



TECHNISCHE  
UNIVERSITÄT  
WIEN  
Vienna | Austria

Dissertation

# Robust and adaptive mechanistic modelling in bioprocessing

Submitted in satisfaction of the requirements for the degree of Doctor of Technical  
Sciences

in the TU Wien, Faculty of Chemical Engineering

---

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines  
Doktors der technischen Wissenschaften  
eingereicht an der Technischen Universität Wien, Fakultät für Technische Chemie

von

**Adnan Jouned**

Matr.Nr.: 01529935

unter der Anleitung von

Prof. Dr.techn. Dipl.-Ing. **Christoph Herwig**

Institut für Verfahrenstechnik, Umwelttechnik und Technische Biowissenschaften  
Getreidemarkt 9  
1060 Wien  
Austria

# Robust and adaptive mechanistic modelling in bioprocessing

# Acknowledgement

This work would not have been possible without the support of many persons. All thanks to Dr. Tilman Barz, who did not spare the assistance and support. I have to thank Dr. Julian Kager and Prof. Christoph Herwig, for their great help and motivation along the project, and Dr. Thomas Fleckl, who have been very supportive. Nobody has been more important to me in the pursuit of this work than my loving wife DI. Nada Eidi, whose love and patience are with me in whatever I pursue.

# Zusammenfassung

Mechanistische Modelle spielen eine wesentliche Rolle bei der Entwicklung von Bioprozessen. Trotz ihrer rigorosen Struktur beschreiben sie die Prozesse mit interpretierbaren Modellparametern und liefern eine mathematische Darstellung der zugrunde liegenden Dynamik. Aus diesem Grund werden sie in großem Umfang bei der experimentellen Planung, Überwachung und Steuerung von Prozessen eingesetzt.

Allerdings stehen ihrer effektiven Nutzung immer noch viele Hindernisse im Wege, Modelle werden je nach ihren Zielen konzipiert. In einem industriellen Kontext werden Modelle mit Vereinfachungen erstellt, von denen einige zu diskontinuierlichen Modellen führen. Dieses Problem sowie die Unzulänglichkeiten und Besonderheiten der Analytik, wie z. B. die Abgas- und Gelöstsauerstoff Signale, und die Unterschiede in den Prozessbedingungen, wie z. B. das Arbeitsvolumen, können zu nicht adaptiven (unflexible Strukturen) und nicht robusten (unzuverlässiges Output) Modellen führen. Außerdem müssen die Modelle ähnlich wie die Prozesse, die sie beschreiben, im Laufe des Entwicklungszyklus angepasst werden. In dieser Arbeit wird die Hypothese aufgestellt, dass diese Herausforderungen durch geeignete wissenschaftliche Methoden gelöst werden können.

Die These zielt darauf ab, adaptive, robuste Modelle zu entwickeln, indem Methoden vorgeschlagen werden, um: die reduzierten Vorhersagefähigkeiten der diskontinuierlichen Modelle zu überwinden, (latente) ungenutzte Informationen aus bereits existierenden Analysen zu extrahieren, Besonderheiten der Analysen angemessen zu berücksichtigen und die Übertragbarkeit zwischen verschiedenen Skalen zu erleichtern.

Die einzelnen Ergebnisse dieser Arbeit lassen sich wie folgt zusammenfassen: Eine Methode, die Dynamic Time Warping (DTW) und die Least-Squares-Algorithmen (LSQ) kombiniert, wurde als Anpassungskriterium für die Kalibrierung von Hefekultivierungsmodellen entwickelt, die Abgasmessungen enthalten. Die Methode führte zu repräsentativeren, unverzerrten Parameterschätzungen. Eine vergleichende Analyse von Hefekultivierungsmodellen mit und ohne explizite Berücksichtigung der plötzlichen Änderungen des Reaktorvolumens ergab eine spürbare Auswirkung auf die Modellergebnisse und die Parameterunsicherheit aufgrund der Fehlerfortpflanzung. Es wird eine Event Driven Methode (EDM) zur angemessenen Berücksichtigung von Volumenänderungen vorgeschlagen. Eine vergleichende Studie zur Untersuchung aller möglichen diskontinuierlichen Verhaltensweisen in Bioprozessmodellen wird für Hefekultivierungen vorgestellt. Die Ergebnisse zeigen im Detail die wichtigsten Arten von Diskontinuitäten. Es wird ein EDM-Workflow vorgeschlagen, um diese angemessen zu behandeln, sowie ein allgemeiner Ansatz zur Umwandlung mechanistischer Modelle in eine matrixartige Form unter Verwendung eines Entscheidungsbaums und boolescher Bedingungsfunktionen. Der Workflow führte zu robusten Modellvorhersagen und geringer Parameterunsicherheit. Es wurde eine Methode zur Segmentierung des Signals der gelösten Sauerstoff (Dissolved Oxygen Tension DOT) für die Kultivierung von *E. coli* in Minibioreaktorsystemen (MBRs) entwickelt. Die identifizierten Segmente ermöglichten die Extraktion von diskreten Merkmalen. Die datengesteuerte Analyse der Merkmale ergab ein Muster von Reaktionen,

insbesondere sind DOT-Segmente mit bestimmten Stoffwechszuständen verbunden. Eine modellbasierte Analyse ermöglichte die Schätzung der wichtigsten Modellparameter für die Umschaltbedingung des Overflow-Metabolismus. Außerdem wurde in der Analyse ein in der Literatur nicht beschriebenes DOT-Segment beobachtet, in dem die Zelle wahrscheinlich den Stoffwechsel drosselt.

Die Wirkung der Arbeit kann auf zwei Ebenen gemessen werden. Auf wissenschaftlicher Ebene unterstreicht sie die Bedeutung eines angemessenen Umgangs mit Diskontinuitäten in der Bioprozesstechnik und bietet solide wissenschaftliche Methoden, um dies zu erreichen. Sie bietet Methoden zur Analyse und richtigen Integration der Abgas- und DOT-Signale in Hefe- und *E. coli*-Kultivierungsmodellen. Für *E. coli* werden außerdem Methoden zur genauen Überwachung der Stoffwechselaktivitäten der Zellen bereitgestellt. Auf industrieller Ebene stellt diese Arbeit Möglichkeiten zur Verfügung, um die Anzahl der für die Kalibrierung von Hefe- und *E. coli*-Modellen erforderlichen Durchläufe zu reduzieren und MBR-Probleme der Kontrolle von gelöstem Sauerstoff und der intermittierenden Feeding zu lösen, was zu einer Reduzierung der Kosten und des Zeitaufwands für die Prozessentwicklung führt.

# Abstract

Mechanistic models play an essential role in the development of bioprocesses. Despite their rigorous structure, they describe the processes with interpretable model parameters and provide a mathematical representation of the underlying dynamics. That's why they are employed extensively in process experimental design, monitoring and control.

However, many obstacles still hinder their effective utilization; models are formalized according to their goals. In an industrial context, models are built with simplifications, some of which lead to discontinuous models. This issue alongside the insufficiency and peculiarities of the analytics such as the off-gas and dissolved oxygen tension (DOT) signals, and the differences in processes conditions such as the working volume, can lead to non-adaptive (inflexible structures) non-robust (unreliable output) models. Further, similarly to the processes they describe, models must be adapted along the development life cycle. This thesis hypothesizes that proper sound scientific methods can address these challenges.

The thesis aims to achieve adaptive robust models by proposing methods to: overcome the reduced predictive capabilities of the discontinuous models, extract (latent) unexploited information from already-existing analytics, properly account for analytics peculiarities, and facilitate the transferability between different scales.

The individual outcomes of this work can be summarized as follows: A method combining Dynamic Time Warping (DTW) and the Least-Squares (LSQ) algorithms is developed as a fitting criterion for the calibration of yeast cultivation models which incorporate off-gas measurements. The method resulted in more representative unbiased parameter estimates. A comparative analysis of yeast cultivation modelling with and without explicit consideration of the sudden changes in reactor volume revealed a tangible effect on model outputs and parameter uncertainty due to error propagation. An event-driven modelling method (EDM) to properly account for volume changes is proposed. A comparative study to investigate all possible discontinuous behaviors in bioprocessing models is presented for yeast cultivations. The results revealed in detail the main sorts of discontinuities. An EDM workflow to properly handle them. The workflow resulted in robust model predictions and low parameters uncertainty. A method to segment the DOT signal for *E. coli* cultivations in minibioreactor systems (MBRs) is developed. The identified segments allowed the extraction of discrete features. Data-driven analysis of the features revealed a pattern of responses in particular DOT segments which are linked to specific metabolic states. A model-based analysis provided a possibility to estimate key model parameters of the metabolic overflow switching condition. Additionally, a DOT segment, not reported in literature, is observed in the analysis in which the cell is likely to attenuate the metabolism.

The thesis impact can be measured on two levels. On a scientific level, it highlights the importance of proper handling of discontinuities in bioprocessing, and provides sound scientific methods to achieve that. It provides methods to analyze and properly integrate the off-gas and DOT signals in yeast and *E. coli* cultivation models. For *E. coli*, it also provides tools for accurate monitoring of cell metabolic activities. On an industrial level, this work provides tools that help to reduce the number of required runs to calibrate yeast and *E. coli* models, and to address MBRs issues of oxygen supply and intermittent feeding planning, leading to a reduction in process development cost and time.

The novelty of this thesis can be summarized by: a) the detailed analysis of the discontinuities in bioprocessing models, and the workflow proposed to properly handle them, b) the new method to improve the estimation of model parameters in yeast cultivation with off-gas measurements, c) the novel analysis of the DOT signal for *E. coli* cultivations in MBRs of which one of the findings was a plausible relationship between metabolic adaptation behavior and a newly-observed DOT segment, and d) the novel workflow to extract model parameters from DOT signals.

The results of the proposed methods encourage further investigations on different organisms and platforms, to evaluate their generic applicability on a wider set of variable conditions.

# Content:

<b>ACKNOWLEDGEMENT</b>	<b>III</b>
<b>ZUSAMMENFASSUNG</b>	<b>IV</b>
<b>ABSTRACT</b>	<b>VI</b>
<b>1 INTRODUCTION, GOALS AND STRUCTURE</b>	<b>1</b>
<b>1.1 INTRODUCTION</b>	<b>2</b>
1.1.1 BIOPROCESS DEVELOPMENT AND MODELLING	2
1.1.2 MECHANISTIC MODELS	6
1.1.3 OBSTACLES TO THE EFFECTIVE DEPLOYMENT OF MECHANISTIC MODELS	12
<b>1.2 GOALS OF THE THESIS</b>	<b>20</b>
<b>1.3 STRUCTURE OF THE THESIS</b>	<b>21</b>
<b>2 RESULTS</b>	<b>22</b>
<b>2.1 ACHIEVEMENTS AND FINDINGS</b>	<b>23</b>
2.1.1 ACCURATE LOCATION OF MODEL EVENTS FOR ADAPTIVE MODELLING	23
2.1.2 IMPROVE MODEL CALIBRATION AND SCALABILITY FOR ROBUST MODELLING	27
<b>2.2 FULL MANUSCRIPTS</b>	<b>31</b>
2.2.1 IMPROVING THE CALIBRATION OF KINETIC GROWTH MODELS USING DYNAMIC TIME WARPING	31
2.2.2 EVENT DRIVEN ANALYSIS TO ENHANCE MODEL CALIBRATION OF EXPERIMENTS WITH HIGH OFFLINE RATE SAMPLING	38
2.2.3 EVENT DRIVEN MODELING FOR THE ACCURATE IDENTIFICATION OF METABOLIC SWITCHES IN FED-BATCH CULTURE OF <i>S. CEREVISIAE</i>	45
2.2.4 ROBUST MODELLING OF <i>S. CEREVISIAE</i> FED-BATCH CULTURES BY PROPER HANDLING OF MODEL DISCONTINUITIES	64
2.2.5 A UNIQUE RESPONSE BEHAVIOR IN THE DISSOLVED OXYGEN TENSION SIGNAL OF <i>E. COLI</i> IN MINIBIOREACTOR SYSTEM EQUIPPED WITH INTERMITTENT BOLUS FEEDING	73
<b>3 CONCLUSION, IMPACT, AND OUTLOOK</b>	<b>108</b>
<b>3.1 CONCLUSIONS, IMPACT AND OUTLOOK</b>	<b>109</b>
3.1.1 CONCLUSIONS	109
3.1.2 IMPACT ON ACADEMIC AND INDUSTRIAL LEVEL	112
3.1.3 OUTLOOK	114
<b>4 APPENDIX</b>	<b>115</b>
<b>4.1 DISCONTINUITIES AND ODE SOLVERS</b>	<b>116</b>
4.1.1 DETAILS ON THE NUMERICAL SOLUTION	116
4.1.2 STATE-OF-THE-ART ODE SOLVERS	118

4.1.3	MATHEMATICAL DISCONTINUITIES IN YEAST GROWTH MODEL	119
4.1.4	VIDEO EXPLANATION	120
<b>4.2</b>	<b>VIDEO CONTRIBUTIONS</b>	<b>121</b>
<b>4.3</b>	<b>POSTERS</b>	<b>122</b>
<b>5</b>	<b>REFERENCES</b>	<b>124</b>

---

# 1 Introduction, Goals and Structure

# 1.1 Introduction

## 1.1.1 Bioprocess development and modelling

The advancement in technology and digitization in the recent decades accompanied with the rapid increase in computational power and the decrease in hardware costs have accelerated the industry transformation. The idea of “Industry 4.0” is well established today [1]. This idea refers to the industry in which smart factories have different units that interact with each other in the real-time, exchange information and perform decisions with little to no human intervention [2]. However, (bio)pharmaceutical industry is still lagging behind this transition. Currently, (bio)pharmaceutical industry in general does not fulfill even the standards of the earlier generation “Industry 3.0” concept [1].

However, a growing emphasis by the regulatory authorities such as the U.S. Food and Drug Administration FDA and the European Medicine Agency EMA to accelerate the industry, to develop the processes, and to reduce the costs has led to initiatives such as Quality by Design (QbD) and Process Analytical Technology (PAT)[3], [4].

These initiatives encourage the voluntary development and implementation of innovative pharmaceutical processes, manufacturing, and quality assurance for adaptive process understanding, and better prediction and consistency [5]–[7].

The biopharmaceutical industry has complex processes, a variant operational environment, strict regulations and high material and operation costs. Therefore, the industry tends to rely on reliable but outdated methods for production rather than risking adopting new approaches.

The development cycle for the production of biopharmaceuticals and biosimilars has three main stages: in the first stage, the screening and characterization of the organisms take place. Second, the reaction conditions (e.g., optimizing medium and process variables) are optimized. In a last stage, the scale up to pilot and production scale takes place [8]. During these stages, a vast number of development cultivations is usually required [9]. It is important that the gathered knowledge in all stages and cultivations is scalable.

Developing a pharmaceutical production cycle requires improvements in production techniques on two levels: one the physical component level and on the virtual component levels. While the first component can be transferred with little effort from other industries, the virtual component including process modelling is highly specific and needs to be developed independently [1], [10]–[13].

In (pharmaceutical) bioprocessing industry, the challenge is to cut costs, labor, and time. However, bioprocess development is often too long and owes a very high failure rate [14]–[16]. In general, each development cycle produces high volumes of variant datasets, generated by the sensing and controlling systems, e.g., by the online and offline measurements methods. The data is usually a result of many experiments that are conducted in the design space in the vicinity of the desired process goal. This data is usually (pre-)processed with chemometric or statistical methods to extract direct process information, which is then compiled to extract process knowledge. A clear and effective hierarchy of data to information to knowledge (DIK)

is of importance for robust bioprocess development [16]–[19]. Pharmaceutical industry is reported to suffer from data rich information poor syndrome DRIP [20].

Hence, effective modelling methods with a broad scope that covers all cycle stages is of a great importance. These methods can promise a huge reduction in the cost and time. Some estimated a possibility to halve the required resources when effective in-silico methods are applied [11], [21]. Supported by the recent advancement of sensing technology and analytics techniques, robotic and high throughput systems, and high and cost-effective computational power, the pharmaceutical industry has all the prerequisites for the digital transformation.

However, the effective modelling methods have to offer accurate, robust and easily interpretable results on the relation between the critical process parameters (CPPs) and the critical quality attributes (CQAs), and to enable efficient data and information management, knowledge transfer, and advanced process monitoring and control [22]. In the context of (pharmaceutical) bioprocessing industry, the models have to describe the most important, but not all, characteristics of the cultivated cells. Having an interpretable model structure is favorable to understand the metabolic interactions complexity. Such a structure is usually achieved with an explicit (not latent) mathematical description of cell metabolic activities and the affecting process conditions. This also facilitates the validation of the models in the absence or scarcity of the targeted analytics. Modelling methods that fit last criterion are rare.

Different modelling methods are available to fulfill the previous requirements. In general, three main modelling approaches are used in bioprocessing: I) data-driven and statistical methods, II) mechanistic methods, and III) hybrid methods which combine both.

Data-driven and statistical methods are gaining popularity as tools to investigate correlations in bioprocessing. The biological nature of the bioprocesses means dynamically changing inter-relationships between process variables [1].

Standard statistical methods are used mostly in the context of Design of Experiments (DoEs) to provide insight into parameter correlations [23]. Multivariate data analysis methods (MVDA) like principal component analysis (PCA) and partial least square regression (PLSR) are widely used to investigate highly correlated datasets. They are widely applied to group: experimental runs according to process conditions and phases, process outliers and product quality [1], [24]–[27], and as a mean to understand the latent relationships between process variables and process products [28].

Data-driven methods require a large set of data in order to completely identify the large number of model parameters resulting from the large number of process critical factors that should be investigated [29]. The lack of such a data volume means an overfitting problem, which is common in data-driven methods. A possible remedy is to generate a large enough data set with enough process and cellular variability. This is a costly and time-consuming task for industries under pressure for short and cost-effective development cycles. Further, the generated data in the corporations in the pharmaceutical industry are not shared, and it is unlikely that one company will diversify its data portfolio enough to overcome the above-mentioned problem [11]. The vaccination development for the recent COVID-19 pandemic shows a possible example of that.

Data-driven methods feature the possibility to run methods without prior knowledge, but this also means that model results are highly dependent on the used dataset. Extrapolation outside the used dataset space is therefore unreliable. In this sense, data-driven methods can be seen as “tailored” rather than “standard” methods.

Mechanistic methods, on the other hand, rely on the prior knowledge gained by different scientific fields like biology and chemistry to form mathematical equations. These models represent the knowledge of the underlying physical characteristics of the process and the physiological behavior of the organisms using mathematical expressions and model parameters [30]–[32]. One could rely on literature to accumulate knowledge of different organisms or platforms without the need to have the experimental data. Due to the validity of the equations representing the underlying phenomena, the interaction between these equations is also valid. Hence, the extrapolation results of mechanistic models are much more reliable than the results of their data-driven counterparts [33]. However, mechanistic models have their own problems of structural rigidity and numerical complexity. Therefore, they need more effort to be implemented properly, they also need some basic knowledge to properly set the numerical solution environment.

During the development of bioprocesses, mechanistic models play an essential role for effective experimental design [34]–[39], real-time monitoring and predictive control [40]–[44]. They could predict quantities which are hard or costly to be measured, e.g., soft sensors [45], also, they are increasingly used in the frame of multi-objective control to promote increased selectivity of products [33], making them indispensable tools in biotechnology.

Hybrid methods which combine data-driven and mechanistic approaches are now gaining more interest as a nice possibility to improve the abstraction level of data-driven approaches with other knowledge sources [1]. A common hybrid model contains two components: a data-driven component; mostly neural networks, and a mechanistic component. The experimental data is fitted to the mechanistic model and the deviations are overcome by a trained neural network [23].

