared to previous versions, in which control values were determined for time periods.
The feed had a sucrose concentration of 216 g L^{-1}, which was consistent with the other experiments in this case-study. The planned experimental duration was 20 h. The optimisation horizon for the model-based optimisation was also set to 20 h. The feed was planned to be turned off after 14 hours, and 6 hours of ethanol reduction were planned, to get additional data for the ethanol uptake phase.
Figure 78 presents the experimental offline and online data for the cultivation with the NMPC in comparison to the resulting simulation of the last optimisation step.

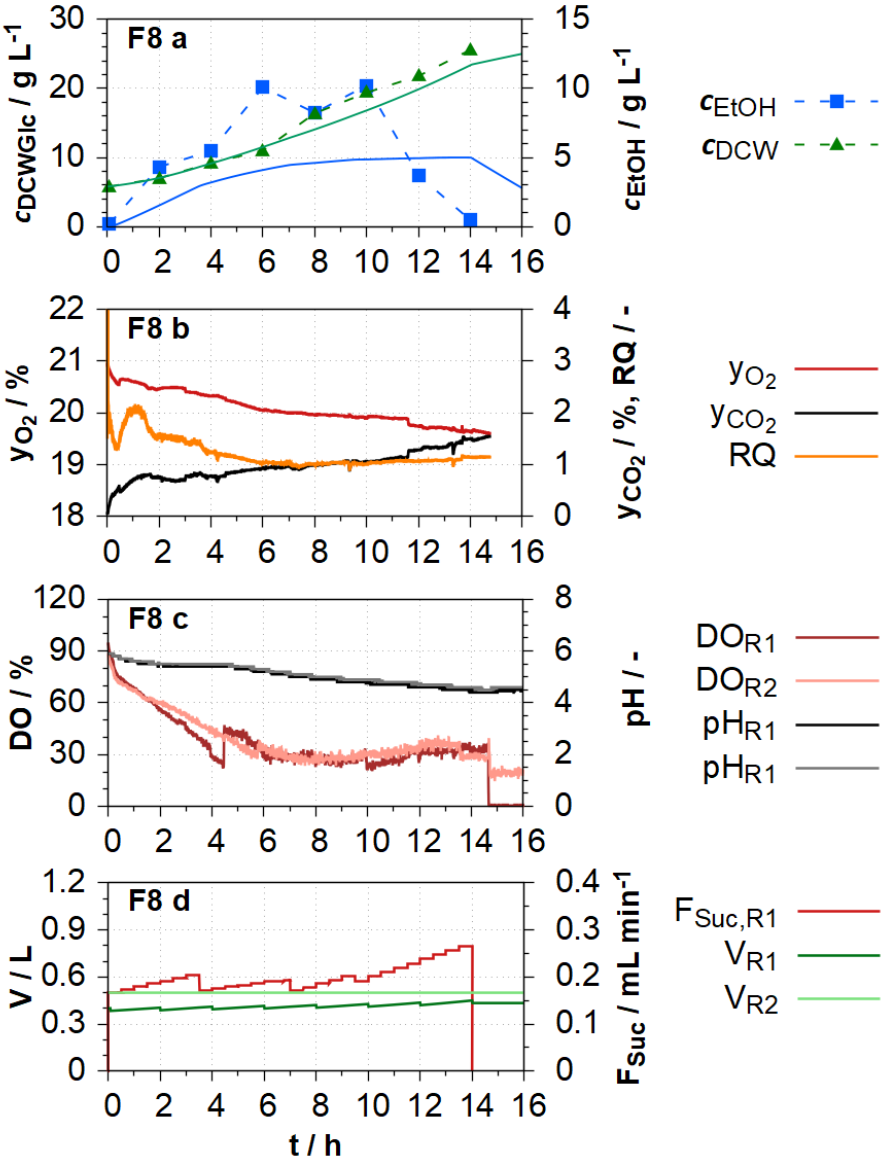


Figure 78: NMPC cultivation F8 (final optimisation step) in the scale-down model in comparison with the final simulation. 1st row: Experimental and simulated data for c_{DCW} and c_{EtOH}. 2nd row: Online data for the exhaust gas composition (y_{O2} and y_{CO2}) and the RQ. 3rd row: Online data for the dissolved oxygen concentration and pH. 4th row: Optimised feeding rate and the calculated volumes of both reactors.

At a process time of 14 h the ethanol concentration reached 0 g L^{-1}. This made the planned ethanol uptake phase from 14 h to 20 h obsolete. For this reason, the final cultivation duration was reduced to 14 hours.

At the end of the process (t = 14 h) a dry cell weight density of 26 g L^{-1} was reached. Peak ethanol concentration was only about 10 g L^{-1} which was reached between 6 h and 10 h. The respiratory quotient is about 1 after 6 h until the end of the process. The exhaust gas fractions of O_2 and CO_2 increased, reflecting the increasing feeding rate. The DO

remained at 20% in both reactors after the initial decline. The pH was decreasing to 4.5 until the end, which was slightly higher than in experiments F1 – F7.

A total of five parameterisation and optimisation steps were performed at 0 h, 3.5 h, 7 h, 9 h and 11 h. The historic data (of experiments F1 – F4 and F7) was weighted with a factor of 0.5 for all parameter identification during the NMPC experiment. The last optimisation at 11 h did not change the feeding rate calculated at 9 h. The first four optimisation steps are presented in Figure 79.

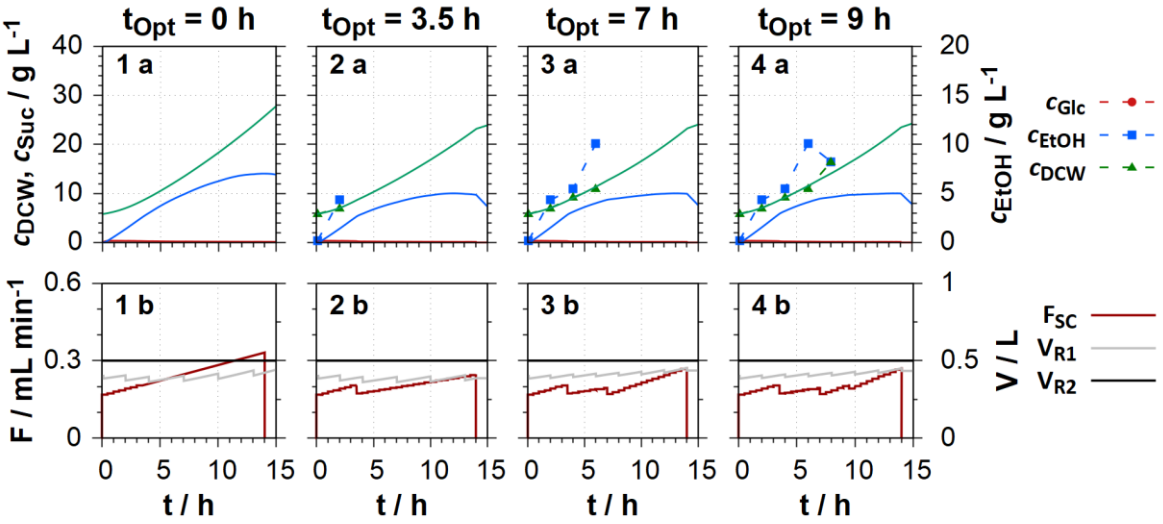


Figure 79: Four (of five) optimisation steps of the NMPC process F8 in the SDM. 1st row: Experimental and simulated data for c_{DCW} and c_{EtOH}. 2nd row: Optimised / calculated feeding rate and the resulting volumes of both reactors.

During the optimisation, the algorithm tried to limit the ethanol concentration to values below 5 g L^{-1} to stay below the activation limit of the penalty term. In this respect, the experiment was successful and only a maximum ethanol concentration of 10 g L^{-1} was produced, which is half of the previous laboratory scale experiments. This was achieved by decreasing the initial feeding rate after each optimisation step, resulting in the apparent “saw pattern” (Figure 79). The reason for this is, that the experimentally determined ethanol concentration was always higher than the simulated concentration. The new parameter set, which was identified prior to the control function optimisation, was therefore tuned to simulate a higher ethanol production in each step.

The mathematical model describes the time course of the biomass density quite well (R^2 = 0.93). However, the experimental ethanol concentrations of F8 were consistently clearly above the simulated values (see Figure 78). This difference might be caused by the low variety in the data in the experiments utilised for the parameter identification. The experimental data from five cultivations (F1 – F4 & F7) were used along with the NMPC

cultivation data for the parameter identifications, which all employed similar feeding strategies. This resulted in a parameter set which is adapted to these historic processes. However, the NMPC cultivation only had a limited influence on the parameter identification since the effect of simultaneous sucrose feeding and ethanol consumption caused by the lower feeding rate is not in the data of the previous experiments. One possible solution would be to increase the weighting of the NMPC experiment which would in turn potentially decrease the quality of the identification for the previous cultivations (F1 – F7).

During the optimisation it is crucial to keep the period between the last sampling and the transfer of the optimised control trajectories as short as possible. The model parameters must be identified, and the process optimised, while the process is performed. Therefore, the number of iterations for the parameter identification had to be limited, which might contribute to the larger deviations between simulated and experimental data. In this case only one hour was planned for a combined parameter identification and the calculation (optimisation) of the feeding strategy.

The result of the NMPC experiment indicates that the effective carbon substrate concentration in the pilot scale reactor might have been lower as compared to the laboratory scale experiments. Most of the yeast in the pilot reactor might be exposed to lower concentrations of sucrose caused by heterogeneous conditions resulting in a reduced ethanol production. This will be tested with the mathematical model in chapter 9.6 after the comparison and discussion of the experimental results.

Overall, this novel application of an NMPC calculating control strategies for an experiment in a physical scale-down model tested the capabilities of an NMPC as a tool supporting the model-assisted scale-down of process conditions. Previously, NMPCs were used for the design, optimisation and control of operational strategies of cultivations in single bioreactors (Frahm et al. 2005; Witte 1996; Wang et al. 2016; Dewasme et al. 2013; Craven et al. 2014).

9.4 Comparison of experiments

The comparison of the pilot scale cultivation V1 and the scale-down cultivations F1 – F8 is mainly based on the biomass density and ethanol concentration (Figure 80). The pilot cultivation V1 (V_{ini} = 1 m³) and the NMPC cultivation F8 (V_{ini} = 0.5 L per reactor) are compared to selected cultivations; cultivation F2 in Biostat B (V_{ini} = 1 L), cultivation F5 in MDX reactor (V_{ini} = 0.5 L) and cultivation F6 in the SDM (V_{ini} = 0.5 L per reactor). One cultivation per reactor type (and cultivation type) was chosen for the comparison.

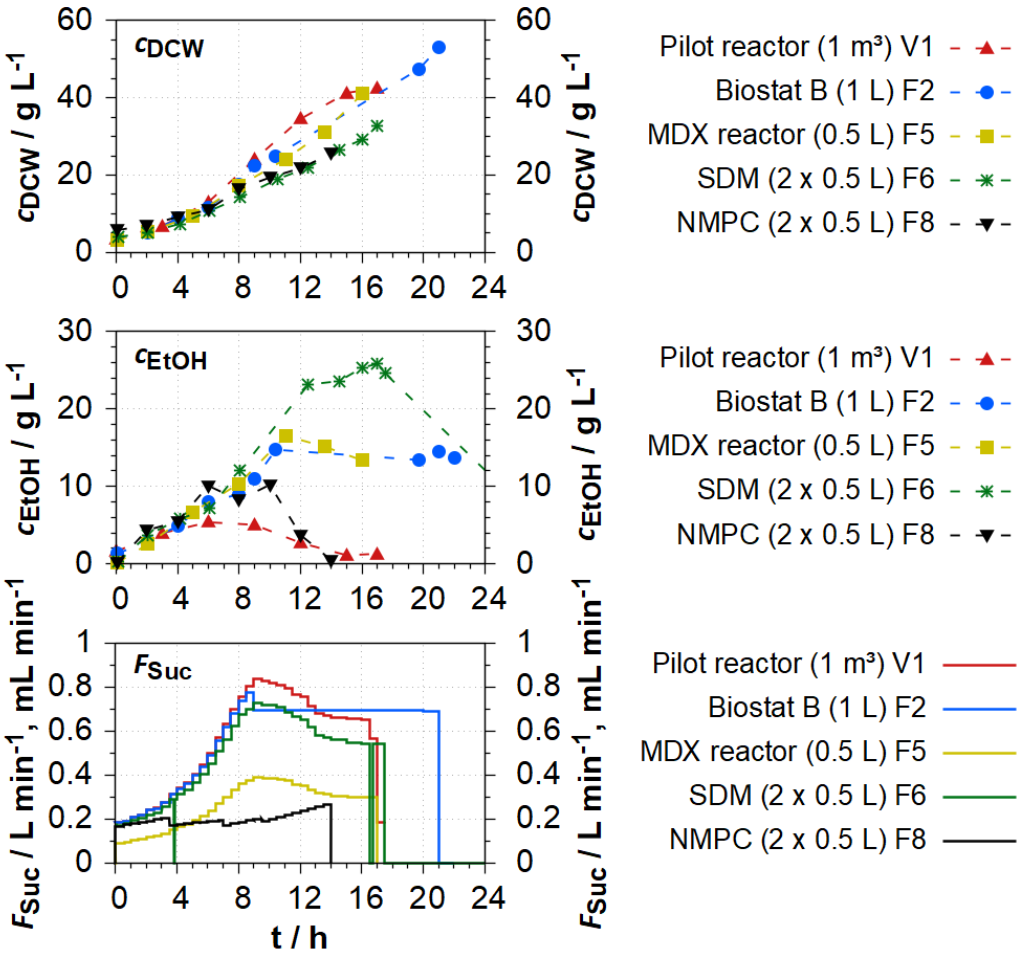


Figure 80: Comparison between experiments in pilot reactor V1 (red), Biostat B F2 (blue), MDX reactor F5 (yellow), SDM F6 (green), and NMPC (black) F8 for c_{DCW}, c_{EtOH}, and feeding rates.

The cultivations in single reactor systems (V1, F2, F5) exhibited similar biomass densities reaching about 40 g L^{-1} after 16 h. At the end of the NMPC process F8 (t = 14 h), a dry cell weight density of 26 g L^{-1} was reached, which is approximately 8 g L^{-1} less than in the single bioreactor processes (F1 - F5) and 12 g L^{-1} less than the pilot reactor process (V1) at the same time. Compared to the other process in the SDM F6, the biomass growth in F8 was about the same in the first 14 h, despite a lower feeding rate.

For the first four hours, c_{EtOH} was similar in all processes. Afterwards, the ethanol production rate in the pilot reactor decreased and c_{EtOH} reached its peak after 6 h, which is four hours earlier than in the single bioreactor experiments (F2 and F5) and six hours earlier compared to cultivation F6 in the SDM. The NMPC cultivation is most similar to V1 regarding c_{EtOH}. A peak ethanol concentration of 10 g L^{-1} (5 g L^{-1} in V1) was reached after 6 h. The ethanol concentration of 10 g L^{-1} is still relatively high, which can be attributed to the fact that the simulated ethanol concentrations were lower compared to the experimentally determined

concentrations. In both cultivations (V1 and F8) ethanol was decreasing after 10 h. Furthermore, the penalty function is only active at simulated ethanol concentrations above 5 g L^{-1}. The maximum ethanol concentration was three times higher in the laboratory scale processes F2 and F5 and even five-times higher in F6 compared to the pilot scale.
Lower biomass densities and higher ethanol concentrations for *S. cerevisiae* cultivations due to heterogeneous glucose concentration were also published in literature by George et. al for a STR-PFR and Sweere et. al for a STR-STR configuration. (George et al. 1993; Sweere and Matla et al. 1988)

9.5 Evaluation of the NMPC application

The NMPC employed a function for the calculation of a linear feeding strategy (see equation 9.2). This utilisation of a function represents a novel approach compared to the other versions of the OLFO algorithm in which control values were determined for multiple, specific time periods (Witte 1996; Frahm and Hass et al. 2003). This function reduced the optimisation factors to only two compared to previous applications of the OLFO-controller and thereby improved the performance of the NMPC (reduction of the computation time for optimisation). Consequently, the time between the last sampling and the transfer of the feeding strategy could be reduced, which is particularly advantageous for fairly fast cultivations, such as those of *S. cerevisiae*.
The aim of the following evaluation is to determine how much the NMPC application has improved the process in comparison to the previously performed cultivations (V1, F1 – F7). The evaluation is based on retrospectively calculated optimisation criteria which are exclusively used for this evaluation and on yield coefficients. Table 19 lists the criteria for all experiments in chapter 9. Furthermore, the yield coefficients for each process of the case study are presented.

Table 19: Retrospectively calculated optimisation criteria (K_{opt}) at closest sample time ($t_{K_{Opt}}$) and yield coefficients ($Y_{X/SC}$) for biomass from sucrose for all processes of the case-study

Process	Reactor	V	$Y_{X/SC}$	K_{opt}	$t_{K_{opt}}$
V1	Pilot	1 m³	0.51	58.8	15.0 h
F1	Biostat B	1 L	0.51	118.3	12.6 h
F2	Biostat B	1 L	0.54	109.2	14.0 h
F3	Biostat B	1 L	0.45	105.1	14.5 h
F4	MDX	0.5 L	0.38	145.0	15.5 h
F5	MDX	0.5 L	0.50	119.5	13.6 h
F6	SDM	2 x 0.5 L	0.37	166.5	14.5 h
F7	SDM	2 x 0.5 L	0.30	153.0	15.0 h
F8 (NMPC)	SDM	2 x 0.5 L	0.58	74.4	14.0 h

A lower K_{opt} corresponds to a higher final biomass density and lower final ethanol concentration. Ethanol concentrations below 5 g L^{-1} do not further reduce K_{opt} (equation (9.1)). K_{opt} was retrospectively calculated for each experiment at the sample time $t_{K_{Opt}}$ closest to the final time of cultivation F8 (14 h). For cultivation F2, the values for the calculation of K_{opt} were linearly interpolated between the two closest samples ($t_{K_{opt},1}$ = 11.6 h and $t_{K_{opt},2}$ = 17.6 h). The criterium is only used as a target value which should ideally be as close as possible to the value of the pilot reactor process V1 (K_{opt} = 58.8).

The best scale-down process without the NMPC yields a K_{opt} of 105.1, which is about 78.7% higher than V1. The higher K_{opt} is caused by the higher ethanol concentration in the laboratory scale experiment. The NMPC (process F8) was able to reduce this relative difference to just 26.5% (K_{opt} = 74.4) by reducing the final ethanol concentration. Ethanol is fully consumed in F8 until no ethanol is determined in the last sample compared to at least 12 g L^{-1} in the other scale-down processes. Since the minimisation of K_{opt} was the optimisation goal, the application of the NMPC can be evaluated as being successful.