Figure 1 shows an illustration of the different modelling approaches in bioprocessing according to the needed pre-knowledge.

The main interest in this thesis is the mechanistic models. The thesis embraces the notion that mechanistic models still have room for improvements. An improved mechanistic model means better modelling output for both mechanistic and hybrid methods.

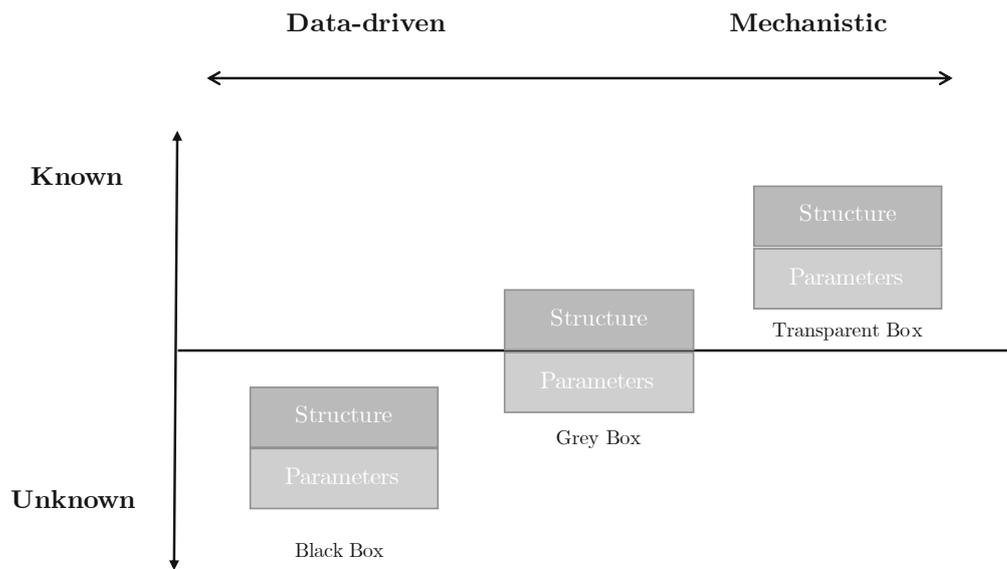


Figure 1: Modelling methods in bioprocessing according to the prior knowledge needed to construct model structure and define model parameters, adapted from [46]

## 1.1.2 Mechanistic models

### The definition of the mechanistic models in bioprocessing

Mechanistic models are traditionally defined as “ab initio” models, physical models, or first principal models. However, more accurately, they are non-empirical models built upon physical, chemical and/or biological principles [11].

Mechanistic kinetic growth models use stoichiometric information, nonlinear reaction rates and mass and concentration balances [47]–[49], and are usually written as a set of deterministic and continuous Ordinary Differential Equations (ODEs).

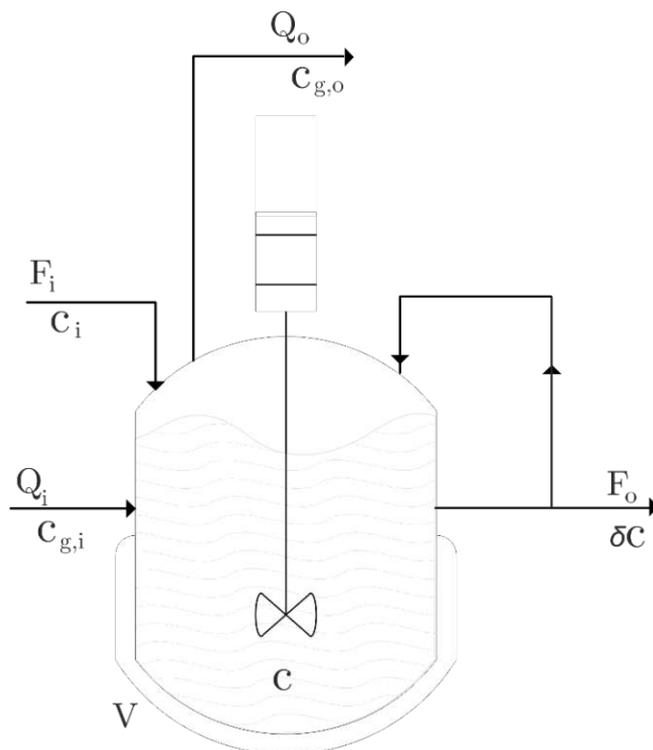


Figure 2: A scheme for an ideal bioreactor system with a homogenic medium and perfect mixing

In a standard generic bioprocess, and for an arbitrary reacted component with a concentration  $c$  in a reactor with a working volume  $V$ , the mass balance ( $[g \cdot h^{-1}]$ ) is written according to [50]:

$$\frac{d(V \cdot c)}{dt} = F_i c_i + Q_i c_{g,i} - F_o \cdot \delta c - Q_o c_{g,o} + V r_c \quad (1.1)$$

Where  $Q_i$  is the inlet gas flow rate,  $Q_o$  is the outlet gas flow rate,  $F_i$  is the medium inlet flow rate, and  $F_o$  is the medium outlet flow rate.  $r$  is the volumetric reaction rate, it gives a positive value for production and a negative value for consumption for the arbitrary component  $c$ . Subscript  $g$  refers to the gas flows,  $i$  and  $o$  refer to the inlet and outlet flows.  $\delta$  is the separation factor, refers to the cells that are recirculated, mainly in the continuous cultivations.

The change in the volume equals the difference between the inlet  $F_i$  and outlet flow  $F_o$ . By dividing both sides of eq (1.1) by the volume, the general mass balance equation that describes the change in the concentration of a component  $c$  is:

$$\frac{dc}{dt} = \frac{F_i}{V}(c_i - c) - \frac{F_o}{V}(\delta c - c) + r_c + \frac{Q_i}{V}c_{g,i} - \frac{Q_o}{V}c_{g,o} \quad (1.2)$$

This equation is valid for common process operations in batch, fed-batch, and continuous processes. The previous equation (1.2) can be shortened in case of batch and fed-batch cultures where  $\delta = 1$ . The Gas Transfer Component can be written as  $GTR = \frac{Q_i}{V}c_{g,i} - \frac{Q_o}{V}c_{g,o}$ . Most of the components in the bioreactor don't involve any gas transfer, and therefore their  $GTR$  is zero. The volumetric reaction rate  $r_c$  consists of the yields coefficients  $Y$  that links the specific reaction rates  $q$  with the reacting components  $c$ , generally as  $r_c = Y \cdot q \cdot c$ .

A generalized matrix-form of equation (1.2) for  $k$  feed,  $j$  gaseous components,  $m$  reacting components (liquid and gaseous),  $n$  independent reaction rates, and a biomass concentration of  $c_x$  is written as:

$$\frac{d}{dt} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} = \begin{bmatrix} Y_{1,1} & \cdots & Y_{n,1} \\ Y_{1,2} & \cdots & Y_{n,2} \\ \vdots & \ddots & \vdots \\ Y_{1,m} & \cdots & Y_{n,m} \end{bmatrix} \cdot \begin{bmatrix} \mp q_1 \\ \mp q_2 \\ \vdots \\ \mp q_n \end{bmatrix} \cdot c_x + \sum_{k=1}^{k^{max}} D_k \cdot \begin{bmatrix} c_{k,1} - c_1 \\ c_{k,2} - c_2 \\ \vdots \\ c_{k,m} - c_m \end{bmatrix} + \sum_{j=1}^{j^{max}} GTR_j \quad (1.3)$$

Where  $Y_{n,m}$  is the yield of a reaction state  $m$  and a reaction rate  $n$ .  $q_n$  denotes the biomass specific reaction rate.  $D_k$  is the dilution rate. This representation of the mechanistic equations is adopted by many authors [51], [52]. It provides a concentrated view on model structure.

The main characteristic of the mechanistic models, aside from having a model structure, is having rates and yields with interpretable parameters. The rate (named sometimes a kinetic or a kinetic link) is formed based on an empirical observation and matched with the simplest and most accurate mathematical representation. That's why there are plenty of rate kinetics that describe the same phenomena. Some of these kinetics describe a continuous behavior, others feature a discrete behavior such as Blackman kinetics. Kinetics usually have "calibration" parameters, e.g., Monod term and its maximum value at saturation and the affinity constant. Kinetics can also differ in terms of the process and cellular condition they cover such as pH and temperature [53]–[55]. [56], [57] recently showed a nice review on bacterial growth rate kinetics. Table 1 shows some of the most common rate kinetics, and Figure 3 shows their behaviors.

The main question when forming and validating a new kinetic is which metabolic activities and process conditions should be taken and which should be smoothed or simply neglected. This can be seen by the numerous publications that propose models and/or kinetics for bio-organisms.

Table 1: Commonly used rate kinetic in mechanistic models,  $c_n$  referred to the limiting substrate for a rate  $q_n$  with a saturation value  $q_n^{max}$  and substrate affinity constant  $K_n$ . Inhibition state and affinity constant are  $c_i$  and  $K_i$  consequently.

Monod [58]	$q_n = q_n^{max} \frac{c_n}{c_n + K_n}$
Moser [59]	$q_n = q_n^{max} \frac{c_n^p}{c_n^p + K_n}$
Blackman [60]	$q_n = \begin{cases} \frac{q_n^{max}}{2k_n} \cdot c_s & \text{if } c_n \leq 2k_n \\ q_n^{max} & \text{if } c_n > 2k_n \end{cases}$
Competitive inhibition	$q_n = q_n^{max} \frac{c_n}{c_n + K_n(1 + \frac{c_i}{K_i})}$
Non-competitive inhibition	$q_n = \frac{q_n^{max}}{(1 + \frac{c_i}{K_i})} \frac{c_n}{c_n + K_n}$
Substrate inhibition	$q_n = q_n^{max} \frac{c_n}{K_n + c_n(1 + \frac{c_i}{K_i})}$

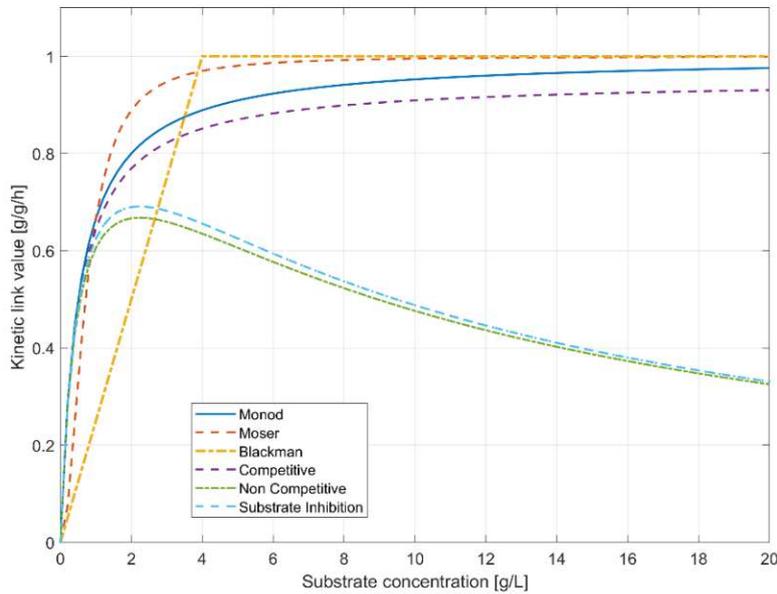


Figure 3: Illustration of some of the common kinetic links shown in Table 1,  $K_i = K_s = 0.5 \left[\frac{g}{L}\right]$  and inhibition state concentration  $c_i = 1 \left[\frac{g}{L}\right]$ .

## Types of the mechanistic models

Mechanistic models are broadly categorized into *structured* or *unstructured* models and into *segregated* or *non-segregated* models. Figure 4 illustrates the differences between these categories.

If the cell population is considered to have equal properties, then the model is categorized as *segregated*. If not, then the model is called *non-segregated*.

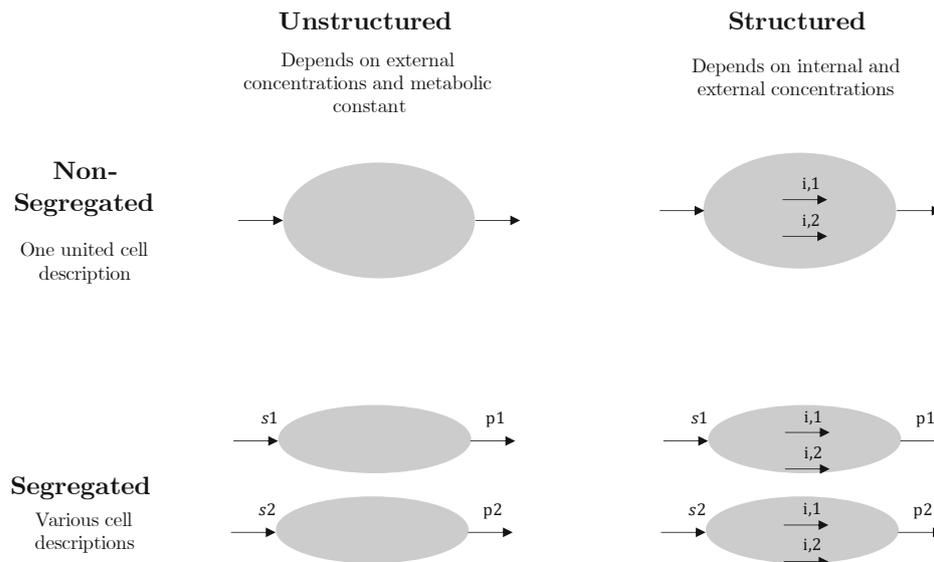


Figure 4: The broad categories of the mechanistic models

*Unstructured* models (sometimes called Macro-models) do not incorporate a detailed metabolic and physiological description of the organism.

*Structured* models (sometimes called Micro-models) provide a detailed description of the intracellular activities. However, due to the complexity and the difficulty of measuring all concentrations, their application in practice is still very limited [30], [61], [62].

This is why, in the context of industrial biotechnology, model-based monitoring, control and characterization rely mainly on *non-segregated unstructured* models [63]–[65].

## Adaptiveness and robustness of the mechanistic model

Mainly, there are two ways to offer model flexibility with interpretable mathematical description in bioprocessing modeling; having a set of kinetics or having a flexible model structure that changes depending on certain process or cellular conditions. A combination of both is possible and helpful if there is no repeated description of the same phenomenon in the structure and the kinetic terms, so the number of model parameters doesn't increase without any additional information added to the model.

The first approach is referred to in this thesis as the “adaptive kinetics” approach, while the second approach is referred to as the “adaptive structure” approach. Figure 5 illustrates the two possibilities.

For the “adaptive structure” approach, an additional degree of model flexibility without increasing the number of parameters can be afforded if the switching conditions are reformed from, or as the same as, some of the model's own kinetics. This is discussed in detail in section 2.1.1.

This thesis has the goals to adopt the simplest and most accurate kinetics description and offer more flexibility on the model's structure level, thus, in the thesis context, the “adaptiveness” is understood by aiming to achieve an “adaptive structure”.

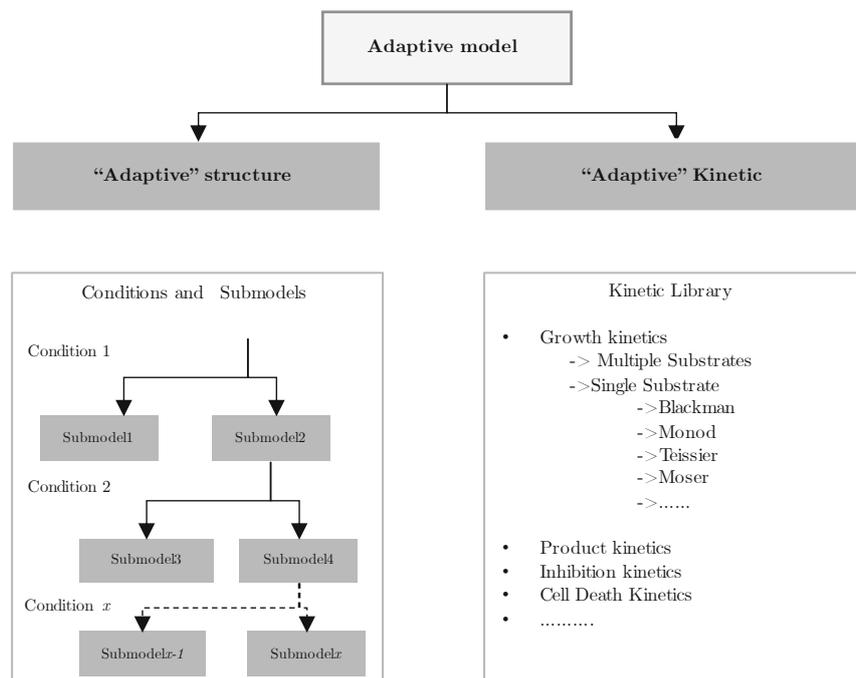


Figure 5: Adaptive modelling has two main approaches: Left) adaptive structure, where the different model equation sets are active over different time windows, and Right) adaptive kinetics, where different kinetic links are proposed to account for the changing cell behavior.

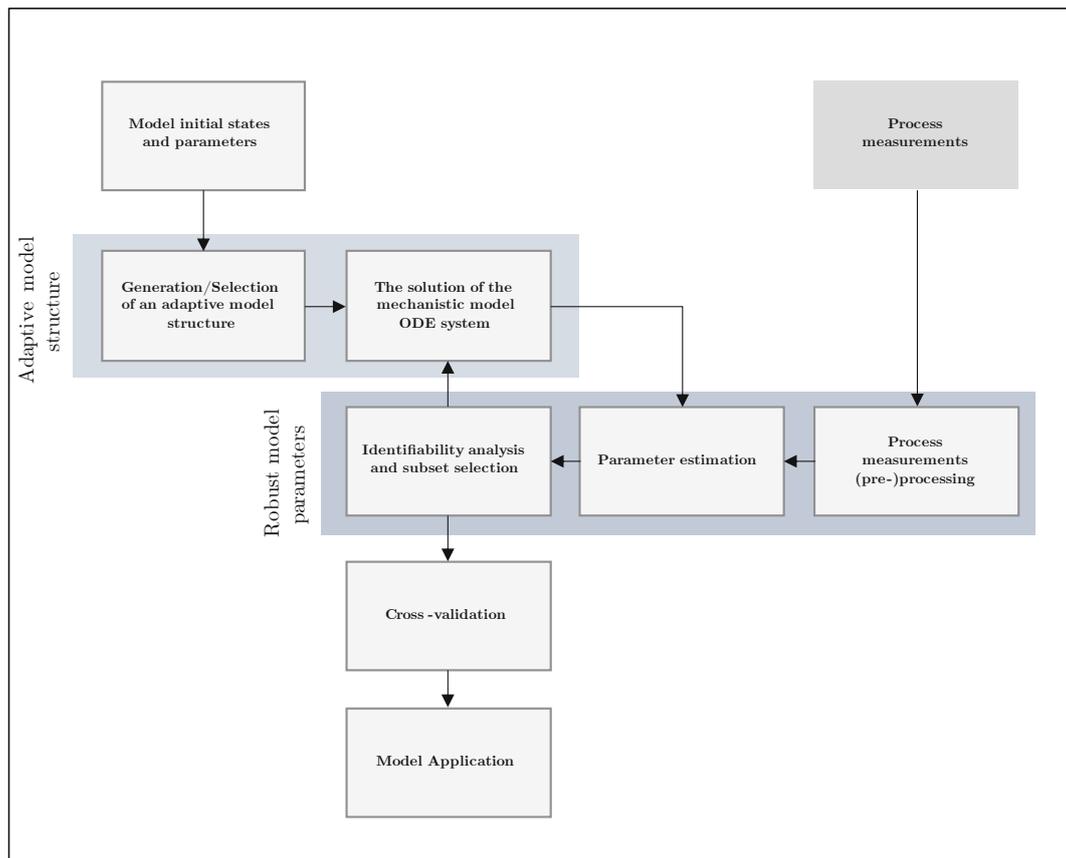
The matrix-form of the model mentioned in (1.3) offers an elegant and uncomplicated way to implement the adaptive structure approach.