The yield coefficients for the single reactor processes are similar to V1. The scale-down model processes without adaptions of the feeding rate produced the lowest yield coefficients. The yield coefficient $Y_{X/SC}$ for the NMPC process in the SDM is the highest with

0.58 out of all cultivations in this study with industrial media. The yield coefficient determined from F8 is 10% higher than the yield coefficients in the single reactor processes.
Furthermore, the objective of acquiring data on reduced ethanol production was successfully accomplished since the NMPC experiment in the physical SDM (F8) produced less than half the ethanol compared to previous experiments (F6 & F7).
Overall, the cultivation was brought closer to the pilot reactor cultivation (as evidenced by the low K_{opt}). The yield coefficient $Y_{P/SC}$ of the NMPC experiment was the highest of all experiments in this study (in chapter 9), which is caused by the low ethanol production during the process. Furthermore, the objective of acquiring data on reduced ethanol production was successfully accomplished since the NMPC experiment in the physical SDM (F8) produced less than half the ethanol compared to previous experiments (F6 & F7). The new data were already incorporated in the parameter set of the mathematical model during the NMPC application. With the parameterised model, model-assisted investigation of scale-dependent effects were performed (chapter 9.6). The application of the NMPC in a non-ideally mixed scale-down system together with the mathematical model was therefore successfully validated.
Further investigation into the behaviour of the NMPC should be carried out in the future. This investigation should quantify the influence of the selected optimisation criterion on the cultivations. Additionally, the requirements for the experimental data used for parameterisation of the mathematical model to successfully identify the parameters should be determined. In this context, it should be investigated how accurate the simulations with the identified parameter set must be in order to successfully apply the NMPC.

9.6 Model-assisted investigation of scale-dependent effects

The experiments on laboratory scale (F1 – F8) demonstrate different behaviour to the pilot scale reactor cultivation. In cultivations conducted in the single reactors (F1 – F5) with volumes between 0.5 L and 1.5 L, higher ethanol concentrations were determined compared to the pilot reactor cultivation (1 m³). These differences between the scales might be caused by non-ideally mixed conditions in the pilot reactor. The mixing times in the laboratory reactors are shorter (< 5 s; visually determined) than in the pilot scale (> 10 s, mathematically approximated). In the SDM with two reactors, longer mixing times lead to more distinct differences (see Figure 32) regarding the sucrose concentrations between the two reactors. The pilot reactor might feature areas around the feeding inlet with higher sucrose concentrations (leading to ethanol production) and areas with lower concentrations (without ethanol production and/or ethanol uptake). The resulting heterogeneous substrate

concentrations also depend on the volumes of the zones and their volume ratios (Sweere and Matla et al. 1988).

Figure 81 displays a comparison between the simulated and experimentally determined ethanol concentrations and biomass densities in the 1 m³ pilot-scale process (cultivation V1) as well as the simulated sucrose concentration. The model parameter values of the simulations are identical to the final parameter set of the NMPC application and were determined in chapter 9.3.3. The model parameter values for the limits of ethanol uptake (Diauxie) and production (Overflow metabolism, Crabtree-effect) are indicate in Figure 81 by horizontal lines.

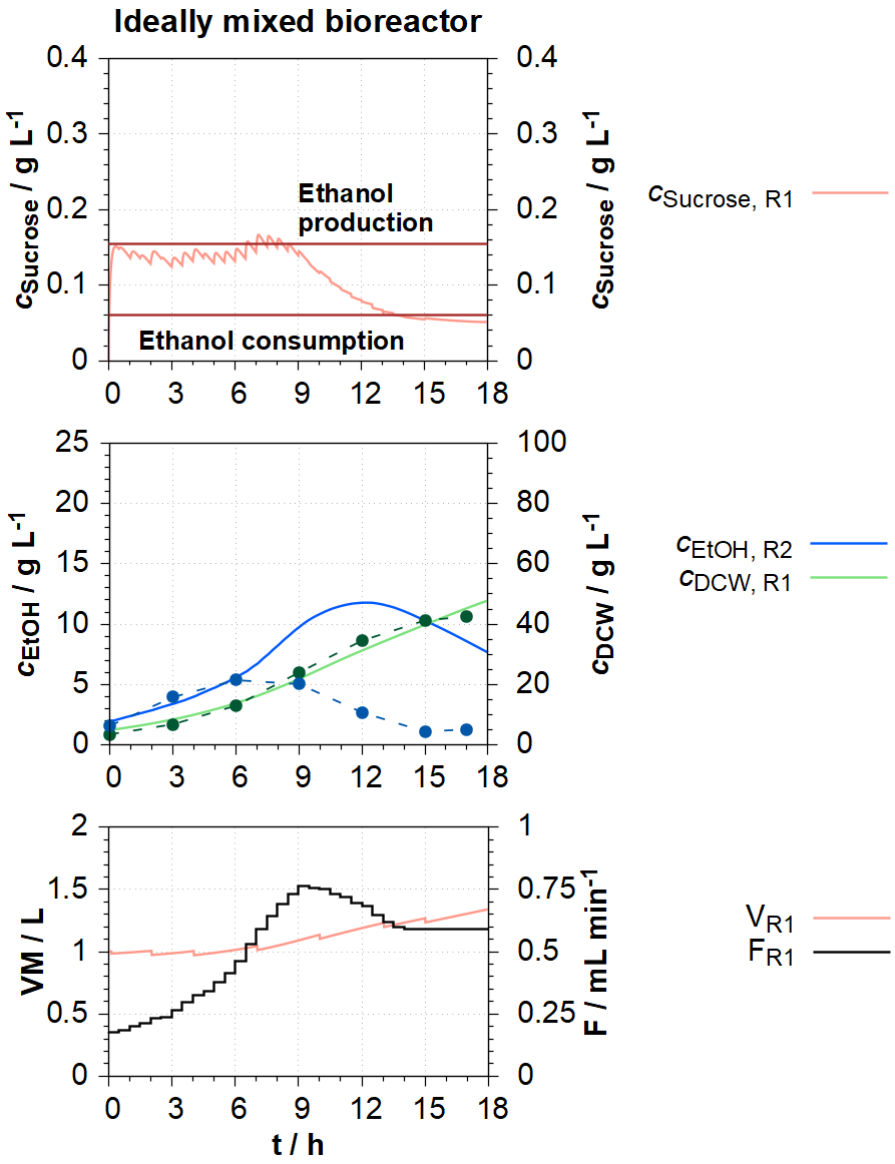


Figure 81: Simulation of the pilot reactor cultivation under ideal mixed conditions compared to experimental data of experiment V1. 1st row: Simulated data of the sucrose concentration; 2nd row: Offline data and simulated data for the biomass density and ethanol concentration; 3rd row: Calculated volume and feeding rate of the simulation.

The simulated sucrose concentration is very close to the boundary concentration of overflow metabolism (0.155 g L^{-1}) for the first 9 h of the experiment. After 9 h, the concentration decreases and from hour 12 h onwards it is close to the boundary concentration (0.06 g L^{-1}) at which ethanol is consumed. Accordingly, ethanol is consumed from 12 h onwards. Therefore, slight differences in the sucrose concentration might have major influences on the simulation and cultivation result.

The simulation shows that ethanol production and uptake are primarily dependent on the sugar concentration in the medium. In the simulation of the ideally mixed system, a higher maximum ethanol concentration is achieved (about 12 g L^{-1}) compared to the pilot reactor process V1 (maximum ethanol concentration of 5 g L^{-1}). This corresponds with the results of the experiments on the laboratory-scale, which feature a faster mixing time than the pilot reactor process.

The mathematical model shall be used in the next simulations to test whether the reduced ethanol concentrations on the pilot scale could be caused by non-ideally mixed conditions. The sensitivity study in chapter 6.2 demonstrated that less ethanol might be produced when the mixing time is longer (see Figure 31). This was not confirmed by simulations using the conditions of the pilot scale cultivation and the identified parameter set (Figure 83 in Appendix). Longer mixing times led to lower ethanol concentrations in the simulation with the identified parameter set (Figure 83). An increase of the mixing time led to higher ethanol concentrations in the laboratory scale experiments in this study (see Figure 80). This difference is caused by the sugar concentration that quickly fell below the limit for ethanol uptake in one of the two reactors. This led to the lower ethanol concentration in the sensitivity study (see Figure 31). However, with the process conditions and identified parameters of this study, the simulated sugar concentration decreased considerably later (see Figure 83).

This result is consistent with the results of the experiments on the lab-scale (F1 – F8). A lower maximum ethanol concentration was reached in the cultivations in the single bioreactor system compared to the cultivation in the SDM with mixing times of about 70 s. The mixing time by itself is therefore not sufficient to explain the lower ethanol concentration at the pilot scale.

Another possible explanation for the lower ethanol concentration on the pilot-scale might be the existence of poorly mixed areas that might occur around the position of the feed inlet. The volume of this zone would only take up a small part of the overall reactor. Increased sucrose concentrations in the zone around the feeding inlet could lead to ethanol production in this zone. Lower sucrose concentrations in the rest of the reactor could show no ethanol

production and/or even ethanol uptake. Considering these two zones and their ratio influences the heterogeneous sucrose concentration and thus the process result.

Three simulations were carried out with different volume ratios of the zones under fully aerobic conditions. Ratios of 330 mL/670 mL, 100 mL /900 mL and finally 50 mL /950 mL were simulated, the feed was always positioned in the reactor with the smaller volume.

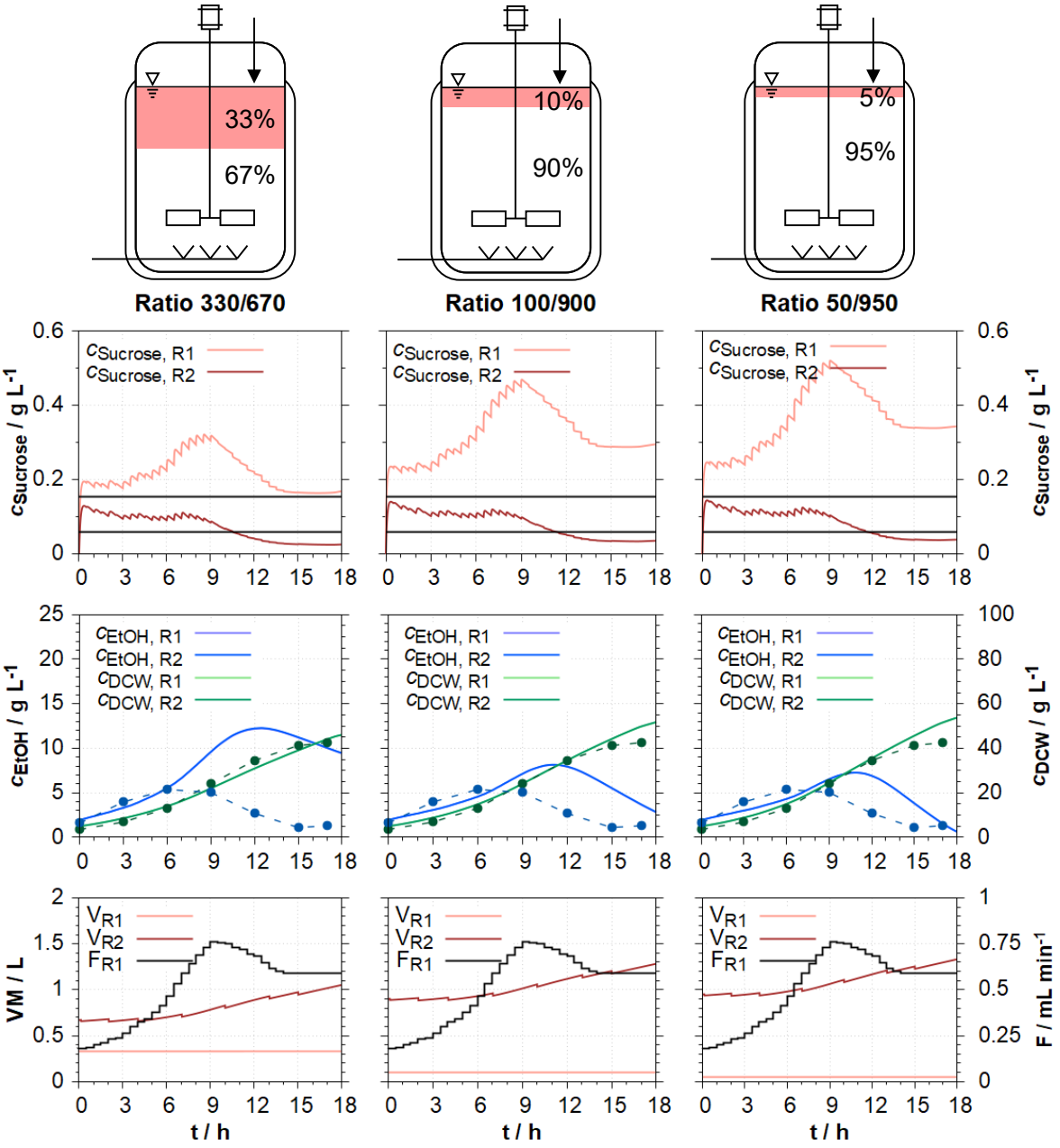


Figure 82: Three simulations at different volume ratios. 1st column: 330 mL/670 mL; 2nd column 100 mL/900 mL; 3rd column: 50 mL/950 mL (volume ratios indicated in red and white above the graphs). 1st row: Offline and simulated data for sucrose concentration; 2nd row: Experimental offline and simulated data for the biomass density and ethanol concentration; 3rd row: Calculated volume as well as the feeding rate of the simulation.

The lowest ethanol concentrations were obtained in the simulated cultivation with the 50 mL/950 mL ratio at approx. 7 g L^{-1}, which is only approx. 2 g L^{-1} above the determined concentration of the experimental pilot reactor cultivation. The biomass production also increased with increasing volume differences and corresponds well with the data from the pilot reactor in all three simulations.

The larger the volume difference between the zones, the greater the difference in the sucrose concentrations between the two zones. At 9 h, the sucrose concentration in reactor zone 1 compared to zone 2 is about 3 times higher for the 330 mL/670 mL ratio and 5 times higher for the 50 mL/950 mL ratio. Literature reports measured concentrations in the vicinity of the feed location that can be between twice as high (Larsson et al. 1996) and 400 times as high compared to the well-mixed zone around the stirrer (Bylund et al. 1998).

A smaller volume ratio of the feeding inlet zone leads to a lower maximum ethanol concentration. This is caused by the smaller volume in which ethanol production takes place. The sucrose concentration within the zone without feeding input is always below the overflow metabolism limit. For this reason, no or only tiny amounts of ethanol are produced in this zone, but ethanol is not consumed either. Ethanol consumption starts at 10.5 h when the feeding rate surpasses its maximum.

For these reasons, the existence of a relatively calm zone with a low exchange rate with the medium of the rest of the reactor in the area of the feed inlet may contribute to the differences between the pilot-scale and lab-scale cultivations.

9.7 Discussion and conclusion

Differences between *S. cerevisiae* cultivations in large scale reactors compared to those in laboratory-scale reactors can probably be attributed to the heterogeneous substrate concentrations and dissolved oxygen concentrations that might occur on the production scale (George et al. 1998; Larsson et al. 1996).

To investigate this issue, data from a 1m³ pilot reactor cultivation were compared to data from lab-scale cultivations (1 L and 2 L bioreactors) using the industrial medium and the same feeding rate to volume ratio. For the pilot reactor cultivation, a relatively low feeding rate was utilised that could have led to low sucrose concentrations and heterogenous conditions within the reactor. The influence of these heterogenous conditions was investigated utilising five experiments in ideally mixed tank reactors (F1 - F5) and two experiments in the SDM with heterogeneous sucrose concentrations (F6 & F7).

Biomass production was similar in the single bioreactor lab-scale experiments but reduced by about 20% for the experiments performed in the physical SDM as compared to the pilot-scale data. Ethanol concentrations were higher in all seven lab-scale experiments F1 to F7

as compared to the pilot reactor process V1. In the ideally mixed experiments about the same dry cell weight density was reached, but the SDM cultivation resulted in 20% less biomass as compared to V1. In experiment F7, exhaust gas measurements in the individual reactors of the SDM revealed that ethanol was simultaneously produced in the reactor with the feeding inlet and consumed in the other reactor. However, a net increase of ethanol concentrations was measured, thus indicating that the overall ethanol production rate exceeds the overall ethanol consumption rate.

Simulations with the mechanistic, mathematical model supported the assumption that cultivation F2 was under aerobic conditions, despite the DO measurement being close to 0%. This was concluded based on the simulated results being more similar to a simulation of an aerobic cultivation compared to the simulation of anaerobic cultivation. However, the ethanol uptake rate in the simulation was higher than observed in the experiment, which suggests that the cultivation was performed most likely under some oxygen limitation (microaerobic conditions).

For the parameter identification before and during the application of the NMPC, the data of five experiments performed on the laboratory scale were utilised (F1 – F4 and F7). The NMPC used is based on the OLFO algorithm. Historical data from other cultivations were used, which represents a modification of previous versions of the OLFO-controller (Witte 1996; Frahm et al. 2005). The modification was enabled by the lower computation time of the new mathematical model.

The implemented criterion for the NMPC was a combination of aiming to limit the ethanol concentration to 5 g L^{-1} (the maximum concentration in the pilot scale was 6 g L^{-1}) and maximising biomass growth. In the experiment, a maximum ethanol concentration was achieved which was only 4 g L^{-1} above the concentration of the pilot reactor process without reduced biomass growth (compared to the other experiments in the SDM F6 and F7). The maximum ethanol in of all other small-scale experiments (F1 – F7) were more than 10 g L^{-1} above the ethanol concentration of the pilot-scale experiment.

For the application of the mathematical model in an adaptive NMPC, short computation times are required, since the model parameters are identified, and the model-based process optimisation is performed during the cultivation. The chosen modelling approach enables the model utilisation in adaptive model predictive controllers due to the fast computation time of about one second per simulation run. Most other models, e.g. CFD simulations or compartmentalised models, utilised for the simulation of processes under heterogeneous conditions feature longer simulation times (Morchain 2017; Wang et al. 2015; Delafosse et al. 2014; Vrábel et al. 2000).

To understand the differences between the scales, further simulations of yeast cultivations were performed with the mechanistic model. These investigations were based on the experiments with the industrial medium and feeding strategy (F1 – F8) utilising the resulting parameter set of the NMPC application (chapter 9.3.3).

The model-based investigations extend the research which was performed by Sweere et al. (1988) who carried out experiments with a heterogeneous glucose concentration in a two-compartment scale-down model. However, Sweere et al. investigated glucose fluctuations with a mathematical model based on Sonnleitner et al. for ideally mixed systems. (Sweere and Mesters et al. 1988; Sonnleitner and Käppeli 1986)

The performed model-assisted investigations of the scale-dependent effects in this work revealed that slight changes of low sucrose concentrations can have a critical influence on the process performance. The simulation of small zones located at the feeding inlet resulted in comparatively low ethanol concentrations. In the small zone with high sucrose concentrations ethanol is produced, whereas in the rest of the reactor, no ethanol production was simulated, because the sucrose concentration was below the limit required for overflow metabolism. Similar simulations with a compartmentalised model considering flow patterns within a reactor were performed by Pigou et. al (2015) for an *E. coli* cultivation. Their simulations led to an acetate production (overflow metabolism) in a small zone of the reactor while in a different zone acetate was consumed (Pigou and Morchain 2015).

Overall, the study presented in this chapter confirms that *S. cerevisiae* cultivations are largely influenced by the sugar concentration. The decisive metabolic changes mainly occur at low sugar concentrations (ethanol production and consumption). The experimental and modelling studies carried out were able to show how non-ideal mixing affects the sugar concentrations within a bioreactor and the consequences on cultivations. Besides the heterogeneous conditions regarding the sugar concentration (and dissolved oxygen concentration), the combination of physical scale-down model and mathematical process model can be used to investigate factors that have a decisive influence on the sugar concentrations and thus the process:

1. Metabolism of the yeast and biomass density
2. Feeding strategy
3. Mixing / mixing time
4. Zones and volumes (within the bioreactor)

The cultivation processes can only be explained by consideration of all influencing factors and their interactions. The mathematical model is able to contribute substantially to the simultaneous investigation of the influence of these factors on the process behaviour. Adjusting these factors in simulations with the mathematical model support the design of

experimental strategies, designed with the aim of understanding different cultivation behaviours that would otherwise be difficult or impossible to explain.
Due to the exciting potential for industrial application, the application and properties of the NMPC should be further investigated and improved (e.g. with regard to the function of the optimisation criterium, the parameter identification, etc.) for cultivations under heterogeneous conditions in the next steps. It should also be examined whether the scale-down system, the mathematical model and the mDoE/NMPC can be transferred to other organisms (e.g. E. coli) or reactor configurations.

10 Conclusion and outlook

The transfer of cultivation processes between scales remains a challenge. Cultivations performed at different scales present different yields and exhibit different process time courses and performances.

The different behaviour of cultivation processes at different scales might be attributed to non-ideal mixing in larger scales. Additionally, the effects resulting from locally different substrate and oxygen concentrations could potentially overlap and interact. These effects and their influence on cultivation processes can be analysed and investigated with the utilisation of various physical scale-down models (SDM), as presented in numerous publications.

Publications have also demonstrated the suitability of dynamic process models for the description of biokinetics and for the investigation and improvement of process control strategies in ideally mixed reactors. The combination of mathematical models with a non-ideally mixed reactor allows for the consideration of industrially relevant heterogenous conditions in process simulations and design. However, the combination of models for the model-based development of process control strategies for processes in non-ideally mixed bioreactor systems with complex cultivation kinetics has yet to be shown.

In this context, two research questions have been formulated, which were addressed in this work:

1. Is a structured, mechanistic mathematical model, calculating several interconnected ideally mixed zones, able to approximate the flow patterns in stirred tank reactors and describe the effects caused by non-ideal mixing in biotechnological processes with complex kinetics?
2. How can feeding strategies for biotechnological processes in non-ideally mixed reactors be systematically designed?

In order to address the research questions, a combined modelling approach was employed, comprising of a mechanistic mathematical process model and a physical scale-down model consisting of two coupled stirred tank reactors. The influence of non-ideal mixed conditions on *S. cerevisiae* cultivations was investigated in laboratory-scale reactors including the physical scale-down model and ideally mixed systems and in a 1 m³ pilot-scale reactor for industrial yeast production.

In order to ensure the effective utilisation of the mathematical model, it was necessary to reduce both the time required for its development and the time taken for computations. This was made possible by developing modelling tools, which also enable the development of user-defined mathematical models. The mathematical model and the aforementioned modelling tools in model-based methods were used to facilitate the application of adaptive

process control and design. Parameter identification and the calculation of a control trajectory were completed within an hour, which is a crucial timescale for yeast processes that typically span only one to a few days.

A total of 27 cultivations were carried out in single, ideally mixed, lab-scale stirred tank reactors and in the non-ideally mixed physical scale-down system comprising of two interconnected stirred tank reactors. The investigations in this work were focused on cultivations with overlapping and interacting effects, such as heterogeneous (locally different) oxygen and substrate concentrations, which are also suspected in large production reactors. These overlapping effects exerted a considerable influence on the process performance of fed-batch *S. cerevisiae* processes and have not yet been published in this combination.

Data from exhaust gas analysis and the calculated respiratory quotient (RQ) have demonstrated that the production of ethanol under aerobic or anaerobic conditions and the aerobic uptake of ethanol can occur simultaneously in the non-ideally mixed SDM. The transitions between these metabolisms occurs at low glucose concentrations (below 0.2 g L^{-1}) and therefore play a key role in non-ideally mixed systems.

The mechanistic mathematical model is consisting of a network of zones model. The network of zones model is formed by linking multiple stirred tank reactor models that are ideally mixed. In each stirred tank reactor model the same biokinetic and physical-chemical submodels are integrated, containing all biokinetic and physico-chemical equations. This enables the simulation of cultivations both in ideally mixed and non-ideally mixed reactors with the same set of model parameters.

The mathematical model was successfully parameterised ($R^2 = 0.81 \pm 0.17$) using the experimental data from 16 laboratory-scale experiments with a complex medium. It is able to describe ideally mixed and non-ideally mixed processes in the scale-down model. Furthermore, the model was able to predict new cultivation courses for three experiments with a high degree of accuracy ($R^2 = 0.89 \pm 0.04$) and was thereby successfully validated. Additionally, the model exhibited proficiency in describing processes utilising other industrial substrates well in a case study ($R^2 > 0.9$).

Two methods were selected for the model-based design of feeding strategies and successfully employed for the process design in both ideally mixed and non-ideally mixed reactor systems. Fixed feeding profiles were designed prior to the start of the cultivation using a modified model-based design of experiments (mDoE). Moreover, adaptive feeding profiles were designed using a non-linear model predictive controller (NMPC) based on a variant of the open-loop feedback optimal (OLFO) algorithm.

The modified model-based design of experiments was employed to design a feeding strategy in a non-ideally mixed system with the objective of maximising the biomass density. This resulted in an approximately 20% increase in biomass density compared to other fed-batch experiments performed in the SDM.

The adaptive NMPC was used to design of a cultivation experiment in the physical scale-down model with the objective of reproducing an industrial process performed on the pilot-scale. By performing only one NMPC experiment, it was possible to achieve a cultivation with a low ethanol concentration that was in a similar order of magnitude compared to that of the pilot-scale experiment.

In conclusion, this work has laid the basis for the simulation of cultivation processes with complex kinetics that are influenced by heterogeneous conditions regarding the dissolved oxygen and substrates concentrations, and for the application of model-based design and optimisation strategies. The mathematical model is capable of simulating yeast cultivations on a laboratory scale under ideally mixed and non-ideally mixed conditions.

The new model structure, which incorporates a generalised biokinetic submodel will facilitate a systematic investigation of other organisms cultivated in non-ideally mixed reactor systems. This result is a mathematical model, which is suitable for analysing cultivations of different organisms in both ideally and non-ideally mixed reactor systems. The potential of this combination of mathematical and physical modelling for process design has been demonstrated in this work using cultivations of *S. cerevisiae*.

Furthermore, a basis for novel advanced scale-down strategies was established. The model-based design of experiments in scale-down systems is now possible. The interconnected mathematical stirred tank reactor models within the process model can be utilised to simulate different reactor configurations, without the need of identifying the model parameter set of the biokinetic submodel, which is parameterised independently from the mixing conditions. Cultivations conducted on the laboratory scale in ideally mixed tank reactors can be used to identify the parameters of the biokinetic submodel. Furthermore, the model can also be employed to derive how a process would need to be configured in a scale-down model in order to mimic the behaviour of the large-scale process. The mathematical model could also be employed to design targeted experiments to address new research questions regarding other microorganisms.

The newly developed approach, comprising of a flexible mathematical process model and experimental physical scale-down system has the potential to form the basis for new potential research and industrial applications for the model-assisted scale-down.

11 References

Abt V, Barz T, Cruz-Bournazou MN, Herwig C, Kroll P, Möller J, et al. Model-based tools for optimal experiments in bioprocess engineering. Current Opinion in Chemical Engineering. 2018;22:244–52. doi:10.1016/j.coche.2018.11.007.

Agger T, Spohr AB, Carlsen M, Nielsen J. Growth and product formation ofAspergillus oryzae during submerged cultivations: Verification of a morphologically structured model using fluorescent probes. Biotechnol. Bioeng. 1998;57(3):321–9. doi:10.1002/(sici)1097-0290(19980205)57:3<321::aid-bit9>3.0.co;2-j.

Akesson M, Karlsson EN, Hagander P, Axelsson JP, Tocaj A. On-line detection of acetate formation in Escherichia coli cultures using dissolved oxygen responses to feed transients. Biotechnol. Bioeng. 1999;64(5):590–8. doi:10.1002/(sici)1097-0290(19990905)64:5<590::aid-bit9>3.0.co;2-t.

Almquist J, Cvijovic M, Hatzimanikatis V, Nielsen J, Jirstrand M. Kinetic models in industrial biotechnology - Improving cell factory performance. Metab Eng. 2014;24:38–60. doi:10.1016/j.ymben.2014.03.007.

Amanullah A, McFarlane CM, Emery AN, Nienow AW. Scale-down model to simulate spatial pH variations in large-scale bioreactors. Biotechnol. Bioeng. 2001;73(5):390–9. doi:10.1002/bit.1072.

Amanullah A, Buckland BC, Nienow AW. Mixing in the Fermentation and Cell Culture Industries: Science and practice. Hoboken N.J.: Wiley-Interscience; 2004.

Andersen KB, Meyenburg K von. Are growth rates of Escherichia coli in batch cultures limited by respiration? J. Bacteriol. 1980;144(1):114–23. doi:10.1128/jb.144.1.114-123.1980.

Anderson MJ. RSM Simplified. 2nd ed. Taylor and Francis; 2016.

Andrews JF. A mathematical model for the continuous culture of microorganisms utilizing inhibitory substrates. Biotechnol. Bioeng. 1968;10(6):707–23. doi:10.1002/bit.260100602.

Antoniewicz MR. Methods and advances in metabolic flux analysis: a mini-review. J Ind Microbiol Biotechnol. 2015;42(3):317–25. doi:10.1007/s10295-015-1585-x.

Appl C. A new digital twin for enzymatic hydrolysis processes applied to model-based process design. University College London (UCL), Department of Biochemical Engineering; 2023.

Appl C, Moser A, Baganz F, Hass VC. Digital Twins for Bioprocess Control Strategy Development and Realisation. Adv Biochem Eng Biotechnol. 2021;177:63–94. doi:10.1007/10_2020_151.

Arnold JB. ggthemes: Extra Themes, Scales and Geoms for ’ggplot2’ 2021.

Ascanio G. Mixing time in stirred vessels: A review of experimental techniques. Chinese Journal of Chemical Engineering. 2015;23(7):1065–76. doi:10.1016/j.cjche.2014.10.022.

Babel W, Mller RH, Markuske KD. Improvement of growth yield of yeast on glucose to the maximum by using an additional energy source. Archiv. Mikrobiol. 1983;136(3):203–8. doi:10.1007/BF00409845.

Baeza JA. Principles of Bioprocess Control. In: Current Developments in Biotechnology and Bioengineering. Elsevier; 2017. p. 527–561. doi:10.1016/B978-0-444-63663-8.00018-5.

Bailer-Jones CAL. Practical Bayesian inference: A primer for physical scientists. Cambridge: Cambridge University Press; 2017.

Bailey JE. Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities. Biotechnol Progress. 1998;14(1):8–20. doi:10.1021/bp9701269.

Bajaj BK, Sharma S. Construction of killer industrial yeast Saccharomyces cerevisiae HAU-1 and its fermentation performance. Braz J Microbiol. 2010;41(2):477–85. doi:10.1590/S1517-83822010000200030.

Bareither R, Pollard D. A review of advanced small-scale parallel bioreactor technology for accelerated process development: current state and future need. Biotechnol Prog. 2011;27(1):2–14. doi:10.1002/btpr.522.

Barnes J, editor. The Cambridge companion to Aristotle. 1st ed. Cambridge: Cambridge Univ. Pr; 1995.

Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. J Stat Softw. 2015;67(1):1–48. doi:10.18637/jss.v067.i01.

Battley EH. Growth-Reaction Equations for Saccharomyces cerevisiae. Physiol Plant. 1960;13(2):192–203. doi:10.1111/j.1399-3054.1960.tb08023.x.

Bayer B, Stosch M von, Striedner G, Duerkop M. Comparison of Modeling Methods for DoE-Based Holistic Upstream Process Characterization. Biotechnol J. 2020;15(5):e1900551. doi:10.1002/biot.201900551.

Benbadis L, Cot M, Rigoulet M, Francois J. Isolation of two cell populations from yeast during high-level alcoholic fermentation that resemble quiescent and nonquiescent cells from the stationary phase on glucose. FEMS Yeast Res. 2009;9(8):1172–86. doi:10.1111/j.1567-1364.2009.00553.x.

Bengtsson H. matrixStats: Functions that Apply to Rows and Columns of Matrices (and to Vectors) 2022.

Bezzo F, Macchietto S, Pantelides CC. General hybrid multizonal/CFD approach for bioreactor modeling. AIChE Journal. 2003;49(8):2133–48. doi:10.1002/aic.690490821.

Birol G, Ündey C, Çinar A. A modular simulation package for fed-batch fermentation: penicillin production. Computers & Chemical Engineering. 2002;26(11):1553–65. doi:10.1016/S0098-1354(02)00127-8.

Blackman FF. Optima and limiting factors. Annals of Botany. 1905;1905(19):281.

Blanch HW, Rogers PL. Production of gramicidin S in batch and continuous culture. Biotechnol. Bioeng. 1971;13(6):843–64. doi:10.1002/bit.260130609.

Blesgen A. Entwicklung und Einsatz eines interaktiven Biogas-Echtzeit-Simulators. Bremen; 2009.

Blesgen A, Hass VC. Operator Training Simulator for Anaerobic Digestion Processes. IFAC Proceedings Volumes. 2010;43(6):353–8. doi:10.3182/20100707-3-BE-2012.0024.

Bohrer B. Reduced process performance of production scale industrial *S. cerevisiae* cultivations. Internal communication received from Dr. B. Bohrer (B+B Biotech), Villingen-Schwenningen; 2020.

Boole G. The mathematical analysis of logic: Being an essay towards a calculus of deductive reasoning. Cambridge [Massachusetts], London: Macmillan, Barclay, & Macmillan; George Bell; 1847.

Brüning S. Development of a generalized process model for optimization of biotechnological processes. Bremen: Information Resource Center der Jacobs University Bremen: Jacobs University Bremen; 2016.

Brüning S, Gerlach I, Pörtner R, Mandenius C-F, Hass VC. Modeling Suspension Cultures of Microbial and Mammalian Cells with an Adaptable Six-Compartment Model. Chem. Eng. Technol. 2017;40(5):956–66. doi:10.1002/ceat.201600639.

Bylund F, Collet E, Enfors S-O, Larsson G. Substrate gradient formation in the large-scale bioreactor lowers cell yield and increases by-product formation. Bioprocess Engineering. 1998;18(3):171. doi:10.1007/s004490050427.

Candioti LV, Zan MM de, Cámara MS, Goicoechea HC. Experimental design and multiple response optimization. Using the desirability function in analytical methods development. Talanta. 2014;124:123–38. doi:10.1016/j.talanta.2014.01.034.

Cappello V, Plais C, Vial C, Augier F. Scale-up of aerated bioreactors: CFD validation and application to the enzyme production by Trichoderma reesei. Chem Eng Sci. 2021;229:116033. doi:10.1016/j.ces.2020.116033.

Caramihai M, Severi I. Bioprocess Modeling and Control. In: Matovic MD, ed. Biomass now. Croatia: InTech; 2013. doi:10.5772/55362.

Carnell R. lhs: Latin Hypercube Samples 2022.

Casey GP, Ingledew WM. Ethanol tolerance in yeasts. Crit Rev Microbiol. 1986;13(3):219–80. doi:10.3109/10408418609108739.

Cenci F, Pankajakshan A, Facco P, Galvanin F. An exploratory model-based design of experiments approach to aid parameters identification and reduce model prediction uncertainty. Computers & Chemical Engineering. 2023;177:108353. doi:10.1016/j.compchemeng.2023.108353.

Cesselin B, Garrigues C, Pedersen MB, Roussel C, Gruss A, Gaudu P. Task Distribution between Acetate and Acetoin Pathways To Prolong Growth in Lactococcus lactis under Respiration Conditions. Appl Environ Microbiol 2018. doi:10.1128/AEM.01005-18.

Chang L, Liu X, Henson MA. Nonlinear model predictive control of fed-batch fermentations using dynamic flux balance models. Journal of Process Control. 2016;42:137–49. doi:10.1016/j.jprocont.2016.04.012.

Charlebois DA, Balázsi G. Modeling cell population dynamics. In Silico Biol. 2019;13(1-2):21–39. doi:10.3233/ISB-180470.

Chassagnole C, Noisommit-Rizzi N, Schmid JW, Mauch K, Reuss M. Dynamic modeling of the central carbon metabolism of Escherichia coli. Biotechnol Bioeng. 2002;79(1):53–73. doi:10.1002/bit.10288.

Chen L, Nguang SK, Chen XD, Li XM. Modelling and optimization of fed-batch fermentation processes using dynamic neural networks and genetic algorithms. Biochemical Engineering Journal. 2004;22(1):51–61. doi:10.1016/j.bej.2004.07.012.

Chmiel H, Takors R, Weuster-Botz D. Bioprozesstechnik. Berlin, Heidelberg: Springer Berlin Heidelberg; 2018.

Colijn C, Brandes A, Zucker J, Lun DS, Weiner B, Farhat MR, et al. Interpreting expression data with metabolic flux models: predicting Mycobacterium tuberculosis mycolic acid production. PLoS Comput Biol. 2009;5(8):e1000489. doi:10.1371/journal.pcbi.1000489.

Colin Cameron A, Windmeijer FA. An R-squared measure of goodness of fit for some common nonlinear regression models. J Econom. 1997;77(2):329–42. doi:10.1016/S0304-4076(96)01818-0.

Contois DE. Kinetics of bacterial growth: relationship between population density and specific growth rate of continuous cultures. J Gen Microbiol. 1959;21:40–50. doi:10.1099/00221287-21-1-40.

Crabtree HG. Observations on the carbohydrate metabolism of tumours. Biochem J. 1929;23(3):536–45. doi:10.1042/bj0230536.

Crater JS, Lievense JC. Scale-up of industrial microbial processes. FEMS Microbiol Lett 2018. doi:10.1093/femsle/fny138.

Craven S, Shirsat N, Whelan J, Glennon B. Process model comparison and transferability across bioreactor scales and modes of operation for a mammalian cell bioprocess. Biotechnol Prog. 2013;29(1):186–96. doi:10.1002/btpr.1664.

Craven S, Whelan J, Glennon B. Glucose concentration control of a fed-batch mammalian cell bioprocess using a nonlinear model predictive controller. Journal of Process Control. 2014;24(4):344–57. doi:10.1016/j.jprocont.2014.02.007.

Daume S, Kofler S, Kager J, Kroll P, Herwig C. Generic Workflow for the Setup of Mechanistic Process Models. Methods Mol Biol. 2020;2095:189–211. doi:10.1007/978-1-0716-0191-4_11.