Robustness in the context of mechanistic models is understood on the level of model prediction.

A robust model is expected to have almost the same output regardless of the deviations in its input. Robustness is defined by the *IEEE* standards glossary as "The degree to which a system or component can function correctly in the presence of invalid inputs or stressful environmental conditions" [66].

Mechanistic model output is deterministically linked to the model parameters. Thus, the robustness is also understood on the parameter level. A robust model is expected to have similar parameter estimates regardless of the deviations in its input, e.g., process measurements. These deviations origin, among other reasons, from the noise in the measurements, and the way the measurements are (pre-)processed and incorporated inside the model. This highlights the need for robust methods for the solution of the model, and for the measurements data (pre-)processing.

Model robustness and adaptiveness are linked to each other. An adaptive structure helps to achieve robust parameter estimates by offering more flexibility to the model in regions rich in information. On the other hand, the switches between the adaptive structure submodels cannot be accurately defined without robust model parameters.



## Model development cycle

*Figure 6: A proposed development cycle for an adaptive and robust mechanistic model*

### 1.1.3 Obstacles to the effective deployment of mechanistic models

An effective deployment of mechanistic models in industry is still hindered by: A) practical challenges such as the availability of analytics, and the transferability of models between different applications; e.g., changes in media composition [67], or changes in reactor working volume [68], and by B) theoretical challenges related to the development of the applied methods that must take into account the peculiarities of bioprocessing, for example the different nature of analytics techniques. Different methods and tools to address these issues have been proposed. For example, methods for the reduction of dynamic models [69], modeling of cellular metabolic pathways [70], structural analysis of model equations [71], dynamic and global optimization [72], [73].

However, this thesis is interested in proposing solutions of the following challenges:

#### 1.1.3.1 Model structure formation & solution method

Mechanistic models are formed according to their objectives. Unstructured models are common in industry and have the aim to provide knowledge on industrial relevant goals where the analytics are limited to save costs and time. Therefore, the model mathematical representation is simplified to the central metabolic pathways of the cultivated organism. The result is usually a discontinuous piecewise model.

##### 1.1.3.1.1 The solution of continuous and discontinuous ODE models

#### A word on ODEs solution, stiffness and discontinuities

From a mathematical point of view, a model simplification is considered when the underlying modelled differential equations have a wide variation in time and/or length. In this case, either the faster or the slower dynamics have to be eliminated (simplified). The decision of which dynamics to be simplified is usually done taking the biological meaning and importance into consideration. The line separating important and unimportant dynamics is itself a blurred one. In bioprocessing, this results in simplifying the continuous metabolism transitions into discrete transitions named “*discontinuities*”, sometimes referred to as “*events*”.

Discontinuities cause difficulties for ODE solvers. Handling a discontinuity should happen on a solver level and a problem formulation level [74].

Initial value problems with a changing state  $c$  over time  $t$  described by a function  $f$ , reaction parameters  $p$ , and initial conditions  $c_0$ , has a numerical time resolved solution, where the step size  $h$  determines the accuracy of the solution. The solution exists if there is a known solution  $c_i$  at each point  $t_i$ . The solution at the next step  $t_{i+1}$  is:

$$c_{i+1} = c_i + \varphi \cdot h = c_i + \int_i^{i+1} f(t, c). dt \quad (1.4)$$

Where  $\varphi$ : slope, and  $h$  step size. The solution is correct if for all time steps  $i$ , the following equality holds:

$$\varphi \cdot h = \int_i^{i+1} f(t, c). dt \quad (1.5)$$

Different methods are proposed to solve (1.5). Generally, the evaluation at different points from the first time point helps to adjust step size correctly. If solver estimates the solutions at multiple distances between  $t_i, t_{i+1}$ , then it is called variable-step solver, i.e., different orders like in Gear's method and Adams-Bashforth-Moulton method. On the other hand, calculating the value of the system at the next time step starting from previous ones splits solvers family into implicit and explicit, where it is known that implicit solvers perform better with stiff problems. Table 2, Table 3, and Table 4 in the appendix show the main solvers and their methods that are used in *MATLAB*, *Python* and *Modelica* environments, these methods among the most effective and widely used in scientific computing [75], same algorithms are commonly implemented in other programming environments [76]–[78]. In an ODE solver context, the relative tolerance is a measure of the local error  $\varepsilon_{rel}$  at each step. The absolute error  $\varepsilon_{abs}$  can be seen as a general threshold. Both must not exceed the predefined (acceptable) value  $\varepsilon_{acceptable}$ :

$$\varepsilon_{acceptable} \leq \max(\varepsilon_{rel} \cdot |c_i|, \varepsilon_{abs}) \quad (1.6)$$

Step size  $h$  selection is set differently for each solver to get accurate solution steps and avoid approximation errors [79],[80]. These criteria that are used for error estimation are designed to provide solutions accurately but not necessarily in a stable manner [81].

In conventional ODE solvers, the exact location of discontinuities is not accurately determined. The solvers detect the discontinuity in the state variable (i.e., on the concentration level in bioprocessing models) resulting from a discontinuous constitutive equation (i.e., on the rate level in bioprocessing models). Because the discontinuity is located at the state level, the interpolating polynomial is not representative of the system behavior [74].

Additionally, stiff problems can form an “almost-discontinuity” behavior. This means although the mathematical representation is not discrete, the behavior is very similar to a discrete behavior. A very common example of that is Monod-like kinetics that have the form of  $\frac{c}{c+K}$ . that mainly link growth and uptake to concentrations [30]. The problem happens with small  $K$  values. Appendix 4.1.3.1 shows an example of: I) an “almost-discontinuity” caused by the stiff behavior of Monod kinetics with low  $K$  value. This leads to “non-physical” solutions. II) Model discontinuity caused by the bottleneck kinetics in the model of [82]. The figure shows exemplarily how the discontinuities in bioprocessing could happen on the rate, not the concentration, level.

Appendix 4.1.3.2 shows a quantitative analysis on the simulation errors associated with “almost-discontinuity” behavior of Monod kinetic.

## Event detection and its issues

Event detection relies on Root Finding Methods. These methods monitor the change of the sign of the solution of algebraic equation between  $[t_i, t_{i+1}]$ . The search for roots  $g(t, f(t, y)) = 0$  is considered solved when there is an interval  $[t_i^*, t_{i+1}^*]$  that satisfies:

$$g(t_i^*) \cdot g(t_{i+1}^*) < 0 \text{ and } |t_i^* - t_{i+1}^*| < \delta \quad (1.7)$$

Which implies that zero crossing is contained in a small offset  $\delta$  between  $[t_i^*, t_{i+1}^*]$  [83]. Error term  $\delta$  is set in most solvers including *MATLAB*'s to machine precision [75].

## 1.1.3.2 Model parameters estimation

### 1.1.3.2.1 Lack of sufficient analytics

The complexity of mechanistic models depends on the availability of process analytics such as online gas analyzers, advanced tools such as automatic liquid handling and sampling, and hardware like HPLC (high-performance liquid chromatography), NIR (near infrared spectroscopy) and FIA (flow injection analysis) [84]. In industrial context these analytics are limited to cut costs and reduce development time. However, a nice opportunity to override this issue, is to seek to extract additional “unexploited” information from already-existing analytics data.

In yeast as well as many other bio organisms, Respiratory Quotient (RQ) derived from off-gas information proved to be used to identify the metabolism. It is defined as:

$$RQ = \frac{CER}{OUR} \quad (1.8)$$

Where CER is the carbon dioxide evolution rate, and OUR is the oxygen uptake rate. A RQ greater than one indicates that the yeast is producing ethanol by Crabtree effect (oxidoreductive growth). A RQ close to one indicates that glucose is mostly oxidatively consumed. RQ values around 2/3 indicate ethanol oxidative consumption.

Dissolved oxygen tension (DOT) is a commonly obtained online signal in *E. coli* cultivations. It can be of a similar usefulness as the off-gas signals for yeast cultures to assess the metabolic activities. However, for *E. coli*, the encoded metabolic activities in the DOT signal are more ambiguous, and signal details have higher frequency, hence, the separation of the useful characteristics from the background noise is more difficult. Also, a combination of sensor time delay and high substrate affinity of *E. coli* shifts the metabolic activities nonlinearly along the DOT signal [85]. This shows the need for an advanced analysis to extract the relevant useful information from the DOT signal.

DOT signal is influenced by two opposing components: cell oxygen demand defined by oxygen uptake rate (OUR) and the oxygen delivery to the medium by reactor aeration and stirring systems defined by oxygen transfer rate (OTR).

$$\frac{dDOT}{dt} = OTR - OUR \quad (1.9)$$

However, DOT and off-gas signals are usually more complex than other liquid concentration states. They contain both low and high frequency information where rapid changes represent mostly certain limitations or metabolic shifts, and slower changes usually are correlated with the active metabolic states.

### 1.1.3.2.2 Methods for parameters estimation, identifiability & scalability

Model calibration is one of the delicate tasks when having a mechanistic model. As the parameters have a physiological meaning, their values have to be carefully estimated and checked. Deviations and biases can affect model parameters and result interpretation.

#### Bias in parameter estimation

Forming an objective function is usually an undervalued step when developing models. Probably, the most common (error) criterion used for fitting model predictions to the data in bioprocessing is the Euclidean Least-Squares (LSQ) criterion. LSQ usage is valid for regression problems where the residuals are assumed to be normally distributed with equal variance (homoscedastic), and independent of one another. If any of these distributional assumptions are violated, several of the desirable properties of a least squares fit may not hold [86]. However, this should not be the only criterion to worry about. The objective function should also consider the nature of the data under investigation. In bioprocessing, the measurement of each single experimental run comes from different sensor and measurement devices. The frequency of the measurements can hugely vary from continuous, semi-continuous to discrete.

Considering all these measurements in one objective function is a tough task. One common example is bioprocesses with on-line off-gas signals and off-line liquid concentrations measurements.

Parameter estimation algorithms use LSQ to evaluate the quadratic fitting error simply by subtracting measurement and prediction values on the y axis for each x timepoint. Therefore, if the measured signals show shifts or have measurement noise, which is so common in bioprocessing, then the LSQ function would report a disproportionate change in the error value. Also, in the presence of structural model simplifications and noise, measurements are usually fitted by smooth curves, which are optimal in the sense of the quadratic error criterion, but do not mimic the shape of the measured signals.

For  $N$  models state,  $M$  measurements at different time points, and with  $\theta$  as unknown parameter vector, the minimization problem using LSQ reads:

$$\begin{aligned} \min_{\theta} \phi^{NRSS}(\theta) \quad \text{with} \\ \phi^{NRSS}(\theta) = \frac{1}{M} \sum_{i=1}^N \sum_{j=1}^N (c_{i,j}(\theta) - c_{i,j}^m)^2 \end{aligned} \quad (1.10)$$

Dynamic Time Warping (DTW) is a method proposed in computer science and signal processing fields for shape recognition and signals alignment. It has a high potential to yield model predictions which mimic the interesting features in the measured signals and provide robustness against some off-gas signal peculiarities.

To get shape preserved matching between two sequences, DTW seeks an optimal path  $k$  through the matrix which minimize the warping cost [87] :

$$\min_{\theta} \phi^{DTW}(\theta) \quad \text{with}$$

$$\phi^{DTW} = \min_k \left( \sum_{i,j} \gamma_{i,j} \right) \quad (1.11)$$

Where  $\gamma$  are the elements of the warping path. The optimal path is found by calculating the minimum cumulative distance of the current element and the other three adjacent cells in DTW matrix following:

$$\gamma_{i,j} = \varepsilon_{i,j} + \min(\gamma_{i-1,j-1}, \gamma_{i-1,j}, \gamma_{i,j-1}) \quad (1.12)$$

Figure 7 shows exemplarily the difference between LSQ and DTW algorithms output when a shift is introduced to the off-gas signal.

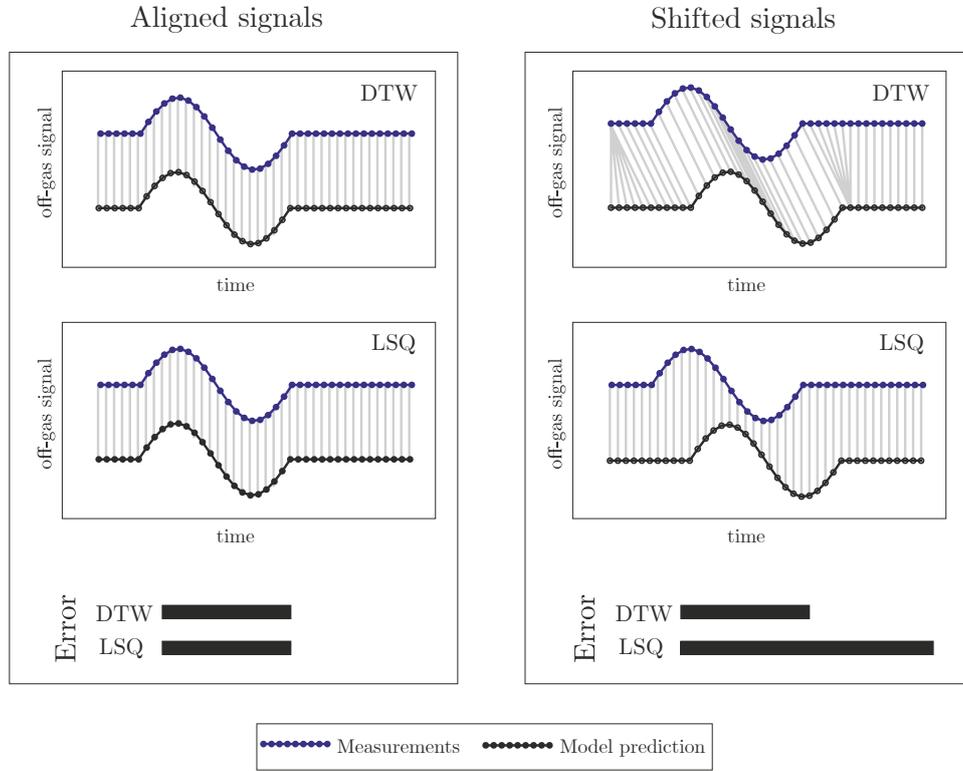


Figure 7: Euclidean Least-Squares (LSQ) and Dynamic Time Warping (DTW) as fitting criteria for off-gas signals in bioprocessing

## Parameters identifiability

Another important issue when estimating mechanistic model parameters is to select an identifiable best fitting parameter (sub-)set. This has been shown to be addressed with different techniques; [88] shows a nice review and proposed a method to deal with this issue.

The scheme of [88] is applied to check the identifiability of the parameters. This scheme requires estimated parameters  $\hat{\theta}$  and the corresponding sensitivity matrix  $S(\hat{\theta})$ .  $\hat{\theta}$  are obtained from repeated numerical solutions of a nonlinear regression, where the initial parameters are defined by stochastic sampling in a reasonable parameter space around initial values.

The sensitivity matrix  $\tilde{S}$  is obtained by normalizing  $S$  with the initial parameters and model output. Singular value decomposition (SVD) is used to detect any linear dependencies in  $\tilde{S}$ . By decomposing

$$\tilde{S} = U\Sigma V^T \quad (1.13)$$

where  $\Sigma$  matrix is found which holds the singular values of  $\tilde{S}$ . The singular values in  $\Sigma$  are then used to calculate:

- 1- the condition number ( $\kappa$ ) which is a measure of the sensitivity of model results to the perturbation of the parameters.
- 2- the collinearity index ( $\gamma$ ) which quantifies the collinearity of the parameters.

The parameters are ranked according to their linear independence and the above metrics are used to perform a parameter subset selection (SsS). The identifiable parameter subset simultaneously satisfies both sensitivity and linear independence conditions.

A stochastic shuffling which involves resampling of the experimental data and re-estimation of the parameters to assess the robustness of the model with the identifiable parameters can be done using bootstrapping technique [89].

## Parameters scalability

Models in bioprocessing should, at least theoretically, work at different scales. In reality, due to process and cell variations, the models don't usually fulfill this assumption.

Models can be adapted to different scales when their parameters are adapted accordingly [9], [90]. However, determining which parameters are changing and how much they change is challenging. Miniaturized bioreactor systems (MBRs) are common platforms for so-called "scaling-down" of different organisms such *E. coli*, *S. cerevisiae* and *Bacillus subtilis* [8], [91]–[93]. When the optimal process conditions and best organisms and/or strain candidates are identified, the cultivations are scaled-up. Therefore, tools that give insights on process dynamics or/and metabolic activities in MBRs to identify critical patterns for scale-up are of great importance [1]. Figure 8 illustrates the concept of scalability. The parameter set  $P$  differs between scales with a quantifiable change  $\pm\sigma$ .

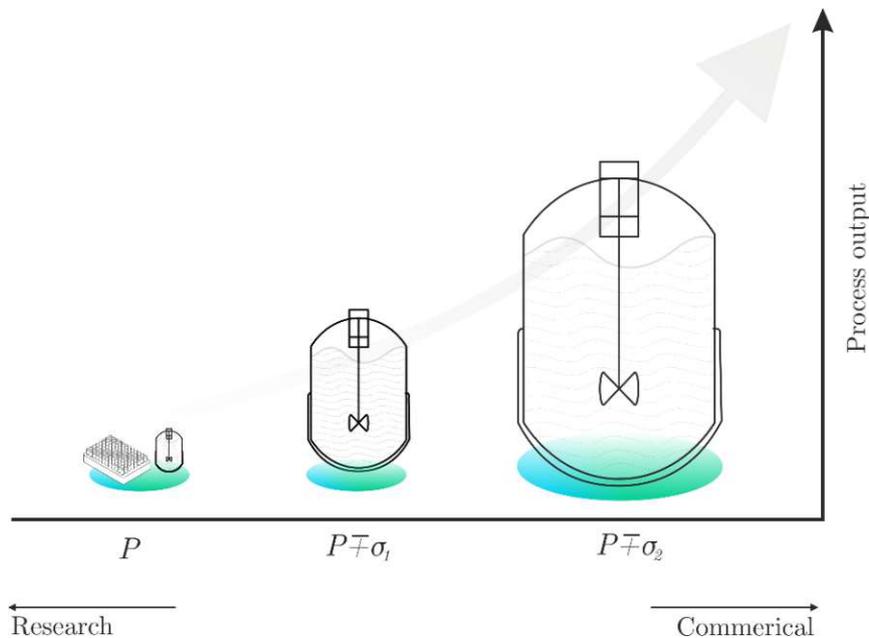


Figure 8: Bioprocess scalability is associated with a change in the model parameter values at each scale. Therefore, methods should correctly quantify the parameter set  $P$  and the change  $\sigma$  to achieve robust modelling across all scales.

## 1.2 Goals of the thesis

The main aim of the thesis is to develop sound science methods for the generation of adaptive, accurate and robust mechanistic models. To achieve that, models have to have a flexible representative structure that describe the underlying biological behavior, and proper methods have to be applied to ensure the robustness of the models' output.