Deflorian M, Zaglauer S. Design of Experiments for nonlinear dynamic system identification. IFAC Proceedings Volumes. 2011;44(1):13179–84. doi:10.3182/20110828-6-IT-1002.01502.

Deken RH de. The Crabtree effect: a regulatory system in yeast. J Gen Microbiol. 1966;44(2):149–56. doi:10.1099/00221287-44-2-149.

Delafosse A, Collignon M-L, Calvo S, Delvigne F, Crine M, Thonart P, Toye D. CFD-based compartment model for description of mixing in bioreactors. Chemical Engineering Science. 2014;106:76–85. doi:10.1016/j.ces.2013.11.033.

Delafosse A, Calvo S, Collignon M-L, Delvigne F, Crine M, Toye D. Euler–Lagrange approach to model heterogeneities in stirred tank bioreactors – Comparison to experimental flow characterization and particle tracking. Chemical Engineering Science. 2015;134:457–66. doi:10.1016/j.ces.2015.05.045.

Delvigne F. Etude de l'impact des conditions hydrodynamiques du bio-réacteur sur la conduite et l'extrapolation des bioprocédés: Dissertation. Gembloux: FACULTE UNIVERSITAIRE DES SCIENCES; 2006.

Delvigne F, Noorman H. Scale-up/Scale-down of microbial bioprocesses: a modern light on an old issue. Microb Biotechnol. 2017;10(4):685–7. doi:10.1111/1751-7915.12732.

Delvigne F, Takors R, Mudde R, van Gulik W, Noorman H. Bioprocess scale-up/down as integrative enabling technology: from fluid mechanics to systems biology and beyond. Microb Biotechnol. 2017;10(5):1267–74. doi:10.1111/1751-7915.12803.

Desouza A, Pike RW. Fluid dynamics and flow patterns in stirred tanks with a turbine impeller. Can. J. Chem. Eng. 1972;50(1):15–23. doi:10.1002/cjce.5450500104.

Dewasme L, Amribt Z, Santos LO, Hantson A-L, Bogaerts P, Wouwer AV. Hybridoma cell culture optimization using nonlinear model predictive control. IFAC Proceedings Volumes. 2013;46(31):60–5. doi:10.3182/20131216-3-IN-2044.00045.

Doherty SK, Gomm JB, Williams D. Experiment design considerations for non-linear system identification using neural networks. Computers & Chemical Engineering. 1997;21(3):327–46. doi:10.1016/S0098-1354(96)00003-8.

Du Y-H, Wang M-Y, Yang L-H, Tong L-L, Guo D-S, Ji X-J. Optimization and Scale-Up of Fermentation Processes Driven by Models. Bioengineering (Basel) 2022. doi:10.3390/bioengineering9090473.

Dusny C, Schmid A. Microfluidic single-cell analysis links boundary environments and individual microbial phenotypes. Environ Microbiol. 2015;17(6):1839–56. doi:10.1111/1462-2920.12667.

Eiteman MA, Altman E. Overcoming acetate in Escherichia coli recombinant protein fermentations. Trends Biotechnol. 2006;24(11):530–6. doi:10.1016/j.tibtech.2006.09.001.

El Saddik A. Digital Twins: The Convergence of Multimedia Technologies. IEEE MultiMedia. 2018;25(2):87–92. doi:10.1109/MMUL.2018.023121167.

Elqotbi M, Vlaev SD, Montastruc L, Nikov I. CFD modelling of two-phase stirred bioreaction systems by segregated solution of the Euler–Euler model. Computers & Chemical Engineering. 2013;48:113–20. doi:10.1016/j.compchemeng.2012.08.005.

Enfors SO, Jahic M, Rozkov A, Xu B, Hecker M, Jürgen B, et al. Physiological responses to mixing in large scale bioreactors. J Biotechnol. 2001;85(2):175–85. doi:10.1016/S0168-1656(00)00365-5.

European Parliament. Regulation (EU) 2018/848 of the European Parliament and of the Council of 30 May 2018 on organic production and labelling of organic products and repealing Council Regulation (EC) No 834/2007; 2018.

Falco B de, Giannino F, Carteni F, Mazzoleni S, Kim D-H. Metabolic flux analysis: a comprehensive review on sample preparation, analytical techniques, data analysis, computational modelling, and main application areas. RSC Adv. 2022;12(39):25528–48. doi:10.1039/d2ra03326g.

Fiechter A, Fuhrmann GF, Käppeli O. Regulation of glucose metabolism in growing yeast cells. Adv Microb Physiol. 1981;22:123–83. doi:10.1016/s0065-2911(08)60327-6.

Filev DP, Kishimoto M, Sengupta S, Yoshida T, Taguchi H. Application of the fuzzy theory to simulation of batch fermentation. J. Ferment. Technol. 1985;1985(63:6):545–53.

Fink M, Cserjan-Puschmann M, Reinisch D, Striedner G. High-throughput microbioreactor provides a capable tool for early stage bioprocess development. Sci Rep. 2021;11(1):2056. doi:10.1038/s41598-021-81633-6.

Fitschen J, Hofmann S, Kursula L, Haase I, Wucherpfennig T, Schlüter M. Advances in Characterization of Industrial Bioreactors for Cell Culture Process. In: Pörtner R, ed. Biopharmaceutical Manufacturing. Cham: Springer International Publishing; 2023. p. 67–111. doi:10.1007/978-3-031-45669-5_3.

Fleet GH. Yeast-growth during fermentation. Wine microbiology and biotechnology. 1993:27–54.

Forbes MG, Patwardhan RS, Hamadah H, Gopaluni RB. Model Predictive Control in Industry: Challenges and Opportunities. IFAC-PapersOnLine. 2015;48(8):531–8. doi:10.1016/j.ifacol.2015.09.022.

Formenti LR, Nørregaard A, Bolic A, Hernandez DQ, Hagemann T, Heins A-L, et al. Challenges in industrial fermentation technology research. Biotechnol J. 2014;9(6):727–38. doi:10.1002/biot.201300236.

Frahm B, Lane P, Atzert H, Munack A, Hoffmann M, Hass VC, Pörtner R. Adaptive, model-based control by the Open-Loop-Feedback-Optimal (OLFO) controller for the effective fed-batch cultivation of hybridoma cells. Biotechnol Prog. 2002;18(5):1095–103. doi:10.1021/bp020035y.

Frahm B, Hass VC, Lane P, Munack A, Märkl H, Pörtner R. Fed-Batch-Kultivierung tierischer Zellen - Eine Herausforderung zur adaptiven, modellbasierten Steuerung. Chemie Ingenieur Technik. 2003;75(4):457–60. doi:10.1002/cite.200390093.

Frahm B, Lane P, Märkl H, Pörtner R. Improvement of a mammalian cell culture process by adaptive, model-based dialysis fed-batch cultivation and suppression of apoptosis. Bioprocess Biosyst Eng. 2003;26(1):1–10. doi:10.1007/s00449-003-0335-z.

Frahm B, Pörtner R, Lane P, Munack A. Optimierung und Steuerung von Zellkultur-Fed-Batch-Prozessen mittels einer Kollokationsmethode. Chemie Ingenieur Technik. 2005;77(4):429–35. doi:10.1002/cite.200407094.

Franceschini G, Macchietto S. Model-based design of experiments for parameter precision: State of the art. Chemical Engineering Science. 2008;63(19):4846–72. doi:10.1016/j.ces.2007.11.034.

Franco-Lara E, Weuster-Botz D. Estimation of optimal feeding strategies for fed-batch bioprocesses. Bioprocess Biosyst Eng. 2005;27(4):255–62. doi:10.1007/s00449-005-0415-3.

Fredrickson AG, Megee RD, Tsuchiya HM. Mathematical Models for Fermentation Processes. In: Fredrickson AG, Megee RD, Tsuchiya HM, eds. Mathematical Models

for Fermentation Processes. Elsevier; 1970. p. 419–465. doi:10.1016/S0065-2164(08)70413-1.

Galvanauskas V, Simutis R, Lübbert A. Hybrid process models for process optimisation, monitoring and control. Bioprocess Biosyst Eng. 2004;26(6):393–400. doi:10.1007/s00449-004-0385-x.

Galvanauskas V, Grigs O, Vanags J, Dubencovs K, Stepanova V. Model-based optimization and pO 2 control of fed-batch Escherichia coli and Saccharomyces cerevisiae cultivation processes. Eng. Life Sci. 2013;13(2):172–84. doi:10.1002/elsc.201200012.

Garcia-Ochoa F, Gomez E. Bioreactor scale-up and oxygen transfer rate in microbial processes: An overview. Biotechnology Advances. 2009;27(2):153–76. doi:10.1016/j.biotechadv.2008.10.006.

Gargalo CL, Cheali P, Posada JA, Carvalho A, Gernaey KV, Sin G. Assessing the environmental sustainability of early stage design for bioprocesses under uncertainties: An analysis of glycerol bioconversion. J Clean Prod. 2016;139:1245–60. doi:10.1016/j.jclepro.2016.08.156.

Gassenmeier V, Deppe S, Hernández Rodríguez T, Kuhfuß F, Moser A, Hass VC, et al. Model-assisted DoE applied to microalgae processes. Current Research in Biotechnology. 2022;4:102–18. doi:10.1016/j.crbiot.2022.01.005.

George S, Larsson G, Enfors S-O. A scale-down two-compartment reactor with controlled substrate oscillations: Metabolic response of Saccharomyces cerevisiae. Bioprocess Biosyst Eng. 1993;9(6):249–57. doi:10.1007/BF01061530.

George S, Larsson G, Olsson K, Enfors S-O. Comparison of the Baker's yeast process performance in laboratory and production scale. Bioprocess Eng. 1998;18(2):135–42. doi:10.1007/PL00008979.

Gerlach I, Hass VC, Brüning S, Mandenius C-F. Virtual bioreactor cultivation for operator training and simulation: application to ethanol and protein production. J. Chem. Technol. Biotechnol. 2013;88(12):2159–68. doi:10.1002/jctb.4079.

Giorgi FM, Ceraolo C, Mercatelli D. The R Language: An Engine for Bioinformatics and Data Science. Life (Basel) 2022. doi:10.3390/life12050648.

Glaessgen EH, Stargel DS. The Digital Twin Paradigm for Future NASA and U.S. Air Force Vehicles. Honolulu, U.S.; 2012.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, et al. Life with 6000 genes. Science. 1996;274(5287):546, 563-7. doi:10.1126/science.274.5287.546.

González-Figueredo C, Alejandro Flores-Estrella R, A. Rojas-Rejón O. Fermentation: Metabolism, Kinetic Models, and Bioprocessing. In: Shiomi N, ed. Current Topics in Biochemical Engineering. IntechOpen; 2019. doi:10.5772/intechopen.82195.

Grahovac J, Jokić A, Dodić J, Vučurović D, Dodić S. Modelling and prediction of bioethanol production from intermediates and byproduct of sugar beet processing using neural networks. Renewable Energy. 2016;85:953–8. doi:10.1016/j.renene.2015.07.054.

Guo M, Song W. The growing U.S. bioeconomy: Drivers, development and constraints. N Biotechnol. 2019;49:48–57. doi:10.1016/j.nbt.2018.08.005.

Gustafsson TK, Skrifvars BO, Sandstroem KV, Waller KV. Modeling of pH for Control. Ind. Eng. Chem. Res. 1995;34(3):820–7. doi:10.1021/ie00042a014.

Hameri T, Fengos G, Hatzimanikatis V. The effects of model complexity and size on metabolic flux distribution and control: case study in Escherichia coli. BMC Bioinformatics. 2021;22(1):134. doi:10.1186/s12859-021-04066-y.

Haringa C, Tang W, Deshmukh AT, Xia J, Reuss M, Heijnen JJ, et al. Euler-Lagrange computational fluid dynamics for (bio)reactor scale down: An analysis of organism lifelines. Engineering in Life Sciences. 2016;16(7):652–63. doi:10.1002/elsc.201600061.

Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020;585(7825):357–62. doi:10.1038/s41586-020-2649-2.

Hartigan JA, Wong MA. Algorithm AS 136: A K-Means Clustering Algorithm. Applied Statistics. 1979;28(1):100. doi:10.2307/2346830.

Hass VC, Kuhnen F, Schoop K-M. An environment for the development of operator training systems (OTS) from chemical engineering models. In: European Symposium on Computer-Aided Process Engineering-15, 38th European Symposium of the Working Party on Computer Aided Process Engineering. Elsevier; 2005. p. 289–293. doi:10.1016/S1570-7946(05)80170-1.

Hass VC, Li M, Kuhnen F, Schoop KM, Pörtner R. Towards the rapid development of control strategies for cultivation processes using mathematical models and training simulators. The Hague, The Netherlands; 2013.

Hass VC. Operator Training Simulators for Bioreactors. In: Bioreactors. John Wiley & Sons, Ltd; 2016. p. 453–486. doi:10.1002/9783527683369.ch16.

Hedengren JD, Shishavan RA, Powell KM, Edgar TF. Nonlinear modeling, estimation and predictive control in APMonitor. Computers & Chemical Engineering. 2014;70:133–48. doi:10.1016/j.compchemeng.2014.04.013.

Heins A-L, Weuster-Botz D. Population heterogeneity in microbial bioprocesses: origin, analysis, mechanisms, and future perspectives. Bioprocess Biosyst Eng. 2018;41(7):889–916. doi:10.1007/s00449-018-1922-3.

Henzler H-J. Verfahrenstechnische Auslegungsunterlagen für Rührbehälter als Fermenter. Chem Ing Tech. 1982;54(5):461–76. doi:10.1002/cite.330540510.

Hernández Rodríguez T, Posch C, Schmutzhard J, Stettner J, Weihs C, Pörtner R, Frahm B. Predicting industrial-scale cell culture seed trains-A Bayesian framework for model fitting and parameter estimation, dealing with uncertainty in measurements and model parameters, applied to a nonlinear kinetic cell culture model, using an MCMC method. Biotechnol Bioeng 2019. doi:10.1002/bit.27125.

Herskowitz I. Life cycle of the budding yeast Saccharomyces cerevisiae. Microbiol Rev. 1988;52(4):536–53. doi:10.1128/mr.52.4.536-553.1988.

Hewitt CJ, Nienow AW. The Scale-Up of Microbial Batch and Fed-Batch Fermentation Processes. 2010;62:105–35. doi:10.1016/S0065-2164(07)62005-X.

Himmelblau DM, Riggs JB. Basic Principles and Calculations in Chemical Engineering, Eight Edition. [Place of publication not identified]: Prentice Hall; 2012.

Hirschmann R. Evaluating the potential of anaerobic production of ethyl(3)hydroxybutyrate for integration in biorefineries. London, United Kingdom: UCL; 2021.

Hodge DB, Karim MN. Modeling and advanced control of recombinant Zymomonas mobilis fed-batch fermentation. Biotechnol Progress. 2002;18(3):572–9. doi:10.1021/bp0155181.

Hollander JA den, Ugurbil K, Brown TR, Bednar M, Redfield C, Shulman RG. Studies of anaerobic and aerobic glycolysis in Saccharomyces cerevisiae. Biochemistry. 1986;25(1):203–11. doi:10.1021/bi00349a029.

Honda H, Kobayashi T. Fuzzy control of bioprocess. J Biosci Bioeng. 2000;89(5):401–8. doi:10.1016/S1389-1723(00)89087-8.

Horiuchi J-I. Fuzzy modeling and control of biological processes. J Biosci Bioeng. 2002;94(6):574–8. doi:10.1016/S1389-1723(02)80197-9.

Hristozov I, Pencheva T, Staerk E, Hitzmann B, Scheper T, St. Tzonkov. Functional States Modelling of Batch Aerobic Yeast Growth Process. Biotechnology & Biotechnological Equipment. 2001;15(2):132–5. doi:10.1080/13102818.2001.10819145.

Humphrey A. Shake Flask to Fermentor: What Have We Learned? Biotechnol Prog. 1998;1998(14):3–7.

Hunter JD. Matplotlib: A 2D graphics environment. Comput Sci Eng. 2007;9(3):90–5. doi:10.1109/MCSE.2007.55.

Ihaka R, Gentleman R. R: A Language for Data Analysis and Graphics. J Comput Graph Stat. 1996;5(3):299–314. doi:10.1080/10618600.1996.10474713.

Iwatani S, Yamada Y, Usuda Y. Metabolic flux analysis in biotechnology processes. Biotechnology Letters. 2008;30(5):791–9. doi:10.1007/s10529-008-9633-5.

Jenzsch M, Gnoth S, Beck M, Kleinschmidt M, Simutis R, Lübbert A. Open-loop control of the biomass concentration within the growth phase of recombinant protein production processes. J Biotechnol. 2006;127(1):84–94. doi:10.1016/j.jbiotec.2006.06.004.

Johnson SG. The NLopt nonlinear-optimization package. 2022.

Ju L-K, Chase GG. Improved scale-up strategies of bioreactors. Bioprocess Eng. 1992;8(1-2):49–53. doi:10.1007/BF00369263.

Junker BH. Scale-up methodologies for Escherichia coli and yeast fermentation processes. J Biosci Bioeng. 2004;97(6):347–64. doi:10.1016/S1389-1723(04)70218-2.

Junne S, Klingner A, Itzeck D, Brand E, Neubauer P. Consistency of Scale-Up from Bioprocess Development to Production. 2012;2012:511–43. doi:10.1002/9783527653096.ch16.

Kandler O. Carbohydrate metabolism in lactic acid bacteria. Antonie Van Leeuwenhoek. 1983;49(3):209–24. doi:10.1007/BF00399499.

Kar T, Delvigne F, Masson M, Destain J, Thonart P. Investigation of the effect of different extracellular factors on the lipase production by Yarrowia lipolityca on the basis of a scale-down approach. J Ind Microbiol Biotechnol. 2008;35(9):1053–9. doi:10.1007/s10295-008-0382-1.

Karakuzu C, Türker M, Öztürk S. Modelling, on-line state estimation and fuzzy control of production scale fed-batch baker's yeast fermentation. Control Engineering Practice. 2006;14(8):959–74. doi:10.1016/j.conengprac.2005.05.007.

Karathia H, Vilaprinyo E, Sorribas A, Alves R. Saccharomyces cerevisiae as a model organism: a comparative study. PLoS One. 2011;6(2):e16015. doi:10.1371/journal.pone.0016015.

Karim M, Yoshida T, Rivera SL, Saucedo VM, Eikens B, Oh G-S. Global and local neural network models in biotechnology: Application to different cultivation processes. Journal of Fermentation and Bioengineering. 1997;83(1):1–11. doi:10.1016/S0922-338X(97)87318-7.

Kim JW, Krausch N, Aizpuru J, Barz T, Lucia S, Neubauer P, Cruz Bournazou MN. Model predictive control and moving horizon estimation for adaptive optimal bolus feeding in high-throughput cultivation of E. coli. Comput Chem Eng. 2023;172:108158. doi:10.1016/j.compchemeng.2023.108158.

Koh BT, Nakashimada U, Pfeiffer M, Yap MGS. Comparison of acetate inhibition on growth of host and recombinantE. coli K12 strains. Biotechnol Lett. 1992;14(12):1115–8. doi:10.1007/BF01027012.