This work seeks to overcome the obstacles that still hinder an effective exploitation of the models in industry and academia. These obstacles are described in detail in section 1.1.3. To overcoming these obstacles, this works aims to achieving adaptive robust models by:

1. Highlighting the importance of proper handling of discontinuities in bioprocessing models, as shown in sections 2.2.2, 2.2.3, and 2.2.4.
2. Providing a workflow to support the solution of discontinuous mechanistic models, and to overcome the problem of the reduced predictive capabilities of the models due to model simplifications as shown in sections 2.2.3 and 2.2.4.
3. Extracting the unexploited knowledge from already-existing analytics, as shown in sections 2.2.1 and 2.2.5.
4. Reducing model parameters uncertainties, by obtaining accurate and representative (unbiased) parameters against numerical and measurement noise, as shown in section 2.2.3.
5. Providing model-based tools to observe process dynamics and metabolic activities in lab and milliliters scales reactors to facilitate the transferability between different scales, as shown in section 2.2.5.

To check the validity and the transferability of the developed workflows and methods, applications had to be tested at the university and by the industrial partner at different process platforms, both at liter and milliliter scale with different production hosts (*Saccharomyces cerevisiae* and *E. coli*).

## 1.3 Structure of the thesis

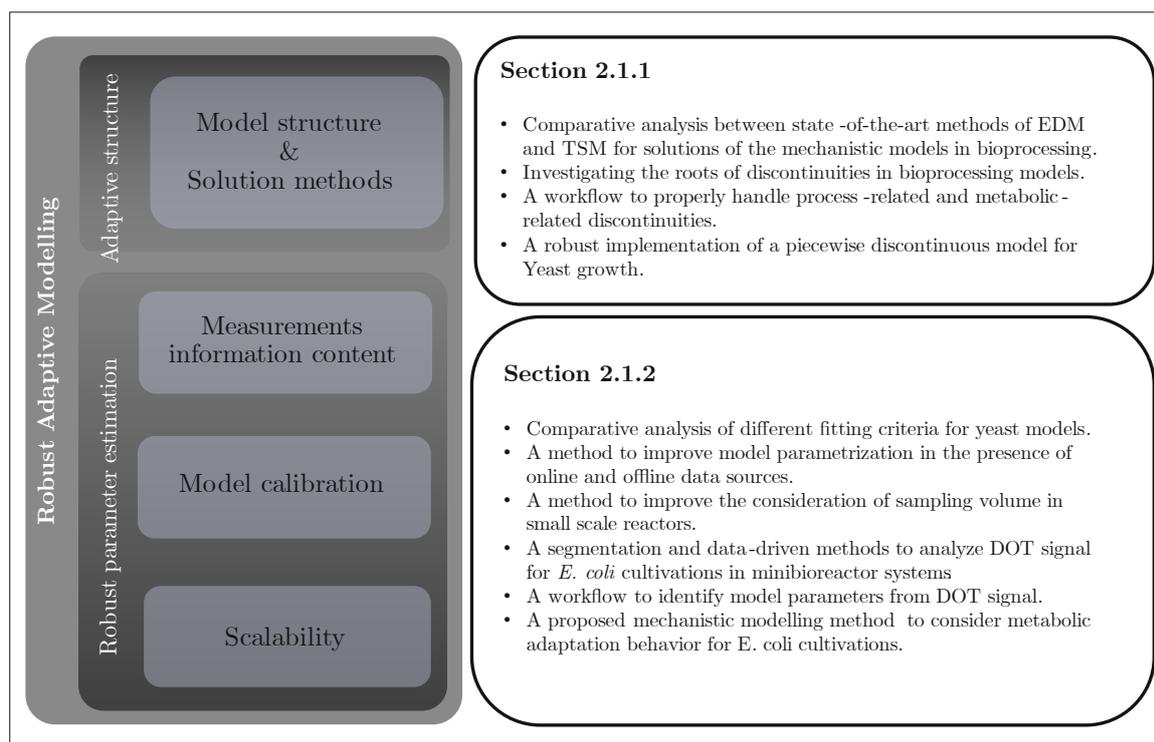


Figure 9: The thesis structure and main elements

The structure of the thesis is presented in Figure 9. The cumulative thesis contains five papers arranged into two main categories. However, a distinctive separation between these categories is not possible. The two challenges of achieving an adaptive and a robust model are strongly correlated. Model structure flexibility (adaptiveness) and the reliability of the model output (robustness) usually influence each other.

The first part (section 2.1.1) has the goal to highlight the importance of proper handling discontinuities. Therefore, a comparative analysis is provided in this part. Also, a workflow and modelling example of yeast in lab-scale are provided in this part. The conclusion of this analysis is seen on three levels: a summary on the origin of the discontinuous behavior in bioprocessing, a comprehensive analysis on the possible consequences of overlooking these discontinuities in bioprocessing, and a workflow to deal with discontinuities in bioprocessing models.

The second part (section 2.1.2) has the goal to achieve better model output in terms of accuracy and robustness. This section is a collection of different methods applied on yeast and *E. coli* cultivations in different scales to improve model calibration and offer more representative parameters that can work hand by hand with the methods shown in the first part.

## 2 Results

## 2.1 Achievements and findings

### 2.1.1 Accurate location of model events for adaptive modelling

#### 2.1.1.1 *Challenges*

Biological processes are difficult to describe by mathematical models, due to the complex biological nature (e.g., tens or thousands of biochemical reactions, different metabolic pathways ...), and the limited understanding of kinetic links between biology and process parameters. The complexity of the mechanistic models depends on the availability of process analytics. In unstructured and simplified structured models, internal reactions are often lumped together and represented as one overall metabolic pathway. Only the most important metabolic pathways are considered. In the context of industrial biotechnology, unstructured or simplified structured models are commonly used [63]–[65].

These models primarily focus on the description of simplified metabolic pathways transitions. However, in reality, transitions are mostly continuous, highly nonlinear, and dependent on intracellular mechanisms, but as dynamics happen in very different timescales, most of the transitions are simplified into discontinuous behaviors. Therefore, switches expressed as logical operations often need to be incorporated in the model [94]. This results in a discontinuous piecewise model. Examples of that are models that describe the Crabtree effects in yeast [95], and overflow metabolism in *E. coli* [96].

Discontinuous (piecewise) models can be mathematically expressed as a combination of a set of continuous differential equations with discontinuous right-hand side, and a set of time-dependent and/or state-dependent conditions.

#### 2.1.1.2 *State of the art*

There are two main approaches to solve ODEs with discontinuous right-hand side [97]: the time stepping method (TSM) and the event driven method (EDM).

The time stepping method: in which solvers assume sufficient smoothness of the right-hand side of the ODEs and rely on the local error estimator to control the step size and keep errors in the generated solution low [80]. This approach can fail or become inefficient in discontinuous and very stiff regions. This is because the solution in these regions does not fulfill the main assumption of smoothness [97].

The event driven method: uses events functions to locate discontinuities (called also events [97]) by defining discontinuity surfaces of the differential system. When the solution reaches a surface, an event is located. The result is the solution of a sequence of initial value problems (IVPs), described by differential equations and interspersed by instantaneous events that cause a discrete change to the initial value problem currently being solved [98].

State of the art IVP solvers in *MATLAB*'s [83], *Python* [99] or *SUNDIALS* [100] provide the option to monitor and locate time and/or state and parameter dependent event functions. To account for the complexity of different events and switches, [101] proposed a general formulation where classical propositional logic is used for the representation of state conditions as it can represent conjunctions and/or disjunctions of relational expressions effectively. This general formulation of models can be used in different modeling languages and software systems. A review on the applications of the formulation for the analysis of differential and hybrid (continuous/ discrete) systems is given by [102].

More information on the solution environments in GPROMS, Modelica/Dymola, Assimulo, deSolve, Mathematician are found in [103]–[108].

A nice review on the available software toolboxes on the solution of differential systems with time or state (and possible parameter) dependent event functions in the context of systems and computational biology is given by [94].

The successful implementation of EDM in other fields highlights the potential of proposing an EDM-based method to properly handle discontinuities in ODEs mechanistic models. Such a method has to be tailored to fit bioprocessing modelling needs.

### 2.1.1.3 Findings

#### Identifying the roots of discontinuities in a generic bioprocessing model

The following discontinuities which affect the solution are identified:

A- *Metabolic discontinuities*: these are the switches that happen between the metabolic pathways. They are located by monitoring the metabolic conditions. The inaccuracy in locating these may lead to deviations in the model output.

B- *Process-related discontinuities*: these are the sudden changes that happen in the process conditions or environment and have a direct effect on the cultivation. For example, when samples or substrates are withdrawn or fed suddenly (especially with intermittent bolus feeding strategies and lower reactor volumes), disturbances in reactor working volume and substrates concentrations happen. When the system is modelled, these disturbances in the corresponding states e.g., volume and substrates, should be modelled with events and conditions, where states are corrected after each change, otherwise the resulting errors might propagate and corrupt the model output.

C- *Discontinuities caused by highly nonlinear kinetic terms (Non-physical solutions)*: nonlinear kinetic such as Monod-like kinetics can exhibit stiff behaviors, especially when the affinity of the organism to the used substrates is high [30], indicated by a small affinity constant. For that, corresponding zero crossing conditions should be considered.

## Identifying the consequences of improper handling of discontinuities in bioprocessing models

The improper handling (or ignoring) model discontinuities in bioprocessing leads to a higher model prediction uncertainty, a higher uncertainty in locating the metabolic switches, a higher model parameters uncertainty, and possibly a lower parameter identifiability. The reason is that the optimization surface generated by the model predictions is distorted and affected by the noise and errors resulting from the above-mentioned discontinuities. A simplified video explanation is shown in 4.1.4.

## Proposing a general formulation for proper handling of discontinuities in bioprocessing models with an adaptive structure

A robust modelling approach has been developed for the growth of *S. cerevisiae*. This has been achieved by the consideration of metabolic switches as events in the framework of a well-established model [82]. A modelling tool (codes library) is built for *MATLAB* to achieve this purpose. The following generic mathematical description is proposed:

$$\dot{c}(t) = Y \cdot \mathcal{A}(C(t)) \cdot r(c(t), u(t), \theta) \quad (2.1)$$

where  $c(t) \in \mathbb{R}^{N_m}$  is the vector of time-dependent state variables,  $t \in \mathbb{R}$  is the independent variable time,  $u(t) \in \mathbb{R}^{N_u}$  is the time-varying input vector, and  $\theta \in \mathbb{R}^{N_p}$  is the parameter vector. The conversion matrix  $Y$  with dimension  $N_n \times N_m$  contains all stoichiometric coefficients for  $N_m$  reacting state and  $N_n$  reaction rate in the network;  $r(\cdot)$  is a reaction vector of dimension  $N_n$  containing the reaction rates. The adaptive switching between different pathways is considered a so-called activation matrix  $\mathcal{A}$  of dimension  $N_n \times N_n$ .  $\mathcal{A}$  is a diagonal matrix whose elements  $\{1,0\}$  are used to activate/deactivate kinetic reactions. Each diagonal element represents a discrete-time variable which activates/deactivates columns in  $Y$ . Changes of  $\mathcal{A}$  are triggered by events, which are located based on the monitoring of certain conditions  $cond_i$  which complement the model:

$$cond_i(c(t), u(t), \theta, t) = 0 \quad \text{for } i = 1, \dots, N_{cond} \quad (2.2)$$

Complementary details on the numerical solution are shown in appendix 4.1.1.

### 2.1.1.4 Impact of the work

The impact of this work can be measured on two levels:

Scientific level: highlighting the importance of proper handling of discontinuities in bioprocesses and proposing a method to properly handle discontinuities in bioprocessing models.

The proposed method improves solution numerical stability against numerical noise, measurements uncertainty, deviations in both initial model states and parameters, and improves the convergence of the optimization algorithm to the optimal solution.

Industrial level: improving the calibration of the models, possibly leading to less calibration runs by providing accurate model predictions and model parameters estimates. This allows for better model deployment in process monitoring, design and control.

### 2.1.1.5 Publications

Citation: Jouned, M. A., Kager, J., Herwig, C., & Barz, T. (2022). Robust modelling of *S. cerevisiae* fed-batch cultures by proper handling of model discontinuities. The 33rd VH-Yeast Conference, self-published by the *Research Institute for Baker's Yeast*, Berlin. Accepted/ to be published soon.

My contribution: Design and conduct the experiments and analytics with Julian Kager. Collect and pre-process the data. Building the model, designing the EDM workflow, performing the identifiability analysis, Monte Carlo simulation, DoE analysis and uncertainty quantification analysis. Writing the paper, managing other authors' contributions. Revising and proofreading.

Citation: Jouned, M. A., Kager, J., Herwig, C., & Barz, T. (2022). Event driven modeling for the accurate identification of metabolic switches in fed-batch culture of *S. cerevisiae*. *Biochemical Engineering Journal*, 180, 108345.

My contribution: Design and conduct the experiments and analytics with Julian Kager. Collect and pre-process the data. Preparing and building the numerics *MATLAB* Libraries with Tilman Barz. Building the model, designing the EDM workflow, performing the identifiability analysis, Monte Carlo simulation, DoE analysis and uncertainty quantification analysis. Writing the paper, managing other authors' contributions. Revising and proofreading with the rest of the authors.

## 2.1.2 Improve model calibration and scalability for robust modelling

### 2.1.2.1 Challenges

Model scalability in bioprocessing is generally associated with a change in values of the model parameter at each scale, therefore, modelling methods should accurately quantify the values in the parameter set  $P$  (i.e., model calibration), and the change  $\sigma$  at each scale to achieve a robust model across all scales.

Inaccuracies in parameters estimation in small scale reactors happen mainly because of the lack of sufficient samples to represent the system, i.e., low sampling frequency. However, unlike larger scales, also when high sampling frequency is possible, e.g., by the means of HTP and autosamplers, the accuracy of the parameters can be affected by the radical changes in the working volume because of the relatively low initial volume. In small scales like minibioreactor, and lab-scale bioreactors systems, the change in the working volume is not neglectable. Sudden changes in the volume due to withdrawal or addition of reacting component or reaction medium do have transient temporary effects on the system dynamics. The nonlinear, stochastic, and propagation properties of these effects can alter model predictions.

These obstacles show the need for methods that can accurately extract parameters information in both low and high sampling frequency scenarios.

A possible remedy, when the sampling frequency is low, is to exploit the information encoded in the already-existed semi-/continuous (online) data resources, e.g., off-gas signals for yeast [46] and dissolved oxygen tension signal for *E. coli*. However, the encoded metabolic activities in these signals (especially in case of *E. coli* cultivations) are ambiguous, and the separation of the useful characteristics from the background noise may be difficult. For the DOT, a combination of sensor time delay and high substrate affinity of the organisms shifts the metabolic activities nonlinearly along the signal [85].

On the other hand, when the sampling frequency is high, solution methods could overlook the high frequency characteristics of the volume signal and look mainly at the overall signal shape, therefore, working volume changes are generally smoothed.

However, in both scenarios a robust criterion to fit model predictions to the measurements of the samples is needed. For online measurements, this is far from being a simple task. Off-gas and dissolved oxygen online signals are usually more complex than other off-line/at-line discrete state measurements, contain both low and high frequency information, and have their own patterns of errors and uncertainties because of the completely different physical and chemical phenomena in their sensors. While for liquid samples, variance in the measurements is witnessed, off-gas and DOT signals suffer from shifts, offsets, and drifts. These are usually caused by different factors including possible sensor delay, possible interactions with other process conditions such as humidity, pH, and temperature.

The commonly applied criteria to fit model prediction to the online measurements use objective functions derived from the Euclidean Least-Squares (LSQ) distance. These criteria don't account for the peculiarities of the online signals. The standard LSQ criterion evaluates the

quadratic fitting error by comparing measurements and predictions point-by-point. Interesting features like kinks and sharp changes in the measured signals are not in the focus.

### 2.1.2.2 *State of the art*

To extract information from already-exist analytics, when the sampling frequency is insufficiently low, the following possibilities are there:

For yeast cultivations, off-gas information proved to be used to identify different metabolic pathways and provide information about some parameters like maximum growth rate [109]–[111].

For *E. coli* cultivations, many contributions reported on the strong correlation between metabolic activities of *E. coli* and certain segments of the DOT signal [112]–[117].

More information on the off-gas signals and DOT signal is provided in section 1.1.3.2.1.

A responsible remedy, when the sampling frequency is high, is to properly account for volume changes, so the volume change errors do not propagate and affect model accuracy.

In both cases, with a low or a high number of samples, a possible solution to overcome the problems associated with using objective function with least squared terms is to use a shape-matching algorithm which is commonly applied in other fields. Dynamic Time Warping is one of the most famous algorithms. It is used as a non-linear mapping tool between signals, which reduces the distance and matches the shape. [111] showed the applicability of DTW to identify different phases of off-gas signals in *S. cerevisiae*. However, this concept has not been integrated yet in a modelling method to match simulation and measurement signals and improve the accuracy of model parameters.

### 2.1.2.3 *Findings*

An analysis and a modelling method to segment DOT signal in minibioreactor systems are proposed and applied to 8 *E. coli* cultivations. The method accounts simultaneously for both high and low frequency characteristics.

As a result of the analysis, a DOT segment, not reported in literature, is observed. The corresponding metabolic state is referred to as the “adaptation state”. In this time window, the cell is likely to pause or attenuate the metabolism.

The quantitative analysis shows the possibility to obtain key model parameters from the DOT signal. The estimation of model parameters of the overflow switching condition was possible using only DOT signal and biomass samples.

A comparative analysis of yeast cultivations modelling with and without explicit consideration of the sudden changes in reactor volume revealed a tangible effect on model outputs and parameter estimation uncertainty. A simple method to account for volume changes correction based on the EDM concept is proposed.

A method combining Dynamic Time Warping and the Least-Squares algorithms is developed as a fitting criterion for the identification of *S. cerevisiae* fermentation models. The method is applied to different lab-scale yeast runs with a pre-planned feeding plan designed to trigger certain metabolic changes. The results showed clearly more reliable parameters estimates of the new methods compared to the standard LSQ fitting approach. It turned out that model predictions generated by only LSQ fitting tend to smooth the off-gas signals, leading to a loss of specific details of the signals shape that might represent important metabolic changes.

#### 2.1.2.4 *Impact of the work*

On a scientific level:

- Reveals a new dissolved oxygen signal segment (characteristics), linked to metabolic adaptations state.
- Provides a systematic method to analyze DOT signal for minibioreactor systems to monitor cell metabolism and calculate model parameters in *E. coli* cultures.
- Proposes a new method to consider the online signals of DOT and off-gas in the mechanistic models of *E. coli* and yeast.

On an industrial level:

- Improves the calibration of an industrially relevant mechanistic model for yeast cultivation to accurately locate metabolic changes, by integrating the commonly obtained off-gas signals.
- Improves the calibration of the model by providing a straightforward method to extract *E. coli* model parameters from the commonly obtained DOT signal.
- Provides model-based tools to better address MBRs issues of oxygen supply and intermittent feeding in *E. coli* cultivations.

Proper analysis and integration of the online signals generally increases the information content of the models, leading to a reduction in costs and time.

#### 2.1.2.5 *Publications*

Citation: Jouned, M. A., Kager, J., Rajamanickam, V., Herwig, C., & Barz, T. (2023). A unique response behavior in the dissolved oxygen tension signal of *E.*

coli in minibioreactor system equipped with intermittent bolus feeding. MDPI Bioengineering Journal. Submitted.

My contribution: Design the experiments with Julian Kager. Collect and pre-process the data. Propose the idea of the analysis and segmentation algorithms, write required *MATLAB* code libraries for the analysis and modelling. Writing the paper, managing other authors' contributions. Revising and proofreading.

Citation: Jouned, M. A., Kager, J., Herwig, C., & Barz, T. (2021). Event driven analysis to enhance model calibration of experiments with high offline sampling rates. In *Computer Aided Chemical Engineering* (Vol. 50, pp. 463-468). Elsevier.

My contribution: Design the experiments with Julian Kager. Collect and pre-process the data. Prepare and build the numerics *MATLAB* Libraries with Tilman Barz. Build model, design the EDM workflow, identifiability analysis, Monte Carlo simulation, DOE analysis and uncertainty quantification analysis. Writing the paper, managing other authors' contributions. Revising and proofreading with the rest of the authors.

Citation: Jouned, M. A., Kager, J., Herwig, C., & Barz, T. (2020). Improving the Calibration of Kinetic Growth Models using Dynamic Time Warping. In *Computer Aided Chemical Engineering* (Vol. 48, pp. 1651-1656). Elsevier.

My contribution: Design and conduct the experiments with Julian Kager. Collect and pre-process the data. Propose the idea, conduct the comparative analysis. Writing the paper, managing other authors' contributions. Revising and proofreading.

## 2.2 Full manuscripts

### 2.2.1 Improving the Calibration of Kinetic Growth Models using Dynamic Time Warping

# Improving the Calibration of Kinetic Growth Models using Dynamic Time Warping

Mhd Adnan Jouned<sup>a,b</sup>, Julian Kager<sup>a,c</sup>, Judit Aizpuru<sup>b</sup>, Christoph Herwig<sup>a,c</sup>, Tilman Barz<sup>b</sup>

<sup>a</sup> ICEBE, TU Wien, Gumpendorfer Straße 1a 166/4, 1060 Wien, Austria

<sup>b</sup> AIT Austrian Institute of Technology GmbH, Giefinggasse 2, 1210 Wien, Austria

<sup>c</sup> CD Laboratory on Mechanistic and Physiological Methods for Improved Bioprocesses, TU Wien, Gumpendorfer Straße 1a 166/4, 1060 Wien, Austria  
[tilman.barz@ait.ac.at](mailto:tilman.barz@ait.ac.at)

## Abstract

Off-gas measurements give valuable information on the respiratory activity of organisms during fermentation processes. Measured oxygen consumption and carbon dioxide production is usually linked to the overall metabolic activity of the cultivated cells. Together with offline measured nutrient and metabolite concentrations reaction parameters of growth models can be determined. Standard algorithms for parameter estimation use the least-squares (LSQ) error criterion for fitting model predictions to measured data. However, their application does not necessarily yield off-gas representative model predictions and parameters. This is especially true for off-gas signals with rapid variability and corresponding sharp bends and kinks. Off-gas signals include clear indicators for nutrient limitations and metabolic shifts of the culture. Using the LSQ error criterion the fitting tends to smooth out these informative details leading to poor model predictions and parameter estimates.

This contribution presents a comparative analysis of the performance of standard nonlinear LSQ algorithms and an adapted algorithm using the Dynamic Time Warping (DTW) criterion. Both algorithms are applied to fit off-gas signals for the calibration of the kinetic model of *Saccharomyces cerevisiae* (Sonnleitner and Käppeli 1986) on three experimental datasets. The data represents high dynamics with rapid variations and covers yeast fermentation through Batch and Fed-Batch phases including time windows where the organisms are forced to produce ethanol through the “Crabtree effect” by overfeeding. It turns out that, compared to results using LSQ criterion, the application of the DTW criterion yields a better shape matching of the data. In addition, results are also discussed comparing the performance in terms of convergence to the best fitting parameters and the robustness of algorithms against structural modelling errors.

**Keywords:** dynamic programming, signals matching, parameter estimation, kinetic modelling

## 1. Introduction

Kinetic modeling plays an essential role in bioprocess development because it provides not only information about changing quantities and rates, but also gives valuable insights about the underlying reactions scheme. Hence, the parameters reflect biological meaning apart from being mathematical coefficients. Model calibration is considered to be a complex task especially for problems with a high number of interdependent parameters and a low number of samples. In addition to that, available information is often

concentrated in certain time ranges or only available at distinct time points. For example, for practical reasons it is often the case that offline samples are concentrated in the Fed-batch phase while no/less data is available for the Batch-phase. The parameters estimated in these cases may not reveal the actual underlying behavior and may result in “Observation Biased” models where the quality of the model in experimental design and control can be questionable. This issue can be improved when semi-/continuous (online) data resources are introduced such as off-gas information or spectral information (Golabgir and Herwig 2016).

In yeast fermentations processes off-gas information proved to be used to identify different metabolic pathways and to provide information, like maximum growth rate (Petkov and Davis 1996, Anderlei et al. 2004, Gollmer and Posten 1996). On the other hand, off-gas signals are usually more complex than other states and contain both low and high frequency information where rapid changes represent mostly certain limitations or metabolic shifts and slower changes are usually correlated to the respiration of the culture. Moreover, off-gas signal errors arise completely from different sources compared to other states such as component concentrations. While for concentration samples, variance in the measurements is witnessed, off-gas signal suffers - among different errors types - from shifts, offsets and drifts. These usually caused by different factors including possible sensors delay, not-proper accounting for humidity and high interactions with other conditions/states such as pH and temperature (Frick and Junker 1999). Fitting criteria able to account for these characteristics are necessary to obtain representative model parameters.

Parameter estimation algorithms using Least-Squares (LSQ) as a criterion to fit off-gas signals, could potentially perform better if another criterion that accounts for these characteristics is used. The reason is that the standard criterion evaluates the quadratic fitting error by comparing measurements and predictions point-by-point. Interesting features like kinks and sharp bends in the measured signals are not in the focus. Accordingly, in the presence of structural model simplifications and measurement noise, off-gas signals are usually fitted by smooth curves (which are optimal in the sense of the quadratic error criterion, but do not mimic the shape of the measured signals). A possible solution to this is parameter estimation implementing an error criterion derived from Dynamic Time Warping (DTW) method (Gollmer and Posten 1996) (Srinivasan and Qian 2007). The method is applied in shape recognition. It has a high potential for yielding model predictions, which mimic the interesting features in the measured signals, and to overcome the limitations of algorithms using LSQ error criterion.

The applicability of DTW as a non-linear mapping tool between signals, which reduces the distance and matches the shape, has been shown in the field of chemistry and bioprocessing (Srinivasan and Qian 2007; González-Martínez, Ferrer, and Westerhuis 2011). (Gollmer and Posten 1996) actually used DTW to identify different phases of off-gas signals in *S. cerevisiae* fermentation two decades ago, but according to author knowledge, this concept has not been exploited to match simulation/observation signals to improve model parameters estimation consequently. This work exemplarily validates this improvement by a comparative analysis of the performance of standard nonlinear LSQ algorithms and an adapted algorithm using the Dynamic Time Warping (DTW) criterion.

## 2. Materials and Methods

To understand the effect of using an error criterion derived from DTW on parameters estimation, two algorithms with different error criteria are used to fit three different

experimental datasets. The well-known model for baker's yeast fermentation (Sonnleitner and Käppeli 1986) is used where we try to get best parameter-set ( $q_{s,max}, \mu_{ethanol,max}, q_{O_2,max}, K_S, Y_{biomass/glucose}^{oxidative}, Y_{biomass/glucose}^{reductive}, Y_{biomass/ethanol}$ ). The selection of the parameters and initial values has been done recursively based on importance ranking and sensitivity information (López et al. 2013) (Ulonska et al. 2018). Details on the model structure, nomenclature and parameter values are shown in the original paper (Sonnleitner and Käppeli 1986).

Around thousand initializations with various initial guesses of selected parameter-set were established, passed to an optimizer ("*fminsearch*", *MATLAB R2017b*, stopping criterion is set to 100 iterations, all other options are set to default) to find parameter values yielding lowest error criterion.

In order to find a reference base to compare both results, the solutions from both algorithms are accepted among best fitting parameters-set, when the metabolic states calculated back from the model match the ones that have been pre-identified by experts, error of the state estimation is between 0-5% NRMSD and the error of off-gas signals estimation is between 0-50% NRMSD of any off-gas signal of each metabolic state.

The three fermentation experiments consist of Batch and Fed-Batch phases where overfeeding is applied after some hours in the Fed-Batch phase to force the cells deliberately to produced ethanol through "Crabtree effect". Data presented in Figure 1 shows clearly high dynamics in the off-gas signals with some rapid changes reflecting the time-varying behavior of the cells.

### 2.1. *S. cerevisiae* fermentation model

In this model derived from (Sonnleitner and Käppeli 1986) the growth is described on two substrates glucose and ethanol with fermentative and oxidative pathways, based on three metabolic pathways with correspondent yield parameters  $Y_{biomass/glucose}^{oxidative}, Y_{biomass/glucose}^{reductive}$  and  $Y_{biomass/ethanol}$ . Total growth is the growth based on all forms of biomass specific intake ( $q_s^{oxidative}, q_s^{reductive}, q_{ethanol}$ ) as:

$$\mu_{total} = Y_{biomass/glucose}^{oxidative} \cdot q_s^{oxidative} + Y_{biomass/glucose}^{reductive} \cdot q_s^{reductive} + Y_{biomass/ethanol} \cdot q_{ethanol} \quad (1)$$

Mass balances equations are ( $x$ : biomass,  $s$ : glucose,  $e$ : ethanol,  $V$ : volume and  $F$ : feed)

$$\begin{aligned} \frac{dV}{dt} &= F_S + F_{Base} + F_{Acid} - F_{Gas} \\ \frac{dC_x}{dt} &= \mu_{total} \cdot C_x - \frac{F_S}{V} \cdot C_x \\ \frac{dC_s}{dt} &= -(q_s^{reductive} + q_s^{oxidative}) \cdot C_x - \frac{F_S}{V} \cdot C_s + \frac{F_S}{(1/C_{s,in})V} \\ \frac{dC_e}{dt} &= (q_s^{reductive} - q_e) \cdot C_x - \frac{F_S}{V} \cdot C_e \end{aligned} \quad (2)$$

Additionally, off-gas equations of Carbon Dioxide Evolution Rate (CER) and Oxygen Uptake Rate (OUR) can be derived from the original model taking into consideration elemental balance, which can be derived from oxidative and reductive reaction stoichiometry:

$$CER = (q_s^{oxidative} \cdot Y_{s,ox}^{CO_2} + q_{ethanol} \cdot Y_e^{CO_2} + q_s^{reductive} \cdot Y_{s,Red}^{CO_2}) C_x \cdot V \quad (4)$$

$$OUR = (q_s^{oxidative} \cdot Y_s^{O_2} + q_{ethanol} \cdot Y_e^{O_2}) C_x \cdot V \quad (3)$$

## 2.2. Objective functions

Two objective functions are formulated, in which the first uses standard LSQ. The second uses a combination of LSQ and DTW derived terms. The two functions differ from each other by how they calculate the similarity of the off-gas signals, i.e. the sampled and simulated *OUR*, *CER*.

### 2.2.1. LSQ criterion

For  $M$  samples,  $N = N_L + N_G$  liquid and gas states, and  $\theta$  as unknown parameter vector, LSQ is defined as  $\phi^{LSQ}$ . Assuming all model states have the same weight in the objective function the fitting problem reads:

$$\min_{\theta} \phi_{L+G}^{LSQ}(\theta); \quad \text{with} \quad \phi_{L+G}^{LSQ}(\theta) = \sum_{i=1}^M \sum_{j=1}^{N_L+N_G} (Y_{i,j}^m - Y_{i,j}(\theta))^2 \quad (5)$$

### 2.2.2. Combined LSQ and DTW criterion

Replacing  $\phi_G^{LSQ}(\theta)$  by  $\phi_G^{DTW}(\theta)$  where all deviations from off-gas measurements are calculated based on DTW. For both, *OUR*, *CER*, two sequences are needed to build DTW distance matrix. The matrix size is defined by the sizes of both sequences  $Y_G^m$  and  $Y_G(\theta)$ , in which each matrix element  $\varepsilon_{i,j}$  represents the distance according to the chosen metric, which is in our case the squared Euclidian distance. To get shape preserved matching between two sequences DTW seeks an optimal path  $k$  through the matrix which minimize the warping cost (Ratanamahatana and Keogh 2004)

$$\phi_G^{DTW} = \min_k \left( \sum_{i,j} \gamma_{i,j} \right)_k \quad (6)$$

Where  $\gamma$  are the elements of the warping path. The optimal path is found by calculating the minimum cumulative distance of the current element and the other three adjacent cells in DTW matrix

$$\gamma_{i,j} = \varepsilon_{i,j} + \min(\gamma_{i-1,j-1}, \gamma_{i-1,j}, \gamma_{i,j-1}) \quad (7)$$

Using  $\alpha$  as a weighting coefficient to scale the DTW term to the same magnitude of the LSQ term, the combined objective function reads

$$\min_{\theta} \left( \phi_L^{LSQ}(\theta) + \alpha \cdot \phi_G^{DTW}(\theta) \right) \quad (8)$$

## 3. Results and discussion

Figure 1 shows model estimation results of two experiments after 100 optimization iterations. While the fitting of the states is similarly good for both objectives, larger differences can be seen when examining the off-gas signals and by comparing the active metabolic states. We can notice that the metabolic states sequence differs. It can be seen that with standard LSQ criterion the optimizer overlooks some intermediate details in order to get a good fitting along the whole time horizon.

This is problematic, as different metabolic states are assigned along the process, where metabolic states are indicated by sharp changes in the off-gas signal.

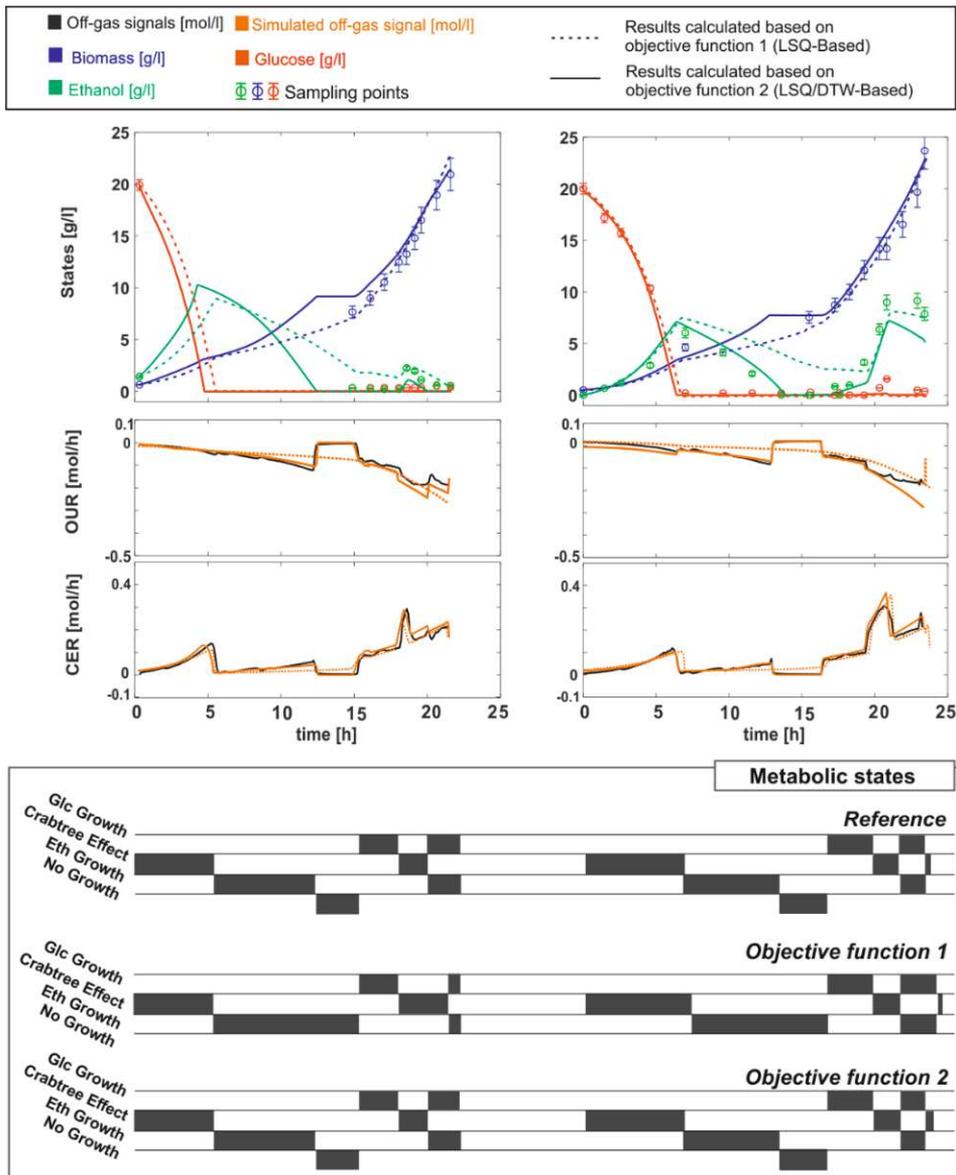


Figure 1: Model fits two experimental datasets out of three after 100 optimization iterations with standard LSQ-based objective function (dashed line) and LSQ/DTW-based objective function (solid line). The corresponding metabolic states are presented below with the reference solution (identified by an experienced field expert based on the visual inspection of CER, OUR signals and offline measurements).

The length of a metabolic state is strongly determined by the parameters related to maximum reaction rates and the corresponding conversion coefficients (yields), which needed to be correctly assigned during model parametrization. Figure 2 shows clearly that the algorithm using DTW/LSQ criterion had a twice higher success rate in finding the exact parameters out of 1000 model calibrations, each with differently perturbed initial parameter sets.

#### 4. Conclusion

Using DTW (Dynamic Time Warping) as a fitting criterion for the identification of *S. cerevisiae* fermentation models clearly leads to more reliable parameters estimates compared with the standard LSQ fitting approach. This has been quantitatively proved for three experimental datasets with different metabolic states. For the presented case study, it is shown that model predictions generated by LSQ fitting tend to smooth out measured off-gas signals losing specific details of the signals shape that might represent important metabolic changes. The results clearly indicate a superior performance using DTW, i.e. the more accurate shape matching of the signals yields improved model predictive performance and provides more accurate model parameters.

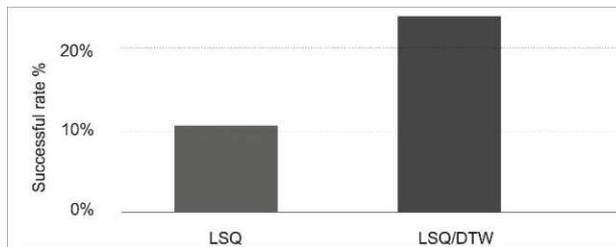


Figure 2: Success rate implementing the LSQ and the combined LSQ/DTW objective functions for parameter identification from off-gas and sampling data.

#### Acknowledgment

This work was partially funded by the Austrian Research Funding Association (FFG) within the program Bridge 1 in the project "AdaMo" (No. 864705).