Kolehmainen M, Rönkkö P, Raatikainen O. Monitoring of yeast fermentation by ion mobility spectrometry measurement and data visualisation with Self-Organizing Maps. Analytica Chimica Acta. 2003;484(1):93–100. doi:10.1016/S0003-2670(03)00307-6.

Konstantinov KB, Yoshida T. An expert approach for control of fermentation processes as variable structure plants. Journal of Fermentation and Bioengineering. 1990;70(1):48–57. doi:10.1016/0922-338X(90)90030-Z.

Kortmann H, Blank LM, Schmid A. Single cell analysis reveals unexpected growth phenotype of S. cerevisiae. Cytometry A. 2009;75(2):130–9. doi:10.1002/cyto.a.20684.

Kosko B. Neural networks and fuzzy systems: A dynamical systems approach to machine intelligence. Englewood Ciffs, NJ: Prentice-Hall; 1992.

Kovárová-Kovar K, Gehlen S, Kunze A, Keller T, Däniken R von, Kolb M, van Loon AP. Application of model-predictive control based on artificial neural networks to optimize the fed-batch process for riboflavin production. J Biotechnol. 2000;79(1):39–52. doi:10.1016/S0168-1656(00)00211-X.

Kreutz C, Timmer J. Systems biology: experimental design. FEBS J. 2009;276(4):923–42. doi:10.1111/j.1742-4658.2008.06843.x.

Kroll P, Hofer A, Ulonska S, Kager J, Herwig C. Model-Based Methods in the Biopharmaceutical Process Lifecycle. Pharm Res. 2017;34(12):2596–613. doi:10.1007/s11095-017-2308-y.

Kuchemüller KB, Pörtner R, Möller J. Digital Twins and Their Role in Model-Assisted Design of Experiments. Adv Biochem Eng Biotechnol 2020. doi:10.1007/10_2020_136.

Kuhnen F. Modellierungssystem eStIM - Kurzeinführung. Unpublished manuscript. Hochschule Bremen, Bremen; 2008.

Kuntzsch S. Energy Efficiency Investigations with a new Operator Training Simulator for Biorefineries. Bremen; 2014.

Kuschel M, Siebler F, Takors R. Lagrangian Trajectories to Predict the Formation of Population Heterogeneity in Large-Scale Bioreactors. Bioengineering (Basel) 2017. doi:10.3390/bioengineering4020027.

Lapin A, Müller D, Reuss M. Dynamic Behavior of Microbial Populations in Stirred Bioreactors Simulated with Euler−Lagrange Methods: Traveling along the Lifelines of Single Cells. Ind. Eng. Chem. Res. 2004;43(16):4647–56. doi:10.1021/ie030786k.

Lapin A, Schmid J, Reuss M. Modeling the dynamics of E. coli populations in the three-dimensional turbulent field of a stirred-tank bioreactor—A structured–segregated approach. Chemical Engineering Science. 2006;61(14):4783–97. doi:10.1016/j.ces.2006.03.003.

Lapin A, Klann M, Reuss M. Multi-scale spatio-temporal modeling: lifelines of microorganisms in bioreactors and tracking molecules in cells. Adv Biochem Eng Biotechnol. 2010;121:23–43. doi:10.1007/10_2009_53.

Lara AR, Galindo E, Ramírez OT, Palomares LA. Living With Heterogeneities in Bioreactors: Understanding the Effects of Environmental Gradients on Cells. MB. 2006;34(3):355–82. doi:10.1385/MB:34:3:355.

Lara AR, Leal L, Flores N, Gosset G, Bolívar F, Ramírez OT. Transcriptional and metabolic response of recombinant Escherichia coli to spatial dissolved oxygen tension gradients simulated in a scale-down system. Biotechnol Bioeng. 2006;93(2):372–85. doi:10.1002/bit.20704.

Lara AR, Palomares LA, Ramírez OT. Scale-Down: Simulating Large-Scale Cultures in the Laboratory. Industrial Biotechnology: Products and Processes (eds C. Wittmann and J.C. Liao). 2016:55–79. doi:10.1002/9783527807833.ch2.

Larsson G, Törnkvist M, Wernersson ES, Trgrdh C, Noorman H, Enfors S-O. Substrate gradients in bioreactors: origin and consequences. Bioprocess Biosyst Eng. 1996;14(6):281–9. doi:10.1007/BF00369471.

Laurí D, Lennox B, Camacho J. Model predictive control for batch processes: Ensuring validity of predictions. Journal of Process Control. 2014;24(1):239–49. doi:10.1016/j.jprocont.2013.11.005.

Ławryńczuk M. Modelling and nonlinear predictive control of a yeast fermentation biochemical reactor using neural networks. Chemical Engineering Journal. 2008;145(2):290–307. doi:10.1016/j.cej.2008.08.005.

Lee K-M, Gilmore DF. Statistical Experimental Design for Bioprocess Modeling and Optimization Analysis: Repeated-Measures Method for Dynamic Biotechnology Process. ABAB. 2006;135(2):101–16. doi:10.1385/ABAB:135:2:101.

Leifheit J, Heine T, Kawohl M, King R. Rechnergestützte halbautomatische Modellierung biotechnologischer Prozesse (Semiautomatic Modeling of Biotechnical Processes). at – Automatisierungstechnik 2007. doi:10.1524/auto.2007.55.5.211.

Lemoigne M, Aubert J-P, Millet J. La production d'alcool et le rendement de croissance de la levure de boulangerie cultivée en aérobiose. Annales de l’Institut Pasteur (Paris). 1954;1954(87(4)):427–39.

Lemoine A, Maya Martínez-Iturralde N, Spann R, Neubauer P, Junne S. Response of Corynebacterium glutamicum exposed to oscillating cultivation conditions in a two- and a novel three-compartment scale-down bioreactor. Biotechnol Bioeng. 2015;112(6):1220–31. doi:10.1002/bit.25543.

Lemoine A, Delvigne F, Bockisch A, Neubauer P, Junne S. Tools for the determination of population heterogeneity caused by inhomogeneous cultivation conditions. J Biotechnol. 2017;251:84–93. doi:10.1016/j.jbiotec.2017.03.020.

Lenth RV. Response-Surface Methods in R, Using rsm. J Stat Softw. 2009;32(7):1–17. doi:10.18637/jss.v032.i07.

Leuenberger HGW. Cultivation of Saccharomyces cerevisiae in continuous culture. Archiv. Mikrobiol. 1972;83(4):347–58. doi:10.1007/BF00425247.

Levenspiel O. Chemical reaction engineering. 2nd ed. New York: Wiley; 1972.

Levermann P, Freiberger F, Katha U, Zaun H, Möller J, Hass VC, et al. NMPC-Based Workflow for Simultaneous Process and Model Development Applied to a Fed-Batch Process for Recombinant C. glutamicum. Processes (Basel). 2020;8(10):1313. doi:10.3390/pr8101313.

Li M. Adaptive Predictive Control by Open-Loop- Feedback-Optimal Controller for Cultivation Processes; 2015.

Lidstrom ME, Konopka MC. The role of physiological heterogeneity in microbial population behavior. Nat Chem Biol. 2010;6(10):705–12. doi:10.1038/nchembio.436.

Limberg MH, Schulte J, Aryani T, Mahr R, Baumgart M, Bott M, et al. Metabolic profile of 1,5-diaminopentane producing Corynebacterium glutamicum under scale-down conditions: Blueprint for robustness to bioreactor inhomogeneities. Biotechnol Bioeng. 2017;114(3):560–75. doi:10.1002/bit.26184.

Lin HY, Neubauer P. Influence of controlled glucose oscillations on a fed-batch process of recombinant Escherichia coli. J Biotechnol. 2000;79(1):27–37. doi:10.1016/s0168-1656(00)00217-0.

Lloyd S. Least squares quantization in PCM. IEEE Trans. Inform. Theory. 1982;28(2):129–37. doi:10.1109/TIT.1982.1056489.

Lukasiewicz J. Untersuchungen über den Aussagenkalkül. Comptes Rendus des Séances de la Sociélé des Scierices et des Lettres des Varsovie Classe III. 1930;1930(23):30–50.

Luttmann R, Munack A, Thoma M. Mathematical modelling, parameter identification and adaptive control of single cell protein processes in tower loop bioreactors. In: Agricultural Feedstock and Waste Treatment and Engineering. Berlin/Heidelberg: Springer-Verlag; 1985. p. 95–205. doi:10.1007/BFb0009526.

Madni A, Madni C, Lucero S. Leveraging Digital Twin Technology in Model-Based Systems Engineering. Systems. 2019;7(1):7. doi:10.3390/systems7010007.

Mandenius C-F, Brundin A. Bioprocess optimization using design-of-experiments methodology. Biotechnol Prog. 2008;24(6):1191–203. doi:10.1002/btpr.67.

Manfredini R, Cavallera V, Marini L, Donati G. Mixing and oxygen transfer in conventional stirred fermentors. Biotechnol. Bioeng. 1983;25(12):3115–31. doi:10.1002/bit.260251224.

Mann R, Pillai SK, El-Hamouz AM, Ying P, Togatorop A, Edwards RB. Computational fluid mixing for stirred vessels: progress from seeing to believing. The Chemical Engineering Journal and the Biochemical Engineering Journal. 1995;59(1):39–50. doi:10.1016/0923-0467(95)02998-2.

Márquez-Baños VE, La Concha-Gómez AD de, Valencia-López JJ, López-Yáñez A, Ramírez-Muñoz J. Shear rate and direct numerical calculation of the Metzner-Otto constant for a pitched blade turbine. Journal of Food Engineering. 2019;257:10–8. doi:10.1016/j.jfoodeng.2019.03.021.

Martins BMC, Locke JCW. Microbial individuality: how single-cell heterogeneity enables population level strategies. Curr Opin Microbiol. 2015;24:104–12. doi:10.1016/j.mib.2015.01.003.

McKay MD, Beckman RJ, Conover WJ. Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics. 1979;21(2):239–45. doi:10.1080/00401706.1979.10489755.

Mears L, Stocks SM, Sin G, Gernaey KV. A review of control strategies for manipulating the feed rate in fed-batch fermentation processes. J Biotechnol. 2017;245:34–46. doi:10.1016/j.jbiotec.2017.01.008.

Megee RD, Kinoshita S, Fredrickson AG, Tsuchiya HM. Differentiation and product formation in molds. Biotechnol. Bioeng. 1970;12(5):771–801. doi:10.1002/bit.260120507.

Mersmann O, Trautmann H, Steuer D, Bornkamp B. truncnorm: Truncated Normal Distribution 2018.

Ming F, Howell JA, Canovas-Diaz M. Mathematical Simulation of Anaerobic Stratified Biofilm Processes. Computer Applications in Fermentation Technology. 1988:69–77. doi:10.1007/978-94-009-1141-3_9.

Moilanen P, Laakkonen M, Aittamaa J. Modelling fermenters with CFD. In: European Symposium on Computer-Aided Process Engineering-15, 38th European Symposium of the Working Party on Computer Aided Process Engineering. Elsevier; 2005. p. 709–714. doi:10.1016/S1570-7946(05)80240-8.

Möller J, Pörtner R, editors. Model-Based Design of Process Strategies for Cell Culture Bioprocesses: State of the Art and New Perspectives. InTech; 2017.

Möller J, Eibl R, Eibl D, Pörtner R. Model-based DoE for feed batch cultivation of a CHO cell line. BMC Proc. 2015;9(Suppl 9):P42. doi:10.1186/1753-6561-9-S9-P42.

Möller J, Kuchemüller KB, Steinmetz T, Koopmann KS, Pörtner R. Model-assisted Design of Experiments as a concept for knowledge-based bioprocess development. Bioprocess Biosyst Eng 2019. doi:10.1007/s00449-019-02089-7.

Möller J, Hernández Rodríguez T, Müller J, Arndt L, Kuchemüller KB, Frahm B, et al. Model uncertainty-based evaluation of process strategies during scale-up of biopharmaceutical processes. Computers & Chemical Engineering. 2020;134:106693. doi:10.1016/j.compchemeng.2019.106693.

Monod J. Recherches sur la croissance des cultures bactériennes. Paris: Hermann; 1942.

Monod J. The Growth of Bacterial Cultures. Annu. Rev. Microbiol. 1949;3(1):371–94. doi:10.1146/annurev.mi.03.100149.002103.

Monsalve-Bravo GM, Garelli F, Mozumder MSI, Alvarez H, Battista H de. Model-based scale-up methodology for aerobic fed-batch bioprocesses: application to polyhydroxybutyrate (PHB) production. Bioprocess Biosyst Eng. 2015;38(6):1179–90. doi:10.1007/s00449-015-1360-4.

Montague G, Morris J. Neural-network contributions in biotechnology. Trends Biotechnol. 1994;12(8):312–24. doi:10.1016/0167-7799(94)90048-5.

Moran EF, Ostrom E. Seeing the Forest and the Trees. The MIT Press; 2005.

Morchain J, Gabelle J-C, Cockx A. A coupled population balance model and CFD approach for the simulation of mixing issues in lab-scale and industrial bioreactors. AIChE Journal. 2014;60(1):27–40. doi:10.1002/aic.14238.

Morchain J. Numerical Tools for Scaling Up Bioreactors. In: Current Developments in Biotechnology and Bioengineering. Elsevier; 2017. p. 495–523. doi:10.1016/B978-0-444-63663-8.00017-3.

Moser H. The dynamics of bacterial populations maintained in the chemostat: Publication 614. Washington; 1958.

Moser A. Kinetics of batch fermentations. Biotechnology. 1985:243–83.

Moser A. Bioprocess Technology: Kinetics and Reactors. New York, NY: Springer New York; 1988.

Moser A, Appl C, Brüning S, Hass VC. Mechanistic Mathematical Models as a Basis for Digital Twins. Adv Biochem Eng Biotechnol 2020. doi:10.1007/10_2020_152.

Moser A, Kuchemüller KB, Deppe S, Hernández Rodríguez T, Frahm B, Pörtner R, et al. Model-assisted DoE software: optimization of growth and biocatalysis in

Saccharomyces cerevisiae bioprocesses. Bioprocess Eng. 2021. doi:10.1007/s00449-020-02478-3.

Moser A, Appl C, Pörtner R, Baganz F, Hass VC. A New Concept for the Rapid Development of Digital Twin Core Models for Bioprocesses in Various Reactor Designs. Fermentation. 2024;10(9):463. doi:10.3390/fermentation10090463.

Müller D, Exler S, Aguilera-Vázquez L, Guerrero-Martín E, Reuss M. Cyclic AMP mediates the cell cycle dynamics of energy metabolism in Saccharomyces cerevisiae. Yeast. 2003;20(4):351–67. doi:10.1002/yea.967.

Nagy ZK. Model based control of a yeast fermentation bioreactor using optimally designed artificial neural networks. Chemical Engineering Journal. 2007;127(1-3):95–109. doi:10.1016/j.cej.2006.10.015.

Najafpour GD. Biochemical engineering and biotechnology. 1st ed. Amsterdam, Oxford: Elsevier; 2007.

Nakano K, Rischke M, Sato S, Märkl H. Influence of acetic acid on the growth of Escherichia coli K12 during high-cell-density cultivation in a dialysis reactor. Appl Microbiol Biotechnol. 1997;48(5):597–601. doi:10.1007/s002530051101.

Nandy SK, Srivastava RK. A review on sustainable yeast biotechnological processes and applications. Microbiol Res. 2018;207:83–90. doi:10.1016/j.micres.2017.11.013.

Nelder JA, Mead R. A Simplex Method for Function Minimization. The Computer Journal. 1965;7(4):308–13. doi:10.1093/comjnl/7.4.308.

Neubauer P, Junne S. Scale-Up and Scale-Down Methodologies for Bioreactors. 2016:323–54. doi:10.1002/9783527683369.ch11.

Neuwirth E. RColorBrewer: ColorBrewer Palettes 2022.

Neves AR, Pool WA, Kok J, Kuipers OP, Santos H. Overview on sugar metabolism and its control in Lactococcus lactis - the input from in vivo NMR. FEMS Microbiol Rev. 2005;29(3):531–54. doi:10.1016/j.femsre.2005.04.005.

Nickel DB, Cruz-Bournazou MN, Wilms T, Neubauer P, Knepper A. Online bioprocess data generation, analysis, and optimization for parallel fed-batch fermentations in milliliter scale. Engineering in Life Sciences. 2017;17(11):1195–201. doi:10.1002/elsc.201600035.

Nielsen J. A simple morphologically structured model describing the growth of filamentous microorganisms. Biotechnol. Bioeng. 1993;41(7):715–27. doi:10.1002/bit.260410706.

Nielsen J, Nikolajsen K, Villadsen J. Structured modeling of a microbial system: I. A theoretical study of lactic acid fermentation. Biotechnol. Bioeng. 1991;38(1):1–10. doi:10.1002/bit.260380102.

Nienow AW. Scale-Up, Stirred Tank Reactors. Encyclopedia of Industrial Biotechnology: Bioprocess, Bioseparation, and Cell Technology. 2010:1–38. doi:10.1002/9780470054581.eib535.

Oliver SG, van der Aart QJ, Agostoni-Carbone ML, Aigle M, Alberghina L, Alexandraki D, et al. The complete DNA sequence of yeast chromosome III. Nature. 1992;357(6373):38–46. doi:10.1038/357038a0.

Olughu W, Deepika G, Hewitt C, Rielly C. Insight into the large-scale upstream fermentation environment using scaled-down models. J. Chem. Technol. Biotechnol. 2019;94(3):647–57. doi:10.1002/jctb.5804.

O'Neil DG, Lyberatos G. Dynamic model development for a continuous culture of Saccharomyces cerevisiae. Biotechnol. Bioeng. 1990;36(5):437–45. doi:10.1002/bit.260360502.

Onyeaka H, Nienow AW, Hewitt CJ. Further studies related to the scale-up of high cell density Escherichia coli fed-batch fermentations: the additional effect of a changing microenvironment when using aqueous ammonia to control pH. Biotechnol Bioeng. 2003;84(4):474–84. doi:10.1002/bit.10805.

Oosterhuis NM. Scale-Up of Bioreactors: A Scale-Down Approach; 1984.

Oosterhuis NM, Kossen NW. Dissolved oxygen concentration profiles in a production-scale bioreactor. Biotechnol. Bioeng. 1984;26(5):546–50. doi:10.1002/bit.260260522.

Oosterhuis NM, Kossen NW, Olivier AP, Schenk ES. Scale-down and optimization studies of the gluconic acid fermentation by Gluconobacter oxydans. Biotechnol. Bioeng. 1985;27(5):711–20. doi:10.1002/bit.260270521.

Palomares LA, Lara AR, Ramírez OT. Bioreactor Scale-Down. 2010;2010:1–15. doi:10.1002/9780470054581.eib142.

Papagianni M. Methodologies for Scale-down of Microbial Bioprocesses. J Microb Biochem Technol 2015. doi:10.4172/1948-5948.S5-001.

Papagianni M, Mattey M, Kristiansen B. Design of a tubular loop bioreactor for scale-up and scale-down of fermentation processes. Biotechnol Prog. 2003;19(5):1498–504. doi:10.1021/bp030002y.