#### References

- Frick, and Junker. 1999. "Indirect Methods for Characterization of Carbon Dioxide Levels in Fermentation Broth." *J. Biosci. Bioeng.* 87 (3): 344–51.
- Golabgir, and Herwig. 2016. "Combining Mechanistic Modeling and Raman Spectroscopy for Real-Time Monitoring of Fed-Batch Penicillin Production." *Chemie Ing. Tech.* 88 (6): 764–76.
- Gollmer, and Posten. 1996. "Supervision of Bioprocesses Using a Dynamic Time Warping Algorithm." *Control Eng. Pract.* 4 (9): 1287–95.
- González-Martínez, Ferrer, and Westerhuis. 2011. "Real-Time Synchronization of Batch Trajectories for on-Line Multivariate Statistical Process Control Using Dynamic Time Warping." *Chemom. Intell. Lab. Syst.* 105 (2): 195–206.
- López, Barz, Peñuela, Villegas, Ochoa, and Wozny. 2013. "Model-Based Identifiable Parameter Determination Applied to a Simultaneous Saccharification and Fermentation Process Model for Bio-Ethanol Production." *Biotechnol. Prog.* 29 (4): 1064–82.
- Ratanamahatana, and Keogh. 2004. "Everything You Know about Dynamic Time Warping Is Wrong." *Third Work. Min. Temporal Seq. Data*, no. January 2004: 22–25.
- Sonnleitner, and Käppeli. 1986. "Growth of *Saccharomyces Cerevisiae* Is Controlled by Its Limited Respiratory Capacity: Formulation and Verification of a Hypothesis." *Biotechnol. Bioeng.* 28 (6): 927–37.
- Srinivasan, and Qian. 2007. "Online Temporal Signal Comparison Using Singular Points Augmented Time Warping." *Ind. Eng. Chem. Res.* 46 (13): 4531–48.
- Ulonska, Kroll, Fricke, Clemens, Voges, Müller, and Herwig. 2018. "Workflow for Target-Oriented Parametrization of an Enhanced Mechanistic Cell Culture Model." *Biotechnol. J.* 13 (4): 1–11.

## 2.2.2 Event driven analysis to enhance model calibration of experiments with high offline sampling rate

## Event driven analysis to enhance model calibration of experiments with high offline sampling rates

M. Adnan JOUNED<sup>a,b</sup>, Julian KAGER<sup>a</sup>, Christoph HERWIG<sup>a</sup>, Tilman BARZ<sup>b</sup>

<sup>a</sup> ICEBE, TU Wien, Gumpendorfer Straße 1a 166/4, 1060 Wien, Austria

<sup>b</sup> AIT Austrian Institute of Technology GmbH, Giefinggasse 2, 1210 Wien, Austria

### Abstract

The use of autosamplers connected to high throughput analytical devices allows for a high sampling frequency and analytics with reduced manual labor, leading to better process characterization (Maurer et al., 2015; Hofer et al., 2020). Increased sampling often leads to a significantly increased information content in the generated data. However, in combination with miniaturized or lab-scale reactors, the effect of volume change by the frequent sampling becomes challenging.

Sampling leads to fast, almost instantaneous volume changes in the reactor. This process represents a discontinuous behavior in the continuous-time kinetic model. A commonly applied so-called “time stepping” method ignores the discrete behavior and relies on the solver’s local error estimator to solve continuous-time differential equations. Therefore, in regions where discontinuities of the solution or its derivative occur, the method may fail to deliver an accurate solution. An alternative is the so-called “event driven” method, which explicitly accounts for discontinuities in the model. During the solution (integration) of the model, the method accurately locates time points, where discontinuities occur (event detection), e.g., volume changes and continues the solution process (Dieci and Lopez, 2012). It is well-known that proper handling of discontinuities can significantly increase the models' accuracy and reduce simulation runtime (Alsoudani, 2016). Still, bioprocess developers often ignore or tolerate discontinuities when implementing models in simulation software. Hence, this contribution highlights the importance of a proper handling of discontinuities in a relevant common case study for bioprocess development. Results are presented for the determination of kinetic parameters of Sonnleitner and Käppli’s (1986) *Saccharomyces cerevisiae* growth model on a lab-scale fed-batch process with fast volume changes caused by frequent sampling. It turns out that the “time stepping” method misses several volume changes. In contrast, the “event driven” method does not. Accordingly, the “event driven” method yields accurate model predictions which are not affected by the reactor volume’s prediction error and thus improves the model calibration, lowers parameters uncertainty, and supports a robust convergence to the best fitting model parameters.

**Keywords:** parameter estimation, kinetic modelling, event driven modelling

### 1. Introduction

The acceleration of bioprocess development for decreasing time to market (TTM) of biopharmaceuticals has not only been known since the COVID-19 pandemic. For this purpose, mechanistic models represented by systems of ordinary differential equations (ODE’s) are indispensable tools for bioprocess design, monitoring, and control (Narayanan et al., 2020). However, the challenge of underfitting these models is present when an insufficient number of observations is used for model calibration. This problem

usually leads to correlated model parameters, which could hinder clear interpretations of the results. Autosamplers and automated analytical devices allows for tackling this challenge by increasing the sampling frequency without additional manual labor (Maurer et al. 2015; Hofer et al. 2020). Frequent sampling inevitably leads to significant volume changes in the reactor, which besides being a bottleneck in miniaturized systems, needs to be properly considered during modelling steps.

Volume changes of an ideal stirred tank reactor are usually modeled by considering mass balance equations. The changes in the volume over time is calculated as the difference between input and output flow rates (Doran, 2012). It is widely accepted to consider sampling volume  $F_{\text{Sampling}}$  as a part of the flows that are leaving the reactor (Rocha, 2003; Callewaert and De Vuyst, 2000; Kager et al., 2020). Hence, volume changes are written usually as (where  $V$ : volume,  $F$ : liquid mass flow rates):

$$\frac{dV}{dt} = F_{\text{in}} - F_{\text{out}} = F_{\text{Feed}} + F_{\text{Base}} + F_{\text{Acid}} - F_{\text{Sampling}} \quad (1)$$

In usual fed-batch fermentation, the measured flow rates (except  $F_{\text{Sampling}}$ ) usually show comparatively smooth curves. These curves are represented by discrete signals and can be transformed to smooth functions with relatively little effort, e.g., by applying a smoothing filter and by subsequent interpolation using piece-wise spline interpolation. Avoiding discontinuities on the right-hand-side of eq. 1 can significantly improve the efficiency and accuracy of its solution (Alsoudani, 2016).  $F_{\text{Sampling}}$  represents strongly discontinuous curves defined by (negative) pulse signals. A transformation of these sharp peaks by smoothing is not a viable solution. The result is highly nonlinear terms, which would require an (inefficient) dense time grid for accurate integration of eq. 1.

In the “time stepping” method (TSM), which is implemented by using standard ODE solvers, the sampling volume is calculated by the integration of eq. 1 considering a (negative) pulse signal. The method relies on the solver error estimator to determine the step size (Dieci and Lopez, 2012). Hence, in regions where discontinuities of the solution or its derivative occur, i.e., sampling timepoints, there is a probability to miss certain events such as sampling volume, causing inaccurate volume calculation. This is exemplarily shown in Figure 1. This results in a wrong volume mass balance and therefore affects subsequent calculations. The extent of the errors resulting from improper handling of samples volumes depends on the ratio of the sampled volume to the reactor volume. The errors are expected to have a bigger influence on smaller platforms, e.g., miniaturized bioreactor systems (< 0.3 L), where often no reactor volume measurement is available.

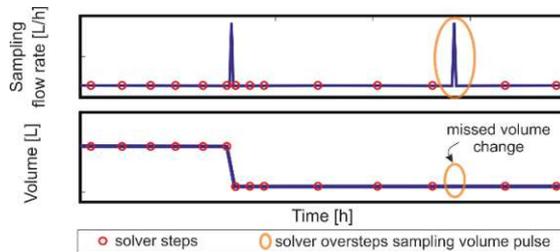


Figure 1: Modeling of volume changes due to sampling. Sampling volumes are considered by a “sampling” flow rate represented by a flow pulse signal. The time stepping method (TSM) does not accurately track the sudden changes in the flow pulse signal. The ODE solver steps miss the second pulse (overstepping problem), which has an immediate effect on the volume.

In contrast, the “event driven” method (EDM) accurately locates time instances (so-called events) where instantaneous sampling happens. At these time points, the integration is stopped, the volume is updated, and the integration is restarted with the updated volume as initial condition. By this, the EDM efficiently prevents missing any sample, leading to a correct and robust volume balance during model simulations.

It is noted that established simulation software like *gProms* and *Modelica/Dymola* use built-in routines to automatically detect discontinuities, locate events, and restart integration (Process Systems Enterprise Limited, 2013; Dynamic Modeling Laboratory, 2004). However, in low-level modeling languages such as the frequently used *MATLAB* (*ODE Suite*, 2020) or *Python* (*SciPy* package, 2020), the proper handling of discontinuities needs tailored solutions and special programming efforts as the ODE solvers provide only the basic functionalities by the so-called “event functions”.

## 2. Materials and Methods

### 2.1. Cultivation process and Sampling

A *S. cerevisiae* fermentation process is considered as an experimental case study. Samples were withdrawn from the fermentation medium at irregular time intervals either by-hand (10 samples of roughly 20 ml per sample) or automatically using (Numerica from Securecell) autosampler (20 samples of roughly 7 ml per sample) and distributed along the time of the experiment. The reactor's initial volume was 1.5 liter. The experiment consists of a batch and a fed-batch phase with different feed regimes.

### 2.2. *S. cerevisiae* fermentation model

The model from (Sonnleitner and Käppeli, 1986) considers the growth on glucose and ethanol substrates. It describes fermentative and oxidative growth based on all forms of biomass specific intake ( $q_s^{oxidative}$ ,  $q_s^{reductive}$ ,  $q_{ethanol}$ ) using ( $Y_{biomass/glucose}^{oxidative}$ ,  $Y_{biomass/glucose}^{reductive}$ ,  $Y_{biomass/ethanol}$ ) yields parameters. Total growth is written as:

$$\mu_{total} = Y_{biomass/glucose}^{oxidative} \cdot q_s^{oxidative} + Y_{biomass/glucose}^{reductive} \cdot q_s^{reductive} + Y_{biomass/ethanol} \cdot q_{ethanol} \quad (2)$$

Mass balances equations are ( $x$ : biomass,  $s$ : glucose,  $e$ : ethanol,  $F$ : flow rate)

$$\begin{aligned} \frac{d C_s}{dt} &= -(q_s^{reductive} + q_s^{oxidative}) \cdot C_x - \frac{F_{in}}{V} \cdot C_s + \frac{F_{Feed}}{V} \cdot C_{s,in} \\ \frac{d C_e}{dt} &= (q_s^{reductive} - q_e) \cdot C_x - \frac{F_{in}}{V} \cdot C_e \\ \frac{d C_x}{dt} &= \mu_{total} \cdot C_x - \frac{F_{in}}{V} \cdot C_x \end{aligned} \quad (3)$$

### 2.3. Time-stepping method (TSM) versus event-driven method (EDM)

Eq. 1 is used to model volume changes in both methods. In EDM,  $F_{Sampling}$  is omitted from eq. 1, and an external algebraic equation is used instead to account for volume changes. Figure 2 illustrates the working principles of the TSM and EDM. In the EDM, to accurately locate each sampling timepoint  $t_s$ , the simulation time interval is separated into  $k$  sub-intervals, where for each sampling point 1, 2, 3 ...  $K$ , the volume is corrected outside the ODE system using the additional algebraic equation;  $V = V - \Delta V$ . Following

the integration is restarted with the updated volume. In TSM, the sampling volume rate is integrated into eq. 1 without special treatment. The same initial conditions applied to both methods.

#### 2.4. Convergence analysis of parameter estimation

To highlight the negative effects of improper handling of sampling volume, the model predictions are fitted to the experimental data. The performance of the fitting is assessed by applying either TSM or EDM. The following parameters

$$\theta = [q_s^{max}, q_{O_2}^{max}, Y_{biomass/glucose}^{oxidative}, Y_{biomass/glucose}^{reductive}, Y_{biomass/ethanol}]$$

are selected for estimation based on the local parameter sensitivities and an identifiability analysis based on importance ranking (López et al., 2013).

The robustness of both methods is assessed through a Monte Carlo approach. 500 uniform distributed initial parameter guesses are generated over a +/- 25% interval around the literature's nominal values (Sonnleitner and Käppeli, 1986). For each parameter realization, an estimation problem was solved using *MATLAB R2017b "ODE15s"* solver and the nonlinear fitting algorithm "*lsqnonlin*". Normalized residual sum of squares (NRSS) between model predictions and measurements was used as an objective function.

### 3. Results and Discussion

Figure 3 (right) shows a visual comparison of one simulation run to highlight the differences in the calculated volume using both methods. It can be noticed that TSM oversteps certain sampling times and consequently doesn't update volume correctly. The wrong volume affects the other model states (concentrations) described in eq. 3 and displayed in Figure 3 (right) for two identical model simulations. The error becomes more pronounced towards the end of the simulation as its effect accumulates over time.

In TSM, to account for sampling volume changes correctly, solver steps must exactly locate sampling times. However, TSM locates solver steps based on the integrator's local error estimator, which is a tool to control approximation error at each step. If the solver oversteps a sampling time interval, the local error estimator does not indicate an approximation error. Consequently, sampling instances are located by chance. Therefore, by evaluating the model at slightly different initial conditions or with slightly different parameter values, the solver may overstep very different sets of sampling timepoints. The accuracy of model simulations might be strongly affected. This is also critical for model parametrization, as the deviations in model volume and predictions are different at each optimization run, which adds artificial noise to the optimization problem. This is shown in Figure 3 (left).

In contrast, EDM accurately locates all sampling events and uses an external algebraic equation to correct the volume and to restart the integration at each sampling timepoint. By doing this, it suppresses the noise and provides accurate and reproducible model simulations. This behavior is illustrated in Figure 3 (left), which shows the optimization

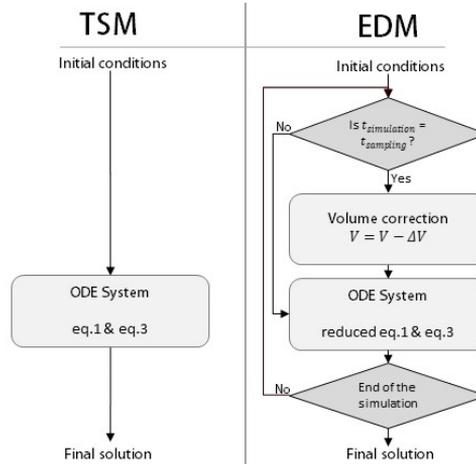


Figure 2: TSM and EDM for handling instantaneous volume changes because of sampling.

surface for two selected parameters. The noisy nature of the surface in TSM reduces the effectiveness of the gradient-based optimizer to seek an optimum. Tables 1 & 2 highlight this observation by comparing the means of prediction errors of the calibrated models obtained by both methods. The two-tailed p-value of the t-test is less than 0.0001, indicating a strong statistical significance. The results clearly indicate a higher probability to obtain better predictions using EDM and to converge to the best fitting parameter estimates. Reducing the solver tolerance in TSM might reduce these effects but not completely avoid them, as the error estimator still does not directly address the underlying reason of the problem, aside from increasing the computational cost.

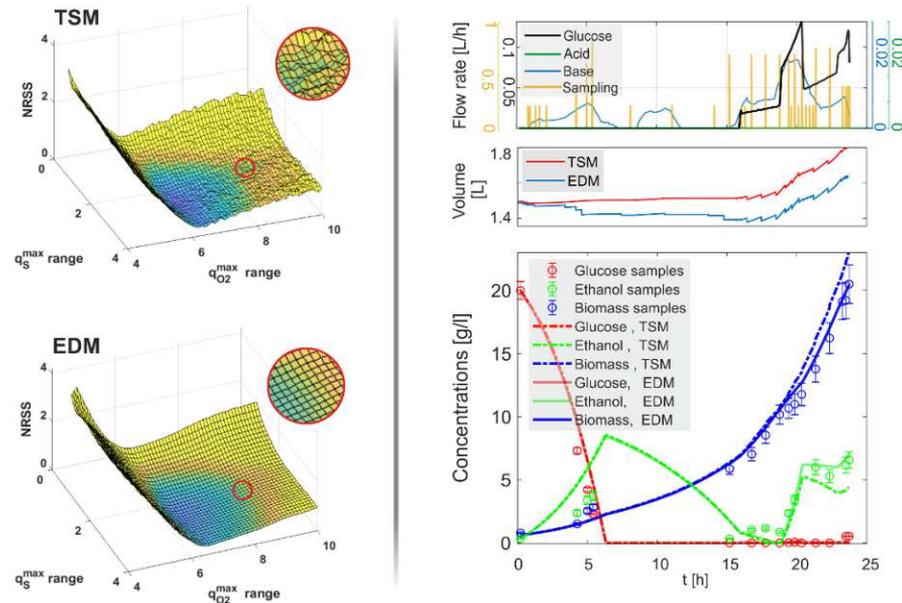


Figure 3: Left) States, flow rates and volumes calculated by the same model and parameters using EDM and TSM. Using TSM certain sampling times are overstepped. EDM delivers accurate results by accurately locating sampling times. Right) Optimization (NRSS) surfaces are smoother in case of EDM compared to standard TSM, allowing for a better convergence and efficiency of gradient-based optimization algorithms.

Table 1 & 2: Results of Monte Carlo procedure ( $n=500$ ). EDM shows a statistically significant lower mean of the model prediction errors.

		<i>TSM</i>	<i>EDM</i>			
<i>Mean of prediction error</i>		3.1	2.3			
<i>Standard deviation (<math>\sigma</math>)</i>		1.65	1.35			
		$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_5$
<i>Mean of Estimated. <math>\theta</math></i>	TSM	3.28	6.81	0.46	0.05	0.70
	EDM	3.30	6.56	0.44	0.05	0.69
<i>95% confidence interval</i>	TSM	0.97	2.50	0.16	0.018	0.24
	EDM	0.92	2.25	0.14	0.016	0.24

#### 4. Conclusion

Using EDM to consider sampling volume in *S. cerevisiae* fermentation models delivers more reproducible and accurate model predictions. Moreover, for model calibration, the quantitative analysis reveals a significantly improved convergence of the parameter estimation algorithm. Accordingly, EDM results are obtained with a lower prediction error and parameters uncertainty. The reason is a smoother optimization surface that leads to more robust convergence to the best fitting model parameters.

EDM is ready-for-use in high-level symbolic modelling languages such as *gProms* and *Modelical/Dymola*. In low-level languages, e.g., *MATLAB* or *Python*, EDM can be implemented with a reasonable effort using available initial value solvers with event detection. Hence, this method is simple to implement for fermentation processes with high sampling rates and platforms where explicit weight measurements are crucial but not available or hard to obtain, e.g., mini multi-bioreactor systems. This will allow for more consistent results, resulting in reduced iterations in bioprocess development and, therefore, a decreased time to market (TTM) of biopharmaceuticals.

#### Acknowledgment

This work was partially funded by the Austrian Research Funding Association (FFG) within the program Bridge 1 in the project "AdaMo" (No. 864705).

#### References

- Alsoudani. 2016. *Discontinuities in Mathematical Modelling : Origin , Detection and Resolution. University College London Department.*
- Callewaert & De Vuyst. 2000. "Bacteriocin Production with *Lactobacillus Amylovorus* DCE 471 Is Improved and Stabilized by Fed-Batch Fermentation." *Applied and Environmental Microbiology* 66 (2): 606–13.
- Dieci & Lopez. 2012. "A Survey of Numerical Methods for IVPs of ODEs with Discontinuous Right-Hand Side." *Journal of Computational and Applied Mathematics* 236 (16): 3967–91.
- Doran. 2012. *Bioprocess Engineering Principles: Second Edition. Bioprocess Engineering Principles: Second Edition.* Vol. 9780080917. Academic Press.
- Dynamic Modeling Laboratory. 2004. "Dymola." *Dynasim AB*. Lund, Sweden: Dynasim AB.
- Hofer, Kroll, Barmettler, & Herwig. 2020. "A Reliable Automated Sampling System for On-Line and Real-Time Monitoring of CHO Cultures." *Processes* 8 (6): 637.
- Kager, Tuveri, Ulonska, Kroll, & Herwig. 2020. "Experimental Verification and Comparison of Model Predictive, PID and Model Inversion Control in a *Penicillium Chrysogenum* Fed-Batch Process." *Process Biochemistry* 90 (March): 1–11.
- López, Barz, Peñuela, Villegas, Ochoa, & Wozny. 2013. "Model-Based Identifiable Parameter Determination Applied to a Simultaneous Saccharification and Fermentation Process Model for Bio-Ethanol Production." *Biotechnology Progress* 29 (4): 1064–82.
- Maurer, Skerker, Arkin, Miller, Biksacky, & Huether-franken. 2015. "Automated Bioreactor Sampling – Process Trigger Sampling for Enhancing Microbial Strain Characterization."
- Narayanan, Luna, von Stosch, Cruz Bournazou, Polotti, Morbidelli, Butté, & Sokolov. 2020. "Bioprocessing in the Digital Age: The Role of Process Models." *Biotechnology Journal.* Wiley-VCH Verlag.
- Process Systems Enterprise Limited. 2013. "Model Developer Guide." London, UK. <http://www.psenderprise.com>.
- Rocha. 2003. "Model-Based Strategies for Computer-Aided Operation of a Recombinant E . Coli Fermentation." Braga: Escola de Engenharia Universidade do Minho.
- Sonnleitner & Käppeli. 1986. "Growth of *Saccharomyces Cerevisiae* Is Controlled by Its Limited Respiratory Capacity: Formulation and Verification of a Hypothesis." *Biotechnology and Bioengineering* 28 (6): 927–37.

### 2.2.3 Event driven modeling for the accurate identification of metabolic switches in fed-batch culture of *S. cerevisiae*



## Event driven modeling for the accurate identification of metabolic switches in fed-batch culture of *S. cerevisiae*

M. Adnan Jouned<sup>a,b</sup>, Julian Kager<sup>c</sup>, Christoph Herwig<sup>a,c</sup>, Tilman Barz<sup>b,\*</sup>

<sup>a</sup> ICEBE, TU Wien, Gumpendorfer Straße 1a 166/4, 1060 Vienna, Austria

<sup>b</sup> Center for Energy, AIT Austrian Institute of Technology GmbH, Giefinggasse 2, 1210 Vienna, Austria

<sup>c</sup> Competence Center CHASE GmbH, Altenbergerstraße 69, 4040 Linz, Austria

### ARTICLE INFO

#### Keywords:

Event driven modeling  
Piecewise kinetic growth models  
Yeast cultivation  
Metabolic pathways  
Metabolic switches

### ABSTRACT

Mechanistic model-based methods are indispensable tools for characterization, monitoring and control in biopharmaceutical industry. However, the complexity of mechanistic models is restricted by the availability of process analytics. As a result, biological reactions are often lumped and only central metabolic pathways and extracellular analytics are considered. Moreover, due to process dynamics during typical batch and fed-batch cultivations, intracellular phenomena can often not be neglected. Typical examples are the Pasteur effect, Crabtree effect, and diauxic growth. A solution to this is to formulate discontinuous (piecewise) growth models and to incorporate metabolic switches expressed as logical operations. This contribution discusses the application of a piecewise kinetic growth model in the context of an industrial relevant case study. Targeted *Saccharomyces cerevisiae* lab scale experiments were conducted with different glucose and ethanol fluxes to trigger switches between metabolic pathways. We propose to use an event driven method to accurately identify the location and sequence of these switches, and the duration of active metabolic pathways during the time course of an experiment. It turns out that, compared with a standard implementation without active event location, the proposed approach leads to more accurate identification of switches and model parameters and thus, to more accurate model predictions.

### 1. Introduction

**Mechanistic growth models:** During the development of biotechnological processes, mechanistic models play an essential role for effective experimental design [1–6], real-time monitoring and predictive control [7–11]. These models represent the knowledge of the underlying physical characteristics of the process and the physiological behavior of the organisms using mathematical expressions and model parameters [12–14]. Mechanistic models usually show better extrapolation compared to data-driven models [15]. They could predict quantities which are hard or costly to be measured, e.g., soft sensors [16], also, they are increasingly used in the frame of multi-objective control to promote increased selectivity of products [15], making them indispensable tools in biotechnology.

Mechanistic kinetic growth models use stoichiometric information, nonlinear reaction rates and mass and concentration balances [17–19], and are usually written as a set of deterministic and continuous Ordinary Differential Equations (ODEs). Unstructured models do not incorporate a

detailed metabolic and physiological description of the organism. They are mainly used to predict the dynamics of cell density, viability, nutrient/metabolite concentrations, and product titer [20], without a detailed description of cell internal reactions or compartmentalization. Internal reactions are often lumped together and represented as one overall metabolic pathway.

In contrast, structured models like metabolic flux analysis models or extended kinetic models derived from (genome-scale) metabolic networks [21] provide a more detailed mathematical description of the intracellular metabolic regulation and control. However, due to the complexity of the metabolic networks, the difficulty of measuring all metabolite concentrations, and the limited understanding of the reaction sequences and enzymes involved in some areas of metabolism, their application in practice is still either impossible or very costly and demanding [12,21,22].

This is why, in the context of industrial biotechnology, model-based monitoring, control and characterization of microbial cultivations rely mainly on unstructured (or purely data-driven) models [23–25]. The

\* Corresponding author.

E-mail address: [tilman.barz@ait.ac.at](mailto:tilman.barz@ait.ac.at) (T. Barz).

<https://doi.org/10.1016/j.bej.2022.108345>

Received 6 September 2021; Received in revised form 10 January 2022; Accepted 17 January 2022

Available online 20 January 2022

1369-703X/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

complexity of these models depends on the availability of process analytics, such as online gas analyzers, advanced tools such as automatic liquid handling and sampling, and hardware like HPLC (high-performance liquid chromatography), NIR (near infrared spectroscopy) and FIA (flow injection analysis) [26]. A major challenge (addressed in this contribution) is related to the difficulty to parametrize and the reduced predictive capabilities of the models due to simplifications in modeling central metabolic pathways.

Unstructured (or simplified structured) models primarily focus on the description of simplified pathways (e.g., product synthesis, oxidative growth, interconversion and degradation of components), and simplified biological transitions. However, in reality, transitions are mostly continuous, highly nonlinear, and dependent on metabolic regulation, gene expression, and other intracellular mechanisms, but as cell dynamics happen in very different timescales, most of the transitions are simplified into discontinuous behaviors. Therefore, switches expressed as logical operations often need to be incorporated in the model [27]. The result is a (discontinuous) piecewise growth model. Examples of such models are the models that describe the Pasteur effect [28], Crabtree effects [29], and diauxic growth, or the models that consider sudden external changes such as pulse feeding and culture induction.

*S. cerevisiae and pharmaceutical production:* In the context of biopharmaceutical production, *S. cerevisiae*, among other organisms, is a good production platform because of its fast growth rate, low cost of medium and downstream processing, its good secretory capacity [30], and the well-understood metabolism. It's used to produce pharmaceuticals like insulin, blood factors, and vaccines [30], and is recently used also for SARS-CoV-2 vaccine production [31].

Biopharmaceutical upstream production processes are usually split into three phases: (I) batch phase with pre-defined initial substrate and biomass concentrations, (II) fed-batch phase, where the substrate is added to the reactor, (III) and production phase which starts usually by an external inducer. The aim of the first two phases is to maximize growth to obtain a high amount of viable cells that are used for production in the third phase. Aerobic growth is preferred as it assures highest biomass conversion yields and growth rates. The aim of the production phase is to maximize product titers and to provide a constant product quality for the subsequent process steps.

Potentially accumulated ethanol in the medium is known to affect growth rate, as it reduces the mitochondrial membrane integrity and therefore impacts cell metabolism [32]. Therefore, for efficient pharmaceutical bioprocesses, it is important to keep high growth rates without the formation of inhibitory by-products (ethanol in case of *S. cerevisiae*) to ensure high amounts of healthy viable cells for the production of the target product. One possible approach for that is to use the predictive power of growth models to optimize the process conditions.

*Solution methods:* Discontinuous (piecewise) growth models can be mathematically expressed as a combination of a set of continuous differential (and algebraic) equations with discontinuous right-hand side, and a set of time-dependent and/or state-dependent conditions, also referred to as event functions. If a condition is fulfilled, an event is triggered, and the model is switched. A switch can mean a change to another model structure, e.g., switching to a different growth model or to a different metabolic pathway. A switch can also mean an update of the system states, e.g., updating the reactor volume after sampling, or an update of model parameters, e.g., accounting for changes in cell affinity during time. The discrete nature of these phenomena can radically change the future evolution of the overall system behavior [33–35].

According to Dieci and Lopez [36], there are mainly two possible approaches to deal with ODEs with discontinuous right-hand sides: the *time stepping method* and the *event driven method*.

The time stepping method simply ignores discontinuities and uses solvers for continuous initial value problems (IVPs). These solvers assume sufficient smoothness of the right-hand side of the ODEs and rely on the local error estimator and the step size control techniques to keep

errors in the generated approximate solution acceptably low [37]. Although this approach is very simple to be implemented, it can be expected to fail (or at least to become inefficient) in discontinuous regions as the solution there violates the crucial assumption of smoothness [36].

In contrast, the event driven method locates discontinuities (defined as events [38]) using event functions which define discontinuity surfaces in the state space of the differential system. When the solution reaches a surface, an event is located. Thus, the solution is a result of a sequence of IVPs, described by differential equations and interspersed by instantaneous events that cause a discrete change to the initial value problem currently being solved [34], i.e., when the solution reaches an event, the solver updates states, parameters or the model structure and restarts at this point. Applications following this approach have been proposed for discontinuous problems in many fields; in mechanics (e.g., see [39,40]), electrical and control engineering (e.g., see [40–42]), chemical engineering and thermodynamics (e.g., see [33,43,44]), ecology (e.g., see [45]) and neuroscience (e.g., see [27]), but are still limited in the bioprocessing context.

*Available software and algorithms:* State of the art IVP solvers in *MATLAB ODESUIT* [46], or *SUNDIALS* [47] provide the option to monitor and locate time and/or state events using parameter dependent event functions. This is realized by a root finding algorithm where the event is defined by a change of sign in the function [48,49]. The user then specifies what is done when an event is found, i.e., the processing of events [49]. To account for the complexity of different events and switches, Park and Barton [38], proposed a general formulation where classical propositional logic is used for the representation of state conditions as it can represent conjunctions and/or disjunctions of relational expressions effectively. This general formulation of models can be used in different modeling languages and software systems. A review on their application for the analysis of general differential and algebraic hybrid (continuous/ discrete) systems is given by Barton et al., [35], where the authors also discuss the consistent reinitialization after detection of state events, the parametric sensitivity analysis, and open problems related to systems with changing sequence of modes. More information and comparison of simulation tools for the analysis of hybrid systems, such as *GPROMS*, *Modelica/Dymola*, *Assimulo*, *deSolve*, *Mathematica* can be found for example in [39,50–54]. Fröhlich et al. [27] presented a recent review in the context of computational biology on available software toolboxes for the solution of differential systems with time or state (and possible parameter) dependent event functions. It turns out that most toolboxes consider only time-dependent events which can be triggered by external changes, such as changes to the reactor feed and are not useful for the consideration of state events that are triggered when certain critical cell internal conditions are reached. In addition, the authors also identify a lack of functionalities for sensitivity analysis with respect to parameters in the model and/or event functions and propose an extension of the maximum likelihood fitting criterion in order to account for model predictions with missing events.

This contribution considers the recent work by Fröhlich et al. [27] on the identification of dynamical biological systems with discrete events and logical operations, where the authors present the development of a mathematical framework and provide an analysis on the accuracy of the numerical simulation and the benefit of accurate sensitivities for parameter estimation. While Fröhlich et al. present applications with rather simple examples (small linear or quadratic ODE's) in neuroscience and mRNA transcription, this contribution presents results from parameter estimation for a more complex nonlinear example for *Saccharomyces cerevisiae* yeast growth [55].

The model in this contribution consists of three different metabolic submodels and is fitted to data from targeted lab scale experiments where ethanol and glucose fluxes are indirectly controlled to trigger switches between different metabolic pathways. The paper highlights the consequences when not explicitly accounting for discontinuities during model implementation and its numerical solution. It turns out

that the solution might be heavily corrupted by noise which can affect not only the accuracy of simulation results but also the convergence of the model fitting algorithm. This has a negative impact on the identifiability of the estimated parameters and increases the model prediction uncertainty.

The novelty of this contribution lies in the systematic thorough quantitative analysis of the performance of the event driven method (EDM) in the context of a biotechnological process and the detailed presentation of a proposed state-of-the-art method for a sound implementation of a typical and well-accepted mechanistic growth model with discontinuous and continuous behaviors for a realistic industrial-relevant use case.

**Structure of the paper:** Differential equation systems with discontinuities and corresponding solution methods are summarized in Section 2, this section also contains the model for yeast growth taken from [55], the model calibration procedure as well as details on the conducted experiments and the reference analytics.

Section 3 presents the results. In Section 3.1 details are given on the numerical implementation following the proposed event driven method (EDM) and the time stepping method (TSM) to account for model discontinuities. In Section 3.2 the EDM is used for the model parametrization of experimental data, considering the quality of the fits and the parameter identifiability.

Section 3.3 presents the comparative analysis of the results obtained by TSM and EDM including the model prediction accuracy, parameters identifiability, convergence of the numerical algorithm for model fitting, and discusses practical aspects in a bioprocessing context. Finally, Section 4 provides the discussion and conclusions.

## 2. Methods

### 2.1. Differential equation systems with discontinuities

#### 2.1.1. Time stepping method

Time stepping methods are widely used for the solution of continuous ordinary differential equation systems (ODEs). The simulation requires the solution of an initial value problem (IVP) described as:

$$\dot{x}(t) = f(x(t), u(t), \theta) \quad \text{with } t \in [t_0, t_f] \quad (2.1)$$

where  $t \in \mathbb{R}$  is the independent time variable,  $x(t) \in \mathbb{R}^{N_x}$  is the vector of dependent state variables,  $u(t) \in \mathbb{R}^{N_u}$  is the time-varying input vector, and  $\theta \in \mathbb{R}^{N_\theta}$  is the parameter vector. Initial conditions are given as  $x(t_0) = x_0$ .

Using time stepping methods for the solution of ODEs with discontinuous right-hand sides means to ignore the discontinuities. Time stepping methods rely on the local error estimator of the solver to ensure that the integration errors remain acceptably small. Therefore, in regions where discontinuities of the solution or its derivative occur, the time stepping method may become either inaccurate or inefficient, or both. The reason is that the local error analysis of the step size control fails because there is not sufficient smoothness of the right-hand side of the ODE [36]. It is noted that modifications of these methods have been proposed to account for ODEs with discontinuities, see e.g., [36]. However, in this contribution, we refer to the “standard” time stepping method that uses the (standard) routines of the *MATLAB ODE SUITE* package for the solution of continuous ODEs without option for event handling.

#### 2.1.2. Event driven method

Event driven methods use event functions, which define the occurrence of discontinuities in the state or time space of the differential system [36]. During numerical solution of the ODE, the exact locations of events are located by solving the conditions equations system outside the ODE system, and the numerical integration is restarted at this point. Barton and Pantelides [33] define the mathematical formulation for this

simulation problem as a sequence of IVP's interspersed by the occurrence of discontinuities (known as events):

$$\dot{x}^{(k)}(t) = f^{(k)}(x^{(k)}(t), u^{(k)}(t), \theta^{(k)}) \quad \text{with } t \in [t^{(k-1)}, t^{(k)}] \quad \forall k = 1, \dots, N_{CD} \quad (2.2)$$

In eq. (2.2) the time domain of interest  $[t_0, t_f]$  is partitioned into  $N_{CD}$  continuous subdomains  $[t^{(k-1)}, t^{(k)}]$ .

While the initial time  $t_0$  is given, the end of each sub-interval is determined by the occurrence of an event. Events are detected during the course of a simulation. The superscript  $k$  indicates that the set of variables and the set of equations may vary from subdomain to subdomain in a completely general manner.

The model equations and initial conditions of the first subdomain are determined by an individual simulation description. For the succeeding subdomains they will be determined from a combination of the final state of the system in the preceding subdomain and the consequences of the corresponding event(s) [33].

**Time and state events:** Discontinuities in ODE models can either be defined by “implicit (or state) events” or by “explicit (or time) events”. In the first, the time of occurrence is not known in advance because it is dependent on the system fulfilling certain conditions. Therefore, the numerical solution of the equations must be advanced speculatively until the state condition becomes satisfied. In contrast, for explicit events the exact time of occurrence is known in advance. Thus, the solution can proceed to these events in time order [33]. Both, implicit and explicit events can trigger (implicit or explicit) switches in the model structure, state variables or parameter values. These switches are triggered by predefined conditional statements (or simply “conditions”) which for explicit switches are defined by exact time points and for implicit switches by a suitable threshold defined by the state variables and parameters.

**Conditions:** Conditions or trigger functions define the time point of occurrence of an event. The general form of these conditions which can trigger both, time and state events, can be defined as:

$$c(x(t), u(t), \theta, t) = 0 \quad (2.3)$$

In the standard mathematical description in eq. (2.3) the “critical threshold” of the condition is zero. However, as the condition  $c(\cdot)$  is a general relation of states, controls and parameters, the critical threshold might also be represented by any other value including nonlinear relations, e.g., for limiting concentrations or uptake rates. In contrast, conditions for explicit events (such as sudden changes in the reactor volume due to sampling) can be simply written as  $t - t_s = 0$ . Details on the numerical implementation of the event-driven method can be found in Appendix A.