Parapouli M, Vasileiadis A, Afendra A-S, Hatziloukas E. Saccharomyces cerevisiae and its industrial applications. AIMS Microbiol. 2020;6(1):1–31. doi:10.3934/microbiol.2020001.

Pigou M, Morchain J. Investigating the interactions between physical and biological heterogeneities in bioreactors using compartment, population balance and metabolic models. Chemical Engineering Science. 2015;126:267–82. doi:10.1016/j.ces.2014.11.035.

Pörtner R, Platas Barradas O, Frahm B, Hass VC. Advanced Process and Control Strategies for Bioreactors. In: Current Developments in Biotechnology and Bioengineering. Elsevier; 2017a. p. 463–493. doi:10.1016/B978-0-444-63663-8.00016-1.

Pörtner R, Platas Barradas O, Frahm B, Hass VC. Advanced Process and Control Strategies for Bioreactors. In: Pörtner R, Platas Barradas O, Frahm B, Hass VC, eds. Advanced Process and Control Strategies for Bioreactors. Elsevier; 2017b. p. 463–493. doi:10.1016/B978-0-444-63663-8.00016-1.

Poulsen BR, Iversen JJL. Mixing determinations in reactor vessels using linear buffers. Chem Eng Sci. 1997;52(6):979–84. doi:10.1016/S0009-2509(96)00466-6.

R Core Team. R: A Language and Environment for Statistical Computing 2023. Vienna, Austria.

Ramkrishna D. Statistical models of cell populations. Advances in Biochemical Engineering. 1979(11):1–47. doi:10.1007/3-540-08990-X_21.

Reiling HE, Laurila H, Fiechter A. Mass culture of Escherichia coli: Medium development for low and high density cultivation of Escherichia coli B/r in minimal and complex media. J Biotechnol. 1985;2(3-4):191–206. doi:10.1016/0168-1656(85)90038-0.

Rizzi M, Baltes M, Theobald U, Reuss M. In vivo analysis of metabolic dynamics inSaccharomyces cerevisiae: II. Mathematical model. Biotechnol. Bioeng. 1997;55(4):592–608. doi:10.1002/(SICI)1097-0290(19970820)55:4<592::AID-BIT2>3.0.CO;2-C.

Robinson A. Did Einstein really say that? Nature. 2018;557(7703):30. doi:10.1038/d41586-018-05004-4.

Rodrigues F, Ludovico P, Leão C. Sugar Metabolism in Yeasts: an Overview of Aerobic and Anaerobic Glucose Catabolism. Biodiversity and Ecophysiology of Yeasts. The Yeast Handbook. 2006:101–21. doi:10.1007/3-540-30985-3_6.

Roeva O, Pencheva T. Functional state modelling approach validation for yeast and bacteria cultivations. Biotechnol Biotechnol Equip. 2014;28(5):968–74. doi:10.1080/13102818.2014.934550.

Roeva O, Pencheva T, Tzonkov S, Arndt M, Hitzmann B, Kleist S, et al. Multiple model approach to modelling of Escherichia coli fed-batch cultivation extracellular production of bacterial phytase. Electron. J. Biotechnol. 2007;10(4):0. doi:10.2225/vol10-issue4-fulltext-5.

Ronen M, Shabtai Y, Guterman H. Hybrid model building methodology using unsupervised fuzzy clustering and supervised neural networks. Biotechnol. Bioeng. 2002a;77(4):420–9. doi:10.1002/bit.10132.

Ronen M, Shabtai Y, Guterman H. Optimization of feeding profile for a fed-batch bioreactor by an evolutionary algorithm. J Biotechnol. 2002b;97(3):253–63. doi:10.1016/S0168-1656(02)00106-2.

Rose AH. The Yeasts: Yeast Technology. 2nd ed. Burlington: Elsevier Science; 1993.

Runge C. Ueber die numerische Auflösung von Differentialgleichungen. Math Ann. 1895;46(2):167–78. doi:10.1007/BF01446807.

Sachs L, Hedderich J. Angewandte Statistik: Methodensammlung mit R ; mit 180 Tabellen. 12th ed. Berlin, Heidelberg: Springer; 2006.

Sandoval-Basurto EA, Gosset G, Bolívar F, Ramírez OT. Culture of Escherichia coli under dissolved oxygen gradients simulated in a two-compartment scale-down system: metabolic response and production of recombinant protein. Biotechnol Bioeng. 2005;89(4):453–63. doi:10.1002/bit.20383.

Santos LO, Dewasme L, Wouwer AV. Nonlinear model predictive control of fed-batch cultures of E. coli: performance and robustness analysis. IFAC Proceedings Volumes. 2010;43(14):1046–51. doi:10.3182/20100901-3-IT-2016.00224.

Santos LO, Dewasme L, Coutinho D, Wouwer AV. Nonlinear model predictive control of fed-batch cultures of micro-organisms exhibiting overflow metabolism: Assessment and robustness. Computers & Chemical Engineering. 2012;39:143–51. doi:10.1016/j.compchemeng.2011.12.010.

Sarkar J, Shekhawat LK, Loomba V, Rathore AS. CFD of mixing of multi-phase flow in a bioreactor using population balance model. Biotechnol Prog. 2016;32(3):613–28. doi:10.1002/btpr.2242.

Scarlat N, Dallemand J-F, Monforti-Ferrario F, Nita V. The role of biomass and bioenergy in a future bioeconomy: Policies and facts. Environmental Development. 2015;15:3–34. doi:10.1016/j.envdev.2015.03.006.

Schaich D, Becker R, King R. Qualitative modelling for automatic identification of mathematical models of chemical reaction systems. Control Engineering Practice. 2001;9(12):1373–81. doi:10.1016/S0967-0661(01)00080-6.

Schilling BM, Pfefferle W, Bachmann B, Leuchtenberger W, Deckwer W-D. A special reactor design for investigations of mixing time effects in a scaled-down industrialL-lysine fed-batch fermentation process. Biotechnol. Bioeng. 1999;64(5):599–606. doi:10.1002/(SICI)1097-0290(19990905)64:5<599::AID-BIT10>3.0.CO;2-C.

Schlüter J-P, Czuppon P, Schauer O, Pfaffelhuber P, McIntosh M, Becker A. Classification of phenotypic subpopulations in isogenic bacterial cultures by triple promoter probing at single cell level. J Biotechnol. 2015;198:3–14. doi:10.1016/j.jbiotec.2015.01.021.

Schmalzriedt S, Jenne M, Mauch K, Reuss M. Integration of physiology and fluid dynamics. Adv Biochem Eng Biotechnol. 2003;80:19–68. doi:10.1007/3-540-36782-9_2.

Schneider R, Hass VC, Munack A. OLFO Controller Performance Study Using Mathematical Fermentation Models of Different Complexity. IFAC Proceedings Volumes. 1993;26(2):229–32. doi:10.1016/S1474-6670(17)48720-9.

Schubert J, Simutis R, Dors M, Havlik I, Lübbert A. Bioprocess optimization and control: Application of hybrid modelling. J Biotechnol. 1994;35(1):51–68. doi:10.1016/0168-1656(94)90189-9.

Sercinoglu O, Platas Barradas O, Sandig V, Zeng A-P, Pörtner R. DoE of fed-batch processes - model-based design and experimental evaluation. BMC Proc. 2011;5 Suppl 8:P46. doi:10.1186/1753-6561-5-S8-P46.

Shi Z, Shimizu K. Neuro-fuzzy control of bioreactor systems with pattern recognition. Journal of Fermentation and Bioengineering. 1992;74(1):39–45. doi:10.1016/0922-338X(92)90265-V.

Shu P. Mathematical models for the product accumulation in microbiolgical processes. Biotechnol. Bioeng. 1961;3(1):95–109. doi:10.1002/jbmte.390030111.

Shuler ML, Leung S, Dick CC. A MATHEMATICAL MODEL FOR THE GROWTH OF A SINGLE BACTERIAL CELL. Ann NY Acad Sci. 1979;326(1 Biochemical E):35–52. doi:10.1111/j.1749-6632.1979.tb14150.x.

Sievert C. Interactive Web-Based Data Visualization with R, plotly, and shiny. Chapman and Hall/CRC; 2020.

Sijpesteijn AK. Induction of cytochrome formation and stimulation of oxidative dissimilation by hemin in Streptococcus lactis and Leuconostoc mesenteroides. Antonie Van Leeuwenhoek. 1970;36(3):335–48. doi:10.1007/BF02069035.

Sin G, Gernaey KV, Lantz AE. Good modeling practice for PAT applications: propagation of input uncertainty and sensitivity analysis. Biotechnol Prog. 2009;25(4):1043–53. doi:10.1002/btpr.166.

Sixt M, Uhlenbrock L, Strube J. Toward a Distinct and Quantitative Validation Method for Predictive Process Modelling—On the Example of Solid-Liquid Extraction Processes of Complex Plant Extracts. Processes. 2018;6(6):66. doi:10.3390/pr6060066.

Sonnleitner B, Käppeli O. Growth of Saccharomyces cerevisiae is controlled by its limited respiratory capacity: Formulation and verification of a hypothesis. Biotechnol Bioeng. 1986;28(6):927–37. doi:10.1002/bit.260280620.

Steel R, Maxon WD. Dissolved oxygen measurements in pilot- and production-scale novobiocin fermentations. Biotechnol. Bioeng. 1966;8(1):97–108. doi:10.1002/bit.260080109.

Stosch M von, Willis MJ. Intensified design of experiments for upstream bioreactors. Eng. Life Sci. 2017;17(11):1173–84. doi:10.1002/elsc.201600037.

Stosch M von, Oliveira R, Peres J, Feyo de Azevedo S. Hybrid semi-parametric modeling in process systems engineering: Past, present and future. Computers & Chemical Engineering. 2014;60:86–101. doi:10.1016/j.compchemeng.2013.08.008.

Sweere A, Luyben K, Kossen N. Regime analysis and scale-down: Tools to investigate the performance of bioreactors. Enzyme and Microbial Technology. 1987;9(7):386–98. doi:10.1016/0141-0229(87)90133-5.

Sweere APJ, Matla YA, Zandvliet J, Luyben KCAM, Kossen NWF. Experimental simulation of glucose fluctuations. Appl Microbiol Biotechnol. 1988;28(2):109–15. doi:10.1007/BF00694297.

Sweere APJ, Mesters JR, Janse L, Luyben KC, Kossen NW. Experimental simulation of oxygen profiles and their influence on baker's yeast production: I. One-fermentor system. Biotechnol. Bioeng. 1988;31(6):567–78. doi:10.1002/bit.260310609.

Tajsoleiman T, Mears L, Krühne U, Gernaey KV, Cornelissen S. An Industrial Perspective on Scale-Down Challenges Using Miniaturized Bioreactors. Trends Biotechnol. 2019;37(7):697–706. doi:10.1016/j.tibtech.2019.01.002.

Takors R. Scale-up of microbial processes: impacts, tools and open questions. J Biotechnol. 2012;160(1-2):3–9. doi:10.1016/j.jbiotec.2011.12.010.

Täuber S, Blöbaum L, Steier V, Oldiges M, Grünberger A. Microfluidic single-cell scale-down bioreactors: A proof-of-concept for the growth of Corynebacterium glutamicum at oscillating pH values. Biotechnol. Bioeng. 2022;119(11):3194–209. doi:10.1002/bit.28208.

Teissier G. Les lois quantitatives de la croissance. Annales de Physiologie et de Physiochimie Biologique. 1936;1936(12):527–73.

The MathWorks Inc. MATLAB R2018b 2018. Natick, Massachusetts, United States.

The pandas development team. pandas-dev/pandas: Pandas 2020. Zenodo: Zenodo. doi:10.5281/zenodo.3509134.

Theobald U, Mailinger W, Baltes M, Rizzi M, Reuss M. In vivo analysis of metabolic dynamics inSaccharomyces cerevisiae : I. Experimental observations. Biotechnol. Bioeng. 1997;55(2):305–16. doi:10.1002/(SICI)1097-0290(19970720)55:2<305::AID-BIT8>3.0.CO;2-M.

Thomas Williams, Colin Kelley. Gnuplot 5.4: an interactive plotting program 2021.

Thomson JM, Gaucher EA, Burgan MF, Kee DW de, Li T, Aris JP, Benner SA. Resurrecting ancestral alcohol dehydrogenases from yeast. Nat Genet. 2005;37(6):630–5. doi:10.1038/ng1553.

Torrez C, André C. Simulation of a Rushton Turbine Mixing Yield Stress Fluids: Application of the Metzner-Otto Concept. Chemical Engineering & Technology. 1999;22(8):701–6. doi:10.1002/(SICI)1521-4125(199908)22:8<701::AID-CEAT701>3.0.CO;2-L.

Tsuchiya HM, Fredrickson AG, Aris R. Dynamics of Microbial Cell Populations. In: Advances in Chemical Engineering Volume 6. Elsevier; 1966. p. 125–206. doi:10.1016/S0065-2377(08)60275-6.

Vallino JJ, Stephanopoulos G. Metabolic flux distributions in Corynebacterium glutamicum during growth and lysine overproduction. Biotechnol. Bioeng. 1993;41(6):633–46. doi:10.1002/bit.260410606.

van Barneveld J, Smit W, Oosterhuis NMG, Pragt HJ. Measuring the liquid circulation time in a large gas-liquid contactor by means of a radio pill. Part 1. Flow pattern and mean circulation time. Ind. Eng. Chem. Res. 1987a;26(11):2185–92. doi:10.1021/ie00071a003.

van Barneveld J, Smit W, Oosterhuis NMG, Pragt HJ. Measuring the liquid circulation time in a large gas-liquid contractor by means of a radio pill. Part 2. Circulation time distribution. Ind. Eng. Chem. Res. 1987b;26(11):2192–5. doi:10.1021/ie00071a004.

van Dijken JP, Weusthuis RA, Pronk JT. Kinetics of growth and sugar consumption in yeasts. Antonie Van Leeuwenhoek. 1993;63(3-4):343–52. doi:10.1007/BF00871229.

van Dijken JP, Bauer J, Brambilla L, Duboc P, Francois JM, Gancedo C, et al. An interlaboratory comparison of physiological and genetic properties of four Saccharomyces cerevisiae strains. Enzyme Microb Technol. 2000;26(9-10):706–14. doi:10.1002/(SICI)1097-0290(19970720)55:2<305::AID-BIT8>3.0.CO;2-M.

van Rossum G. The Python language reference. 3rd ed. Hampton, NH, Redwood City, Calif.: Python Software Foundation; SoHo Books; 2010.

van't Riet K, Tramper J. Basic Bioreactor Design. CRC Press; 1991.

Varma A, Palsson BO. Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. Biotechnology (N Y). 1994;12(10):994–8. doi:10.1038/nbt1094-994.

Vaseghi S, Baumeister A, Rizzi M, Reuss M. In vivo dynamics of the pentose phosphate pathway in Saccharomyces cerevisiae. Metab Eng. 1999;1(2):128–40. doi:10.1006/mben.1998.0110.

Venus J, Richter K. Development of a Pilot Plant Facility for the Conversion of Renewables in Biotechnological Processes. Engineering in Life Sciences. 2007;7(4):395–402. doi:10.1002/elsc.200720199.

Villadsen J, Nielsen J, Lidén G. Bioreaction Engineering Principles. Boston, MA: Springer US; 2011.

Vilums S, Kozlinskis E, Brusbardis V. Application of functional state modelling approach for yeast Sacharomyces cerevisiae fed-batch fermentation modelling: Tbilisi, Georgia, 17 - 19 October 2012. Piscataway, NJ: IEEE; 2012.

Visser D, van der Heijden R, Mauch K, Reuss M, Heijnen S. Tendency modeling: a new approach to obtain simplified kinetic models of metabolism applied to Saccharomyces cerevisiae. Metab Eng. 2000;2(3):252–75. doi:10.1006/mben.2000.0150.

Vrábel P, van der Lans R, Cui YQ, Luyben K. Compartment Model Approach. Chemical Engineering Research and Design. 1999;77(4):291–302. doi:10.1205/026387699526223.

Vrábel P, van der Lans RG, Luyben KC, Boon L, Nienow AW. Mixing in large-scale vessels stirred with multiple radial or radial and axial up-pumping impellers: modelling and measurements. Chem Eng Sci. 2000;55(23):5881–96. doi:10.1016/S0009-2509(00)00175-5.

Wang G, Tang W, Xia J, Chu J, Noorman H, van Gulik WM. Integration of microbial kinetics and fluid dynamics toward model-driven scale-up of industrial bioprocesses. Engineering in Life Sciences. 2015;15(1):20–9. doi:10.1002/elsc.201400172.

Wang R, Unrean P, Franzén CJ. Model-based optimization and scale-up of multi-feed simultaneous saccharification and co-fermentation of steam pre-treated lignocellulose enables high gravity ethanol production. Biotechnol Biofuels. 2016;9(1):88. doi:10.1186/s13068-016-0500-7.

Wang G, Haringa C, Noorman H, Chu J, Zhuang Y. Developing a Computational Framework To Advance Bioprocess Scale-Up. Trends Biotechnol. 2020;38(8):846–56. doi:10.1016/j.tibtech.2020.01.009.

Wechselberger P, Sagmeister P, Engelking H, Schmidt T, Wenger J, Herwig C. Efficient feeding profile optimization for recombinant protein production using physiological information. Bioprocess Biosyst Eng. 2012;35(9):1637–49. doi:10.1007/s00449-012-0754-9.

Wechselberger P, Sagmeister P, Herwig C. Model-based analysis on the extractability of information from data in dynamic fed-batch experiments. Biotechnol Prog. 2013;29(1):285–96. doi:10.1002/btpr.1649.

Weuster-Botz D. Parallel Reactor Systems for Bioprocess Development. In: Kragl U, Aivasidis A, eds. Technology transfer in biotechnology: From lab to industry to production. Berlin: Springer; 2005. p. 125–143. doi:10.1007/b98916.

Wickham H. Reshaping Data with the reshape Package. J Stat Softw. 2007;21(12):1–20.

Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016.

Wiechert W. 13C metabolic flux analysis. Metab Eng. 2001;3(3):195–206. doi:10.1006/mben.2001.0187.

Williams FM. A model of cell growth dynamics. Journal of Theoretical Biology. 1967;15(2):190–207. doi:10.1016/0022-5193(67)90200-7.

Witte VC. Mathematische Modellierung und adaptive Prozeßsteuerung der Kultivierung von Cyathus striatus. Düsseldorf: VDI-Verl.; 1996.

Woehrer W, Roehr M. Regulatory aspects of bakers' yeast metabolism in aerobic fed-batch cultures. Biotechnol. Bioeng. 1981;23(3):567–81. doi:10.1002/bit.260230308.

Yang X, editor. Scale-Up of Microbial Fermentation Process. 3rd ed. Washington, DC: ASM Press; 2007.

Yawalkar AA, Heesink ABM, Versteeg GF, Pangarkar VG. Gas-Liquid Mass Transfer Coefficient in Stirred Tank Reactors. Can. J. Chem. Eng. 2002;80(5):840–8. doi:10.1002/cjce.5450800507.

Ying H. Essentials of fuzzy modeling and control. J. Am. Soc. Inf. Sci. 1995;46(10):791–2. doi:10.1002/(SICI)1097-4571(199512)46:10<791::AID-ASI12>3.0.CO;2-H.

Zadeh LA. Fuzzy sets. Information and Control. 1965;8(3):338–53. doi:10.1016/S0019-9958(65)90241-X.

Zahradník J, Mann R, Fialová M, Vlaev D, Vlaev SD, Lossev V, Seichter P. A networks-of-zones analysis of mixing and mass transfer in three industrial bioreactors. Chem Eng Sci. 2001;56(2):485–92. doi:10.1016/S0009-2509(00)00252-9.

Zakhartsev M, Yang X, Reuss, Matthias, Pörtner, Hans Otto. Metabolic efficiency in yeast Saccharomyces cerevisiae in relation to temperature dependent growth and biomass yield. J Therm Biol. 2015;52:117–29. doi:10.1016/j.jtherbio.2015.05.008.

Zeigler BP, Weinberg R. System theoretic analysis of models: computer simulation of a living cell. J Theor Biol. 1970;29(1):35–56. doi:10.1016/0022-5193(70)90117-7.

Zeileis A, Grothendieck G. zoo: S3 Infrastructure for Regular and Irregular Time Series. J Stat Softw. 2005;14(6):1–27. doi:10.18637/jss.v014.i06.

Zhang X-C, Visala A, Halme A, Linko P. Functional state modeling and fuzzy control of fed-batch aerobic baker's yeast process. J Biotechnol. 1994;37(1):1–10. doi:10.1016/0168-1656(94)90196-1.

Zhang H, Lennox B. Integrated condition monitoring and control of fed-batch fermentation processes. Journal of Process Control. 2004;14(1):41–50. doi:10.1016/S0959-1524(03)00044-1.

Zobel-Roos S, Schmidt A, Mestmäcker F, Mouellef M, Huter M, Uhlenbrock L, et al. Accelerating Biologics Manufacturing by Modeling or: Is Approval under the QbD and PAT Approaches Demanded by Authorities Acceptable Without a Digital-Twin? Processes. 2019;7(2):94. doi:10.3390/pr7020094.

12 Appendix

12.1 Material and methods

12.1.1 Software libraries and packages

Table 20 gives an overview of the software libraries and packages that are used within the modelling environment C-eStIM.

Table 20: External software libraries used by C-eStIM.

Library	Usage	Developer & Link to Website
argtable2-13	Provides a command line parser	Stewart Heitmann, https://argtable.sourceforge.io/
boost_1_43_0	Container, for the shared memory management (shared_ptr).	Boost organisation, https://www.boost.org/
CppUnit_1_12_1	Systematic testing of software components (not necessary)	Baptiste Lepilleur, https://sourceforge.net/projects/cppunit/
fMinimize	Algorithms for the minimisation of functions	City University of Applied Sciences, Bremen based on work of Prof. Dr. Virginia Torczon, College of William and Mary, Williamsburg, Virginia, United States
galib247	Genetic algorithms for Minimisation of a scalar function of several variables	Matthew Wall, Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts, United States, http://lancet.mit.edu/ga/
Minuit2_5_22_00	Minimisation of a scalar function several variables, with particular attention to the Least Square's parameter estimation	Fred James and Matthias Winkler, CERN, Geneva, Swiss Confederation https://root.cern.ch/root/htmldoc/guides/minuit2/Minuit2.html
nTools	Multiple simple numeric algorithms	City University of Applied Sciences, Bremen
tnt_3_0_12	Translation of Fortran-code to C++	Roldan Pozo, National Institute of Standards and Technology Gaithersburg, Maryland, United States, math.nist.gov/tnt/
tnt_helper	Further algorithms. Addition to tnt_3_0_12	City University of Applied Sciences, Bremen

12.1.2 Specification of used computer hardware

Table 21 lists the computer hardware parts and specifications used for this work.

Table 21: List of the used PC hardware parts: The central processing unit (CPU), the graphics processing unit (GPU), the motherboard, dynamic random-access memory (DRAM), hard-disk-drive (HDD), solid-state-drive (SSD), and the power supply.

Part	Specification
CPU	AMD Ryzen 5 5600X
GPU	MSI GeForce GTX 960 Gaming 2G, 2GB GDDR5
Mainboard	Gigabyte B550 Aorus Pro V2 AMD B550 So.AM4 Dual Channel DDR4 ATX Retail
DRAM	32GB G. Skill Aegis DDR4-3200 DIMM CL16 Dual Kit
HDD	Seagate Barracuda 7200 2000GB, SATA 6Gb/s
SSD	1000GB Samsung 970 Evo Plus M.2 2280 PCIe 3.0 x4 NVMe 1.3 3D-NAND TLC (MZ-V7S1T0BW)
Power supply	Corsair CX Series Modular CX430M, 430W

12.2 Experiments

12.2.1 Parameter identifications

12.2.1.1 Parameter identification of single-bioreactor system

The parameters describing the maximum uptake rates under ideal conditions for glucose $r_{SC,mx,X}$, nitrogen source $r_{SN,mx,X}$ and ethanol $r_{PC1,mx,X,Di}$ are identified. These uptake rates for the main biomass compartment X are part of the Monod-like kinetics of the process model.

A total of four yield coefficients of metabolic pathways were identified which are yields of metabolic pathways instead of ratios of metabolites in the model. $y_{SC,X,1}$ describes the first split of the glucose metabolism; ratio of produced biomass & ethanol to energy formation through respiration. $y_{SC,X,2}$ is the yield coefficient for the second split between biomass and ethanol (for example if $y_{SC,X,2} = 1$, only biomass is formed). The last two selected coefficient

$y1_{PC1,X}$ describes the split between ethanol consumption for biomass growth and energy formation for the primary and structural biomass compartment. This coefficient is part of the respiration metabolism of ethanol and oxygen to carbon dioxide and water.

The parameters K_{act}, K_i and K_d describe the standard rates for activation, inactivation, and mortality of the biomass compartments. These rates are also modifiable by double sigmoid functions, like the Monod-like kinetics. The double-sigmoid function parameters of the influence of glucose limitation ($xl50_{Ki,SC}$ and $yl_{Ki,SC}$) and nitrogen-source inhibition ($xh50_{Ki,SN}$ and $yh_{Ki,SN}$) on the inactivation rate were also identified. If the glucose concentration falls closer to $xl50_{Ki,SC}$ the inactivation rate is increasing until being multiplied by $yl_{Ki,SC}$, if the nitrogen source concentration exceeds $xh50_{Ki,SN}$ the rate is increasing until being multiplied by $yh_{Ki,SN}$. The influence of the nitrogen source concentration on the inhibition rate was used by the model to simulate the transitions from exponential growth to the stationary and finally death phase.

The ethanol inhibition strength $y_{Ch,0,PC}$ and boundary concentration $h500_{PC}$ through modifying the glucose uptake rate was also identified. The last parameters ($h502_{SC}$ and $y_{Ch,2,SC}$) describe the sigmoid function which belongs to the Crabtree-effect.

Table 23 lists the model parameters selected for the parameter identification, their initial and identified values, the unit, and a brief description of each parameter.

Table 22: Chosen model parameters for the Monte-Carlo parameter uncertainty quantification and parameter identification, as well as their initial value and mean value after identification, unit, and a description.

Parameter	Initial value	Mean value	Unit	Description
$r_{SC,mx,X}$	5.25E-04	1.29E-04	s^{-1}	Uptake rate of glucose
$r_{SN,mx,X}$	5.33E-05	5.41E-05	s^{-1}	Uptake rate of nitrogen source
$r_{PC1,mx,X,Di}$	4.26E-05	1.36E-05	s^{-1}	Uptake rate of ethanol
$y_{SC,X,1}$	8.50E-01	7.26E-01	-	Yield coefficient glucose (split 1)
$y_{SC,X,2}$	1.10E-01	7.72E-02	-	Yield coefficient glucose (split 2)
$y1_{PC1,X}$	7.00E-01	2.01E-01	-	Yield coefficient ethanol (X)
$y1_{PC1,Xs}$	6.60E-01	4.28E-01	-	Yield coefficient ethanol (Xs)
K_{act}	1.30E-06	9.60E-05	s^{-1}	Activation rate
K_i	1.43E-05	1.08E-05	s^{-1}	Inactivation rate
K_d	1.31E-06	4.76E-06	s^{-1}	Mortality rate
$xh50_{Ki,SN}$	2.00E+01	2.09E+01	kg m^{-3}	Nitrogen source inhibition: -boundary
$yh_{Ki,SN}$	1.00E+00	1.01E+01	-	- strength
$xl50_{Ki,SC}$	1.00E-02	1.36E-02	kg m^{-3}	Glucose limitation: - boundary
$yl_{Ki,SC}$	1.00E+00	1.03E+00	-	- strength
$h500_{PC}$	5.80E+00	8.62E+01	-	Ethanol inhibition: - boundary
$y_{Ch,0,PC}$	1.00E+00	3.23E-01	-	- strength
$h502_{SC}$	8.50E+00	3.83E-01	kg m^{-3}	Overflow metabolism: - boundary
$y_{Ch,2,SC}$	1.00E+00	5.06E-02	-	- strength

The identified mean nitrogen concentration limit ($xh50_{Ki,SN}$) above which nitrogen inhibition occurs was identified at 20.9 g L^{-1}. This concentration is only reached at the end of the processes. The nitrogen concentration is low in the initial medium and gradually increases during the process due to feeding. When exceeding $xh50_{Ki,SN}$, the inhibition constant is increased up to tenfold. This limit would be too low and unrealistic when simulating batch processes, where the initial nitrogen source concentration would be higher than 20 g L^{-1}.

The mean glucose concentration limit for the occurrence of overflow metabolism is identified at about 0.39 g L^{-1}. Above that glucose concentration the Crabtree effect of glucose to ethanol occurs, which is in the range of literature values (Woehrer and Roehr 1981; Crabtree 1929; Deken 1966; Rose 1993).

The identified mean boundary concentration $h500_{PC}$ concentration is about 86 g L^{-1}. Since neither the experimental nor the simulation reach that concentration, ethanol inhibition was

not identified in the simulations. The final identified value for the ethanol inhibition strength $y_{Ch,0,PC}$ was therefore not relevant.

12.2.1.2 Parameter identification for model validation

A total of 27 model parameters were selected for the identification in chapter 8.3.1. As in chapter 7.1.2 (chapter 12.2.1.1 in the appendix), the parameters describing the maximum uptake rates under ideal conditions for glucose $r_{SC,mx,X}$, nitrogen source $r_{SN,mx,X}$ and ethanol $r_{PC1,mx,X,Di}$ are identified. These uptake rates for the main biomass compartment X are part of the Monod-like kinetics of the process model.

A total of three yield coefficients of metabolic pathways were identified which are yields of metabolic pathways instead of ratios of metabolites in the model. $y_{SC,X,1}$ describes the first split of the glucose metabolism; ratio of produced biomass & ethanol to energy formation through respiration. $y_{SC,X,2}$ is the yield coefficient for the second split between biomass and ethanol (for example if $y_{SC,X,2} = 1$, only biomass is formed).

In addition to the four yield coefficients identified in chapter 7.1.2 ($y_{SC,X,1}$, $y_{SC,X,2}$, $y1_{PC1,X}$, and $y1_{PC1,Xs}$; description in chapter 12.2.1.1), two additional yield coefficients were identified ($y_{SC,Xs,1}$ and $y_{SC,X,2,an}$). $y_{SC,Xs,1}$ describes the first split of the glucose metabolism for the structural biomass compartment; ratio of produced biomass & ethanol to energy formation through respiration. $y_{SC,X,2,an}$ describes the second split between biomass and ethanol under anaerobic conditions. This coefficient is important since processes and a heterogeneous oxygen concentration are part of the experimental data in chapter 8.

The parameters K_{act}, K_i and K_d are again identified. They describe the standard rates for activation, inactivation, and mortality of the biomass compartments.

Ethanol concentration of over 50 g L^{-1} are in the experimental data utilised for the identification of the model parameters. At this concentrations inhibition and even an increased death rate might occur for *S. cerevisiae*. For this reason, the parameter of two sigmoid functions were identified. The first function describes the influence of the ethanol concentration on the death rate K_d. $xh50_{Kd,PC}$ describes the boundary concentration above which the death rate is influenced, the value of $yh_{Kd,PC}$ is the multiplication factor which modifies the standard mortality rate, and $K_{Sl,Kd,PC}$ is the slope of the transition from uninfluenced to the higher death rate. The second sigmoid function describes the influence of the ethanol concentration on the first carbon yield coefficient. When ethanol concentration is above $xh50_{1,PC}$, the first split of carbon substrate to biomass and ethanol or energy is influenced by factor $yh_{1,PC}$ and the slope $K_{Sl,1,PC}$. This function results in a less efficient metabolism of the yeasts when ethanol gets close the boundary limit $xh50_{1,PC}$.

Furthermore, the influence of the ethanol concentration on the carbon substrate rate is identified as in the previous chapter (chapter 7.1.2 and 12.2.1.1). The identified parameters of this sigmoid function are the boundary concentration $h500_{PC}$, the multiplied for the modification of the uptake rate $y_{Ch,0,PC}$ and the slope $K_{Sl,0,PC}$.

The next parameters describe a sigmoid function which describes the influence of low carbon substrate concentrations (glucose concentration) on the first metabolic split of the carbon metabolism (biomass and ethanol or energy).

When glucose decreases to the boundary concentration of $xl50_{1,SC}$, the yield coefficient $y_{SC,X,1}$ is multiplied up to factor $yl_{1,SC}$ (slope $K_{Sl,1,SC}$). This function results in a potentially less efficient metabolism of the yeasts.

Last two parameter describe the multiplicator for the change of the carbon substrate uptake rate under anaerobic conditions $yl_{Cl0,PO}$ and the change of yield (split 2) under anaerobic conditions $yl_{Cl2,PO}$. These two parameters were selected since experiments in the data set are performed under anaerobic conditions and mixed oxygen conditions.

Table 23 lists the parameter name of the model parameter selected for the parameter identification, their initial and identified values, the unit, and a brief description of each parameter.

Table 23: Chosen parameters for the parameter identification of the SDM experiments, as well as their initial value and value after identification, unit, and description.

Parameter	Initial value	Final value	Unit	Description
$r_{SC,mx,X}$	4.00E-04	3.83E-04	s^{-1}	Uptake rate of glucose
$r_{SN,mx,X}$	7.60E-05	7.22E-05	s^{-1}	Uptake rate of nitrogen source
$r_{PC1,mx,X,Di}$	6.26E-05	1.06E-05	s^{-1}	Uptake rate of ethanol
$y_{SC,X,1}$	8.00E-01	9.90E-01	-	Yield coefficient SC&X / energy (split 1)
$y_{SC,X,2}$	6.50E-01	9.85E-01	-	Yield coefficient SC / X (split 2)
$y_{SC,Xs,1}$	9.50E-01	9.89E-01	-	Yield coefficient SC&Xs / energy (split 1)
$y_{SC,X,2,an}$	5.00E-01	8.51E-01	-	Yield coefficient anaerobic SC / X (split 2)
$y1_{PC1,X}$	2.00E-01	9.89E-01	-	Yield coefficient ethanol (X)
$y1_{PC1,Xs}$	2.00E-01	3.00E-01	-	Yield coefficient ethanol (Xs)
K_{act}	1.30E-06	1.23E-05	s^{-1}	Activation rate
K_i	1.48E-05	2.41E-05	s^{-1}	Inactivation rate
K_d	1.31E-07	1.80E-06	s^{-1}	Mortality rate
$xh50_{Kd,PC}$	1.80E+02	2.68E+01	kg m^{-3}	Ethanol increased death rate: - boundary
$yh_{Kd,PC}$	1.00E+01	1.72E+02	-	- strength
$K_{Sl,Kd,PC}$	1.00E-01	4.50E-01	-	- slope
$h500_{PC}$	7.00E+01	3.19E+01	kg m^{-3}	Change of SC uptake rate based on PC (ethanol inhibition): - boundary
$y_{Ch,0,PC}$	1.00E+00	1.00E+00	-	- strength
$K_{Sl,0,PC}$	5.05E+01	6.86E+01	-	- slope
$xl50_{1,SC}$	1.55E-01	3.22E+00	kg m^{-3}	Change of yield coefficient (split 1) depending on SC: -boundary
$yl_{1,SC}$	1.00E+00	8.41E-01	-	- strength
$K_{Sl,1,SC}$	3.0E+01	2.79E+01	-	- slope
$xh50_{1,PC}$	1.00E+00	7.71E+01	kg m^{-3}	Change of yield coefficient (split 1) depending on PC: - boundary
$yh_{1,PC}$	1.00E+00	3.57E-01	-	- strength
$K_{Sl,1,PC}$	1.00E+00	5.04E+00	-	- slope
$yl_{Cl0,PO}$	1.08E+00	8.51E-01	-	Change of SC uptake rate under anaerobic conditions
$yl_{Cl2,PO}$	3.00E-01	2.15E-01	-	Change of yield (split 2) under anaerobic conditions

$xh50_{1,PC}$ was identified at 77 g L^{-1} which is more than 20 g L^{-1} higher than the highest determined ethanol concentration in the data set. This means that the ethanol concentration has no identified influence on the yield coefficient of the first metabolic carbon substrate split.

12.2.1.3 Parameter identification in case study

A total of 25 parameters were selected for the identification during industrial process application with an industrial medium (chapter 9.3.2). These 25 parameters are identical to the parameters of the previous identification (chapters 8.3.1 and 12.2.1.2) without $yl_{Cl0,PO}$ and $yl_{Cl2,PO}$ since no anaerobic process conditions were employed in the processes of chapter 9

Table 24: Chosen parameters for the parameter identification of the case study during the application of the NMPC, their initial and final value after the last identification and a description.

Parameter	Initial value	Final value	Unit	Description
$r_{SC,mx,X}$	3.83E-04	8.64E-04	s^{-1}	Uptake rate of glucose
$r_{SN,mx,X}$	7.22E-05	1.07E-04	s^{-1}	Uptake rate of nitrogen source
$r_{PC1,mx,X,Di}$	1.06E-05	2.70E-05	s^{-1}	Uptake rate of ethanol
$y_{SC,X,1}$	9.90E-01	8.88E-01	-	Yield coefficient SC&X / energy (split 1)
$y_{SC,X,2}$	9.85E-01	1.00E+00	-	Yield coefficient SC / X (split 2)
$y_{SC,Xs,1}$	9.89E-01	9.99E-01	-	Yield coefficient SC&Xs / energy (split 1)
$y_{SC,X,2,an}$	8.51E-01	1.00E+00	-	Yield coefficient anaerobic SC / X (split 2)
$y1_{PC1,X}$	9.89E-01	8.62E-01	-	Yield coefficient ethanol (X)
$y1_{PC1,Xs}$	3.00E-01	1.81E-02	-	Yield coefficient ethanol (Xs)
K_{act}	1.23E-05	0.00E+00	s^{-1}	Activation rate
K_i	2.41E-05	4.96E-05	s^{-1}	Inactivation rate
K_d	1.80E-06	8.20E-07	s^{-1}	Mortality rate
$xh50_{Kd,PC}$	2.68E+01	2.19E+01	kg m^{-3}	Ethanol increased death rate: - boundary
$yh_{Kd,PC}$	1.72E+02	2.10E+02	-	- strength
$K_{Sl,Kd,PC}$	4.50E-01	4.35E-01	-	- slope
$h500_{PC}$	3.19E+01	5.95E+01	kg m^{-3}	Change of SC uptake rate based on PC (ethanol inhibition): - boundary
$y_{Ch,0,PC}$	1.00E+00	9.95E-01	-	- strength
$K_{Sl,0,PC}$	6.86E+01	3.45E+00	-	- slope
$xl50_{1,SC}$	3.22E+00	7.48E-01	kg m^{-3}	Change of yield coefficient (split 1) depending on SC: -boundary
$yl_{1,SC}$	8.41E-01	9.40E-01	-	- strength
$K_{Sl,1,SC}$	2.79E+01	1.19E+02	-	- slope
$xl50_{1,PC}$	7.71E+01	7.45E+01	kg m^{-3}	Change of yield coefficient (split 1) depending on PC: - boundary
$yl_{1,PC}$	3.57E-01	6.16E-01	-	- strength
$K_{Sl,1,PC}$	5.04E+00	8.93E-01	-	- slope

12.2.2 Further notes for each experiment

Each experiment performed is presented in this work and was used as sources for data and process information (except process D6, see Table 26), even if the deviations from the planned process operation are significant. This is a great advantage of models, since those

differences can be implemented, and the data still used. The following Table 25 (process C1-C5),Table 26 (process D1-D6), Table 27 (process E1&E2) and Table 28 (process F1-F8) are listing deviations from the planned process operation, performed changes and compensations as well as possible explanations for these deviations. For most deviations no or only small compensations were performed, either because not influence is expected, or the deviation can be considered in the simulations. In cases, when compensations are performed, they are described in the following tables.

Table 25: Deviations from planned process operation for SDM-characterisation cultivations with batch operation (C1-C5)

Exp.	Explanation of the deviation from planned process operation
C1	• pH probe in reactor 2 malfunctioned; no pH measurement possible. • Exhaust gas membrane filter of reactor 1 was block from 4.5 h until 5.5 h resulting in lower CO_2 and higher O_2 concentrations
C2	• Ethanol concentrations at 3.5 h and 4.5 h had to be determined again, because of impossible values in the first set of determinations
C3	• Ethanol concentrations after 3 h had to be determined again, because of impossible values in the first set of determinations
C5	• Exhaust gas membrane filter was starting to block from at 6 h until 8 h resulting in lower CO_2 and higher O_2 concentrations. The slightly higher pressure in the reactor had a positive influence on the DO.

Table 26: Deviations from planned process operation for SDM-characterisation cultivations with fed-batch operation (D1-D7)

Exp.	Explanation of the deviation from planned process operation
D1	• Exhaust gas membrane filter was blocked starting at 7 h until 9.5 h
D2	• Initial dry cell weight density concentration was determined to low since dry yeast was stuck to the reactor wall and over time dissipated into the volume. To compensate this, the initial dry cell weight density was set at 10 g L^{-1}, which is the planned density. • The exhaust gas membrane filter was blocked at 10 h throughout the next 14 hours which leads to: • Aerobic reactor emptied because of the blocked membrane and resulting pressure in aerobic reactor; total volume in anaerobic reactor. • Air coming through circulation pump pipe; was not resulting in increasing DO. • Process extended by 4 hours to get enough values to calculate ethanol consumption rate and biomass growth. • The pH probe in reactor 2 malfunctioned and was not measuring.
D3	• DO is dropping to 0% indicating oxygen limitation, but not fully anaerobic conditions since ethanol was still consumed
D4	• DO probe malfunctioned, offline data indicate that the process was fully anaerobic; DO was set to 0.0 in visualisations
D5	• Execution as planned.
D6	• This experiment is the only experiment that was performed for this work but is not used nor presented. • It was planned as a comparison experiment of D5 with same feeding rates, aeration per volume etc. in a 20 L Biostat C bioreactor. • The ethanol production is unrealistically high during the first feed resulting into a yield coefficient $Y_{X/SC}$ of above 1.4, which is higher than stoichiometrically and physically possible. This was observed during cultivation with contamination, when the determination method for the dry cell weight concentration is producing false values. • The mathematical model would not be able to simulate these processes since stoichiometric equations are implemented into the biological submodel

Table 27: Deviations from planned process operation for SDM-characterisation cultivations with fed-batch operation (D1-D7)

Exp.	Comment
E1	• Execution as planned.
E2	• Exhaust gas membrane filter was blocked around 10 h • Fourth ethanol concentration was determined a second time

Table 28: Deviations from planned process operation for cast study cultivations with fed-batch operation (F1-F8)

Exp.	Explanation of the deviation from planned process operation
F1	• Setting on syringe feed pump were wrong initially. Adapted twice the feeding to feed the correct total amount, resulting ultimately in a higher peak feeding rate
F2	• Reactor was not airtight during experiment, therefore the exhaust gas showing data for ambient air.
F3	• One falsely determined ethanol concentration (five-times higher than other values) was removed from data used for the graphical visualisation and parameter identification.
F4	• Reactor not completely airtight. Exhaust gas volume varying throughout the experiment. During experiment outage of exhaust gas measurement because air volume was too low for equipment.
F5	• Reactor was not completely airtight, resulting in lower exhaust gas flows to the gas analyser. The gas analyser did not always receive necessary volume flows for operation.
F6	• Not enough air flow through gas analyser at around 10 h and between 18 h and 26 h.
F7	• Air bubble in front of DO sensor resulting in higher measured DO between 6 h and 9 h.
F8	• Exhaust gas measurement of the 2nd reactor malfunctioned.

12.2.3 Mixing time calculation

The method for the calculation of the mixing time is described in chapter 4.3.3.2. Equation (4.8) for calculating the Newton number is only valid under turbulent conditions when the Reynolds-Number (Re) is greater than 10^4 and the Froude-Number (Fr) lower than 0.07 $(D/d)^3$ (Henzler 1982). The reactor has a maximum volume of 2 m³ and a height to diameter ratio H/D of approximately 3, resulting in H = 2.95 m and D = 0.95 m. The density of the nitrogen medium in the bioreactor was 1017 kg m^{-3}. The density was assumed to be constant throughout the process. The dynamic viscosity was assumed to be independent from cell growth and that of water at 30 °C at approximately 0.8 mPa s. The aeration rate was set at 100 $m^3\ h^{-1}$ and the volume 1 m³. The Reynolds number is calculated for the lowest stirrer speed at 99 rpm and Froude number for the highest stirrer speed at 199 rpm (worst case scenarios):

$$Re = \frac{(0.25\ m)^2 \cdot {}^{99}/_{60\ s} \cdot 1017\ {}^{kg}/_{m^3}}{0.8 \cdot 10^{-3}\ {}^{kg}/_{m \cdot s}} = 1.31 \cdot 10^5 \tag{12.1}$$

$$Fr = \frac{\left({}^{199}/_{60\ s}\right)^2 \cdot 0.25\ m}{9.81\ {}^{m}/_{s^2}} = 0.28 \tag{12.2}$$

$$with\ \ 0.07 \cdot \left(\frac{0.95\ m}{0.25\ m}\right)^3 = 3.84 > 0.28 \tag{12.3}$$

The stirrer has a diameter of approximately 0.25 m resulting in a stirrer diameter to reactor diameter ratio (d/D) of 0.25. Furthermore, there are two stirrer levels in the reactor at a approximate distance of 1 m resulting in a sufficient $\Delta h/D$ ratio of 1.05.

Since all conditions are for the application of equation (4.8) are satisfied, the aeration number and Newton number were calculated with the average stirrer speed of 149 rpm and with Ne_0 ≈ 4.9.

$$Q = \frac{{}^{100\ m^3}/_{3600\ s}}{(0.25\ m)^3 \cdot \left({}^{149}/_{60\ s}\right)} = 0.72 \tag{12.4}$$

$$Ne = 2 \cdot \frac{4.9 + 187 \cdot 0.72 \cdot 0.28^{-0.32} \cdot \left(\frac{0.25\ m}{0.95\ m}\right)^{1.53} - 4.6 \cdot 0.72^{1.25}}{1 + 136 \cdot 0.72 \cdot \left(\frac{0.25\ m}{0.95\ m}\right)^{1.14}} = 2.51 \tag{12.5}$$

The Newton number can subsequently be used to calculate the power input P (equation (4.7)), the specific average energy dissipation $\bar{\varepsilon}_T$ and lastly we get a rough approximation of the mixing time t_M with H_L of 1.41 m.

$$P = 2.51 \cdot 1017\,{}^{kg}/_{m^3} \cdot \left({}^{149}/_{60\ s}\right)^3 \cdot (0.25\ m)^5 = 19.1\ W \tag{12.6}$$

$$\bar{\varepsilon}_T = \frac{19.1\ W}{1\ m^2 \cdot 1017\,{}^{kg}/_{m^3}} = 0.019\ \frac{m^3}{s^3} \tag{12.7}$$

$$t_M = 6 \cdot 0.95\ m\ \cdot \left(0.019\ \frac{m^3}{s^3}\right)^{-\frac{1}{3}} \cdot \left(\frac{0.25\ m}{0.95\ m}\right)^{-\frac{1}{3}} \cdot \left(\frac{1.41\ m}{0.95\ m}\right)^{2.5} = 71\ s \tag{12.8}$$

This resulting approximation of the mixing time is 71 s. This time was set und utilised as the mixing time in the following SDM experiments. This calculation method using dimensionless key figures is of course speculative and relatively imprecise. However, the calculated mixing time of 71s is within a realistic range for the pilot reactor.

12.2.4 Mixing times investigation of chapter 9.6

Three simulations at different mixing times at 550 s, 280 s, and 70 s for the model-assisted investigation of the scale-dependent effects.

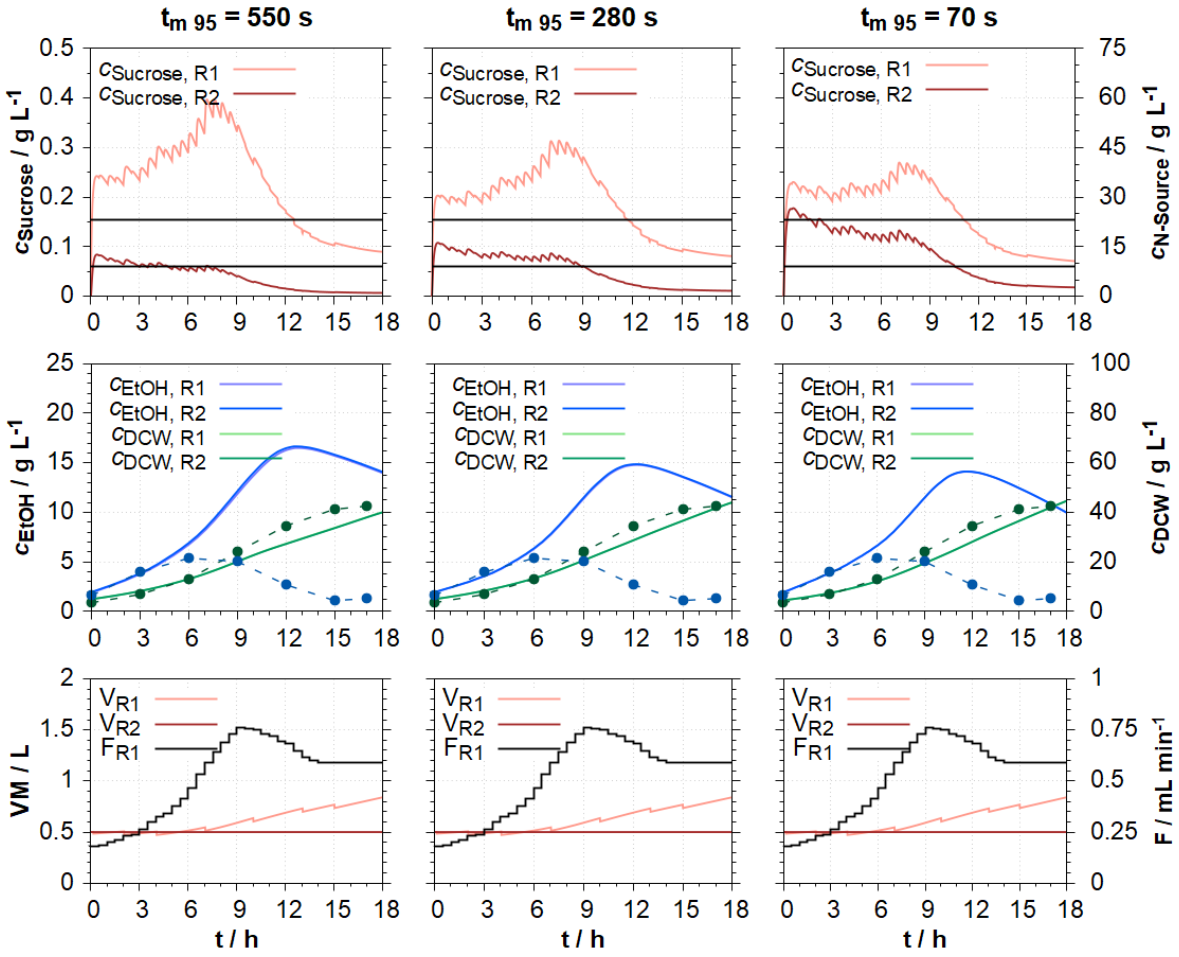


Figure 83: Three simulations of the pilot reactor process. 1st column: $t_{M,90}$ = 550 s; 2nd column; $t_{M,90}$ = 280 s; 3rd column: $t_{M,90}$ = 70 s. 1st row: Offline data and simulated data for the sucrose and nitrogen concentration. The boundaries for ethanol production (upper line) and consumption (lower line) are marked in grey; 2nd row: Offline data and simulated data for biomass density and ethanol concentration; 3rd row: Calculated and measured volume and feeding rate of the simulations.

The difference in sucrose concentrations between the two reactors decreases with shorter mixing times. However, the maximum ethanol concentration also decreases with shorter mixing times, which is different from the previous simulations. This was already indicated by the experimental investigation since the single bioreactor experiments (close to ideally mixed) produced less ethanol than the experiments in the SDM. A longer mixing time therefore does not fulfil the purpose of lowering the simulated ethanol concentration and bringing the process closer to the pilot scale.

List of publications

Publications:

Moser A; Appl C; Pörtner R; Baganz F; Hass, Hass VC (2024): A New Concept for the Rapid Development of Digital Twin Core Models for Bioprocesses in Various Reactor Designs. In: *Fermentation* 10 (9), S. 463. DOI: 10.3390/fermentation10090463.

Gassenmeier V, Deppe S, Hernández Rodríguez T, Kuhfuß F, Moser A, Hass VC, Kuchemüller KB, Pörtner R, Möller J, Ifrim G, Frahm B (2022) Model-assisted DoE applied to microalgae processes. Current Research in Biotechnology 4:102–118. https://doi.org/10.1016/j.crbiot.2022.01.005

Moser A, Kuchemüller KB, Deppe S, Hernández Rodríguez T, Frahm B, Pörtner R, Hass VC, Möller J (2021) Model-assisted DoE software: optimization of growth and biocatalysis in Saccharomyces cerevisiae bioprocesses. Bioprocess Eng. https://doi.org/10.1007/s00449-020-02478-3

Appl C, Moser A, Baganz F, Hass VC (2020) Digital Twins for Bioprocess Control Strategy Development and Realisation. In: *Adv. Biochem. Eng. Biotechnol.* 177, S. 63–94. DOI: 10.1007/10_2020_151.

Moser A, Appl C, Brüning S, Hass VC (2020) Mechanistic Mathematical Models as a Basis for Digital Twins. Adv Biochem Eng Biotechnol. https://doi.org/10.1007/10_2020_152

Presentations:

Moser A, Pörtner R, Hass VC (2023) Investigating the effects of bioreactor inhomogeneities on cell growth and productivity using a combined physical and mathematical model approach, 14th European Congress of Chemical Engineering and 7th European Congress of Applied Biotechnology, Berlin, Germany

Moser A, Appl C, Pörtner R, Hass VC (2022) Development tool for rapid, user-defined design of Digital Twins for different biotechnological processes, 26th International Congress of Chemical and Process Engineering, Prague, Czech Republic

Moser A, Pörtner R, Hass VC (2022) Investigating the effects of bioreactor inhomogeneities on cell growth and productivity using a combined physical and mathematical model approach, Himmelfahrtstagung on Bioprocess Engineering 2022 - Future Bioprocesses for a Sustainable Industry, Mainz, Germany

Moser A, Hass VC (2021) Development of a new NMPC for the design and optimization of dynamic bioeconomic processes, 13th European Congress of Chemical Engineering and 6th European Congress of Applied Biotechnology, Virtual Conference

Moser A, Pörtner R, Hass VC (2021) Model-based optimization of small scale bioreactors as a tool for scale up of biotechnological processes, 13th European Symposium on Biochemical Engineering Sciences (ESBES 2021), Virtual Conference

Moser A, Möller J, Kuchemüller KB, Deppe S, Hernández Rodríguez T, Gassenmeier V, Ifrim G, Frahm B, Pörtner R, Hass VC, (2021) Model-assisted Design of Experiments - Concept, Software and Application, International Congress of Chemical and Process Engineering (CHISA 2021 Virtually)

Poster presentations:

Moser A, Appl C, Pörtner R, Hass VC (2021) Modelling and optimizing small-scale parallelized bioreactors as a tool for scale-up of biotechnological processes at CHISA 2021; 24th International Congress of Chemical and Process Engineering, Prague, Czech Republic (awarded with poster price)

Relevant supervised student theses and projects:

Osdoba M, (2023) Entwicklung eines mechanistischen mathematischen pH-Teilmodells für die Verbesserung des Bäckerhefeprozesses, Master-Thesis

Bail G, (2023) Modellgestützte Optimierung eines Scale-Down-Prozesses in gekoppelten Reaktoren, Bachelor-Thesis

Maier D, (2022) Modellgestützte Optimierung des Bäckerhefeprozesses im Scale-Down Reaktorsystem, Master-Thesis

Maier D, (2021) Modellgestützte Charakterisierung eines Scale Down Reaktorsystems, Project work

Menzel J, (2021) Charakterisierung von Bioreaktoren bezüglich Sauerstoffübergang und Wärmetransport, Master-Thesis

Beltram B, (2020) Erstellung eines mathematischen Modells der Fermentation mit Saccharomyces cerevisiae, Project work

Meissner M, (2020) Modellbasierte, statistische Versuchsplanung zur effizienten Prozessführung von Hefekultivierungen, Bachelor-Thesis

Jasoliya H, (2020) Measurements of pO_2 during cultivations of *S. cerevisiae*, Project work

Ntanko Tainkeu S, (2019) Modellbasierte Experiment für die Kultivierung *von Saccharomyces cerevisiae*, Bachelor-Thesis

Gockel V, (2019) Anwendung von modellgestütztem Design of Experiments (mDoE), Project work