### 2.2. Model fitting, identifiability analysis and uncertainty quantification

The model is fitted to the experimental data by nonlinear regression considering the normalized residual sum of squares (NRSS) of the measured and predicted liquid and gas concentrations. For  $N_L$  liquid concentrations which were measured in  $M_L$  samples taken from the reactor at different time points, and for  $N_G$  gas concentrations which were continuously monitored and evaluated at  $M_G$  time points, and with  $\theta$  as unknown parameter vector, the unconstrained and unbounded minimization problem reads:

$$\begin{aligned} \min_{\theta} \phi^{NRSS}(\theta) \quad \text{with} \\ \phi^{NRSS}(\theta) = \frac{1}{M_L} \sum_{i=1}^{N_L} \sum_{j=1}^{M_L} (Y_{ij}(\theta) - Y_{ij}^m)^2 \\ + \frac{1}{M_G} \sum_{k=1}^{N_G} \sum_{l=1}^{M_G} (Y_{kl}(\theta) - Y_{kl}^m)^2 \end{aligned} \quad (2.4)$$

Parameter initial guesses and estimates are given in Table 4. CO<sub>2</sub> and O<sub>2</sub> content in the off-gas was obtained from online gas analyzer and

**Table 1**  
List of symbols.

Term	Description	Unit
$\alpha_1$	Consumed O <sub>2</sub> for oxidative growth on glucose	mol/mol
$\alpha_{10}$	Produced biomass for oxidative growth on ethanol	mol/mol
$\alpha_{11}$	CO <sub>2</sub> yield for oxidative growth on ethanol	mol/mol
$\alpha_{12}$	H <sub>2</sub> O yield for oxidative growth on ethanol	mol/mol
$\alpha_2$	Produced biomass for oxidative growth on glucose	mol/mol
$\alpha_3$	CO <sub>2</sub> yield for oxidative growth on glucose	mol/mol
$\alpha_4$	H <sub>2</sub> O yield for oxidative growth on glucose	mol/mol
$\alpha_5$	Produced biomass for fermentative growth on glucose	mol/mol
$\alpha_6$	CO <sub>2</sub> yield for fermentative growth on glucose	mol/mol
$\alpha_7$	H <sub>2</sub> O yield for fermentative growth on glucose	mol/mol
$\alpha_8$	Ethanol yield for fermentative growth on glucose	mol/mol
$\alpha_9$	Consumed O <sub>2</sub> for oxidative growth on ethanol	mol/mol
$C_e$	Ethanol concentration	g/L
$C_{s,in}$	Glucose concentration in the feed	g/L
$C_s$	Glucose concentration	g/L
$C_x$	Biomass concentration	g/L
$CER$	Carbon dioxide evolution rate	mol/h
$F_{Acid}$	Acid feed rate	L.h <sup>-1</sup>
$F_{Base}$	Base feed rate	L.h <sup>-1</sup>
$F_{gas}$	Gas in/out flow	L.h <sup>-1</sup>
$F_S$	Glucose feed rate	L.h <sup>-1</sup>
$HX$	Mass fraction of hydrogen in biomass	mol H/mol C
$K_e$	Time affinity constant of the ethanol	g/L
$K_i$	Inhibition parameter of ethanol consumption because of glucose	g/L
$K_s$	Time affinity constant of the glucose	g/L
$\mu_{total}$	Total growth rate	h <sup>-1</sup>
$Mw_e$	Molecular weight of the ethanol	g/mol
$Mw_s$	Molecular weight of the glucose	g/mol
$Mw_x$	Molecular weight of the biomass	g/mol
$NX$	Mass fraction of nitrogen in biomass	mol N/mol C
$OUR$	Oxygen uptake rate	mol/h
$OX$	Mass fraction of oxygen in biomass	mol O/mol C
$q_{O_2}$	Specific oxygen uptake rate per unit of biomass	mmol.h <sup>-1</sup> .g <sup>-1</sup>
$q_s$	Specific glucose uptake rate per unit of biomass	g.h <sup>-1</sup> .g <sup>-1</sup>
$q_e$	Specific ethanol uptake rate per unit of biomass	g.h <sup>-1</sup> .g <sup>-1</sup>
$V$	Liquid volume	L
$Y_{o_2/e}$	Oxygen (stoichiometric) yield on ethanol	mmol/g
$Y_{o_2/s}$	Oxygen (stoichiometric) yield on glucose	mmol/g
$Y_{e/s}$	Ethanol (stoichiometric) yield from glucose fermentation	g/g
$Y_{x/(ox)}$	Biomass yield for oxidative growth on ethanol	g/g
$Y_{x/(ox)}$	Biomass yield for oxidative growth on glucose	g/g
$Y_{x/(red)}$	Biomass yield for reductive growth on glucose	g/g

mass flow measurements. The carbon evolution rate (CER) and the oxygen uptake rate (OUR) were calculated from these measurements and considered in eq. (2.4). Glucose, ethanol and biomass liquid concentrations were measured offline by sampling, see section 2.4 for details.

Fitting of the growth model and parameter identifiability analysis are carried out following the scheme of [56]. The scheme requires estimated parameters  $\hat{\theta}$  and the corresponding sensitivity matrix  $S(\hat{\theta})$ .  $\hat{\theta}$  are obtained from repeated numerical solutions of a nonlinear regression, where the initial parameters of each run are defined by stochastic sampling in a reasonable parameter space around values taken from literature [55], [57]. The sensitivity matrix  $\tilde{S}$  is obtained by normalizing  $S$  with the initial parameters and model output. Singular value decomposition (SVD) is used to detect any linear dependencies in  $\tilde{S}$ . By decomposing  $\tilde{S} = U\Sigma V$ ,  $\Sigma$  matrix is found which holds the singular values of  $\tilde{S}$ . The singular values in  $\Sigma$  are then used to calculate: 1- the condition number ( $\kappa$ ) which is a measure of the sensitivity of model

results to the perturbation of the parameters. 2- the collinearity index ( $\gamma$ ) which quantifies the collinearity of the parameters. Empirical values for thresholds of  $\kappa$  and  $\gamma$  are chosen based on [58].

The parameters are ranked according to their linear independence and the above metrics are used to perform a parameter subset selection (SsS). The identifiable parameter subset simultaneously satisfies both sensitivity and linear independence conditions. Based on the results some parameters are set to active while the others are deactivated and not considered for nonlinear regression. The solution of the nonlinear regression problem and the SsS are computed repeatedly until convergence to the best overall parameter values.

**Parameter's uncertainty quantification:** The uncertainty of the parameters is analyzed using bootstrapping technique [59], which involves resampling of the experimental data and re-estimation of the parameters. For each experiment 500 *Monte Carlo* (MC) datasets are generated based on the measurement's uncertainty. The perturbation is chosen to be three standard deviations of the nominal values of the error for each measurement device for both liquid concentrations and off-gas signals. The measurement error (normal non-correlated error, given as three standard deviations) for biomass is 5%, for ethanol 4%, glucose 1, 3% and for the off-gas is 3.75%. The model is fitted, and parameters are estimated for all 500 datasets individually. The probability distribution of the resulting 500 parameter estimates is assumed to be normal. The 95% parameter confidence regions are used to quantify the accuracy of the estimates.

**Convergence analysis:** Bootstrapping is also used to analyze the convergence of the parameter estimation algorithm for the solution of eq. (2.4). For each experiment a set of 500 *Monte Carlo* (MC) datasets is generated based on the measurement's uncertainty (same as above). In addition, the initial guesses of the parameters were perturbed. The perturbations are chosen by uniform sampling in a  $\pm 10\%$ ,  $\pm 30\%$  and  $\pm 50\%$  interval around the parameters' nominal values which were defined by the best estimates. The model is fitted 500 times and the 95% parameter confidence regions are calculated. The confidence regions are used as a measure for the robustness of the convergence of the fitting algorithm.

**Prediction uncertainty:** The distribution of the simulated output is calculated by a sampling considering the 500 parameter estimates from the convergence analysis. Results in Section 3.3.1 are given for  $\pm 30\%$  perturbation of the initial parameter guess. The depicted ranges of the prediction uncertainty correspond to  $\pm 2\sigma$  (95% confidence interval).

The calculated parameters and prediction uncertainties as well as the results of the convergence analysis are affected not only by perturbations in the measured data sets but also by possible errors in the approximate numerical solution of the model. As mentioned before, using TSM the accuracy of the numerical solution might be low as the errors in the event location are not controlled. As a consequence, the solution of the model can be corrupted by significant "numerical noise", and the parameter estimation problem is characterized by so-called "noisy functions" [60]. Applying bootstrapping and repeatedly solving the parameter estimation problem the results depend on two factors: the perturbations in the measurements, and the numerical noise in the model prediction. Using bootstrapping technique, the impact of both factors is analyzed for the TSM and EDM.

### 2.3. Hardware and software

All computations were carried out in *MATLAB R2017b* on an *Intel (R) Xeon(R)* (CPU E5-2690 V4@ 2.60 GHZ) with 64 GB RAM using 64x-bit operating system. Parallel processing is used. The ODEs (initial value problems) have been solved using *MATLAB "ODE suite"*, mainly by *ODE15S* solver. On average, the computation time for one simulation using the solver's default settings for the absolute and relative error tolerances, takes roughly 1.6 [s] for EDM, and 1.4 [s] for TSM. A detailed comparison of the computation times is given in Appendix E. The unconstrained nonlinear regression problem eq. (2.4) was solved

**Table 2**

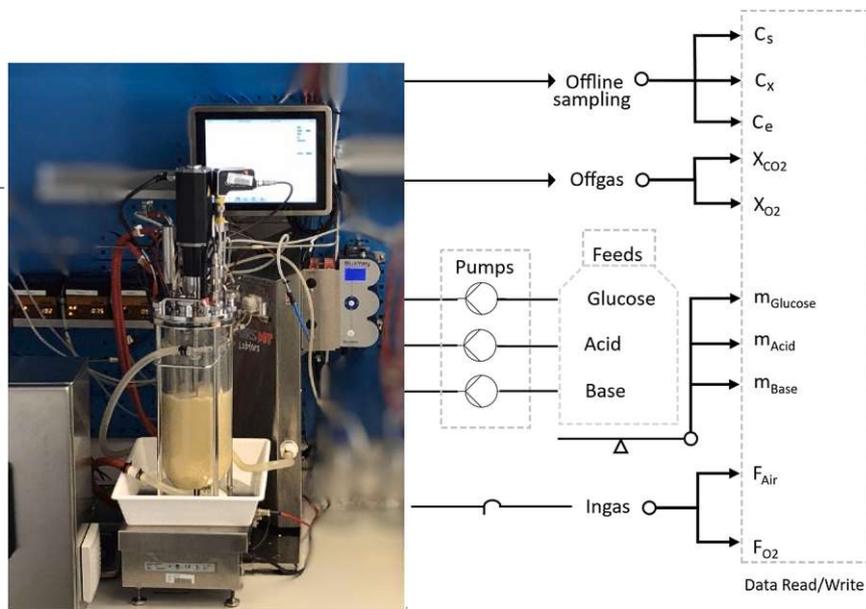
Fed-batch experimental conditions given as initial biomass and glucose concentrations  $C_{X0}$  and  $C_{S0}$ , glucose feed concentration  $C_{s,in}$ , feed start time  $t_{0,feed}$ , duration of overfeeding  $t_{overfeed}$ , and total volume of all samples taken.

No.	$C_{X0}$ [g/L]	$C_{S0}$ [g/L]	$C_{s,in}$ [g/L]	$t_{0,feed}$ [h]	$t_{overfeed}$ [h]	Total sampling volume [L]
Experiment 1	0.7	18.9	220	15	1	0.25
Experiment 2	0.5	19.9	200	16.3	1.8	0.3
Experiment 3	0.5	18.7	198	15.7	2.1	0.39

using MATLAB's "Optimization Toolbox" lsqnonlin/trust-region-reflective algorithm.

#### 2.4. Fed-batch experiments

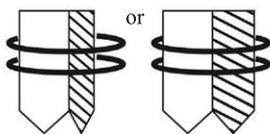
Three *Saccharomyces cerevisiae* (wildtype, CBS8340) experiments



**Fig. 1.** Bioreactor image, and scheme indicating all collected data to simulate and to parametrize the model. Time-dependent inputs are determined as glucose, acid and base feed rates and are calculated from the respective balance signal ( $m_{Glucose}$ ,  $m_{Acid}$ ,  $m_{Base}$ ). Online CER and OUR is calculated based on gas composition ( $X_{CO_2}$ ,  $X_{O_2}$ ) in the off-gas stream as well as input gas stream ( $F_{Air}$ ) and oxygen content ( $F_{O_2}$ ). Biomass, ethanol and glucose concentrations ( $C_S$ ,  $C_X$ ,  $C_E$ ) are measured from offline samples.

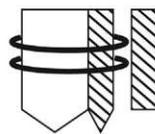
#### Submodel A

**Glucose and ethanol oxidation**  
below glucose and ethanol  
oxidative maximum uptake rates



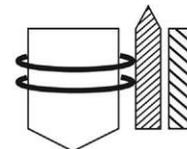
#### Submodel B

**Glucose and ethanol oxidation**  
below glucose but above ethanol  
oxidative maximum uptake rate



#### Submodel C

**Crabtree effect**  
above maximum glucose oxidative  
uptake rate, ethanol production



□ Metabolized glucose    ◻ Metabolized ethanol    ▨ Un-metabolized ethanol    ▩ Produced ethanol

**Fig. 2.** The "bottleneck" concept in Sonnleitner and Käppli [51] yeast fermentation model. If the sum of substrate fluxes does not exceed, or equals, the cell's maximum oxidative uptake rate (shown as two rings in the figure), then the flux is subcritical, or critical, (left). If the glucose flux is higher than the maximum oxidative uptake rate, then the flux is supercritical. The residual part of glucose is metabolized reductively to produce ethanol (right). If the sum of substrate fluxes exceeds the maximum oxidative uptake rate then the flux is supercritical, but ethanol uptake is limited to the maximum oxidative capacity (middle).

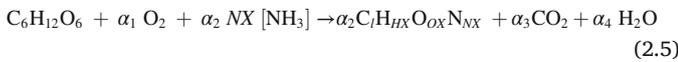
The figure is adapted from [55].

supernatant were analyzed by HPLC (*Thermo Fischer, USA*) with a Supelco gel C-610 H ion exchange column (*Sigma-Aldrich, USA*) and a refractive index detector (*Thermo Fischer, USA*). The mobile phase was 0.1%  $H_3PO_4$  with a constant flow rate of 0.5 mL/min at a temperature of 4 °C. Biomass concentration was determined gravimetrically by separating the cells from 5 mL culture broth via centrifugation at 4800 rpm for 10 min at 4 °C. The cell pellet was dried at 105 °C after a washing step with 5 mL of water in weighted glass tubes and the weight of the dried pellet was determined on an analytical balance. Fig. 1 shows a scheme of the used setting.

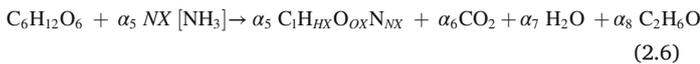
## 2.5. Growth model of *S. cerevisiae*

In Sonnleitner and Käppeli [55] yeast fermentation model, the authors describe how *Saccharomyces cerevisiae* grows using different metabolic pathways. Three reactions (metabolic routes) are distinguished by the following equations with the stoichiometric coefficients ( $\alpha_1$ - $\alpha_{12}$ ):

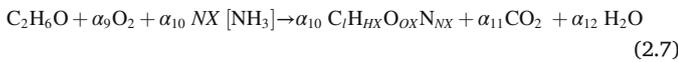
-Oxidative conversion of glucose (into biomass and CO<sub>2</sub>)



-Reductive conversion of glucose (into biomass, CO<sub>2</sub> and ethanol)



-Oxidative conversion of ethanol (into biomass and CO<sub>2</sub>)



The formula  $C_1H_{HX}O_{OX}N_{NX}$  denotes the biomass, where the molecular composition  $HX$ ,  $OX$ ,  $NX$  can be determined by elemental analysis of the dried biomass. Note that it is assumed that the elemental compositions of ethanol-grown biomass and glucose-grown biomass are the same as the difference is within the analytical errors [62]. The yield coefficients  $Y_{x/s(ox)}$ ,  $Y_{x/s(red)}$  and  $Y_{x/e(ox)}$  are determined as model parameters by fitting the model predictions to the measurements. The stoichiometric coefficients  $\alpha_2$ ,  $\alpha_5$ , and  $\alpha_{10}$ , can be determined retrospectively, assuming that the molecular weights ( $Mw$ ) of biomass, glucose and ethanol are known using the relations:  $Y_{x/s(ox)} = \alpha_2 Mw_x / Mw_s$ ,  $Y_{x/s(red)} = \alpha_5 Mw_x / Mw_s$ ,  $Y_{x/e(ox)} = \alpha_{10} Mw_x / Mw_e$ . All remaining stoichiometric coefficients in eqs. (2.5), (2.6) and (2.7) and the corresponding (stoichiometric) yield coefficients that are used later ( $Y_{O_2/s}$ ,  $Y_{O_2/e}$  and  $Y_{e/s}$ ) can be determined by considering the elemental balance of carbon, oxygen, and hydrogen and solving a linear system as described in [55] (see Appendix B). The values of  $\alpha_i$  are later used in Eqs. (2.19) and (2.20).

The specific uptake rates of glucose (substrate)  $q_s$ , ethanol  $q_e$  and oxygen  $q_{O_2}$  are assumed to follow Monod kinetics:

$$q_s = q_s^{max} \frac{C_s}{C_s + K_s} \quad (2.8)$$

$$q_e = q_e^{max} \frac{C_e}{C_e + K_e} \frac{K_i}{K_i + C_s} \quad (2.9)$$

$$q_{O_2} = q_{O_2}^{max} \frac{C_o}{C_o + K_o} \quad (2.10)$$

where  $q_s^{max}$ ,  $q_e^{max}$  and  $q_{O_2}^{max}$  represent the maximum rates with the respective half saturation rates  $K_s$ ,  $K_e$  and  $K_o$  in dependence of the respective concentrations  $C_s$ ,  $C_e$  and  $C_o$ . In addition, ethanol uptake is inhibited by glucose concentration  $C_s$  via competitive inhibition with  $K_i$  as an inhibition constant.

The main concept to switch between these regimes is the “bottleneck” of the respiratory capacity of the cells. The maximum glucose oxidation capacity  $q_{s(ox)}^{max}$  is determined by the current oxygen uptake and the

stoichiometric conversion yield  $Y_{O_2/s}$ :

$$q_{s(ox)}^{max} = \frac{q_{O_2}}{Y_{O_2/s}} \quad (2.11)$$

Based on  $q_{s(ox)}^{max}$  it is possible to differentiate between *subcritical*/*supracritical* substrate flux. The bottleneck to select between a rate limitation by oxygen and substrate availability can be written as:

$$q_s \leq q_{s(ox)}^{max} \quad (2.12)$$

If (2.12) is fulfilled, then the substrate flux is subcritical and can be entirely converted by the oxidative pathways,  $q_{s(ox)} = q_s$  is given by Eq. (2.8), i.e., the actual rate  $q_{s(ox)}$  is equal to the specific rate. Moreover, whether or not ethanol is present in the medium, no reductive reaction happens and therefore neither ethanol production nor fermentative growth exists, i.e.,  $q_{s(red)} = 0$ , according to eq. (2.6). Similar to glucose, maximum ethanol oxidation capacity can be computed as:

$$q_{e(ox)}^{max} = \frac{q_{O_2} - Y_{O_2/s} \cdot q_{s(ox)}}{Y_{O_2/e}} \quad (2.13)$$

where  $q_{s(ox)}$  is the oxidatively consumed glucose.  $Y_{O_2/s}$  and  $Y_{O_2/e}$  are the respective O<sub>2</sub> stoichiometric conversion yields for glucose and ethanol. Similar to glucose oxidation two cases can be distinguished for the ethanol consumption:

$$q_e \leq q_{e(ox)}^{max} \quad (2.14)$$

If (2.14) is fulfilled, then ethanol can be oxidized at current maximum rate  $q_{e(ox)} = q_e$  as described in equation (2.9). Otherwise, the potential ethanol flux exceeds the oxidative capacity, and ethanol uptake rate  $q_{e(ox)}$  is limited to maximum oxidative capacity  $q_{e(ox)} = q_{e(ox)}^{max}$ . On the other hand, if:

$$q_s > q_{s(ox)}^{max} \quad (2.15)$$

then glucose substrate flux is higher than the oxidative capacity, and growth based on glucose corresponds to maximum possible oxidative capacity  $q_{s(ox)} = q_{s(ox)}^{max}$ , the remaining sugar uptake is reduced to ethanol and can be determined by:

$$q_{s(red)} = q_s - q_{s(ox)}^{max} \quad (2.16)$$

Once the cells are in the reductive pathway, ethanol cannot be used as a substrate for growth anymore  $q_{e(ox)} = 0$ .

Overall, the model considers an “oxidative” growth by a co-metabolized glucose and ethanol under the conditions of *subcritical* substrate flux (reaction routes in eqs. (2.5) and (2.7)), “oxidoreductive” growth under aerobic conditions of *critical* and *supracritical* glucose flux (eqs. (2.5) and (2.6)), and “reductive” growth under anaerobic conditions only (eq. (2.6)).

Based on the selected regimes the overall growth can be expressed by the usage of the respective conversion yields of the single pathways:

$$\mu_{total} = Y_{x/s(ox)} \cdot q_{s(ox)} + Y_{x/s(red)} \cdot q_{s(red)} + Y_{x/e(ox)} \cdot q_{e(ox)} \quad (2.17)$$

The mass balances for a fed-batch reactor with a glucose feed, assuming an open system, isothermal operation and homogenous medium result in the following system of ODE's:

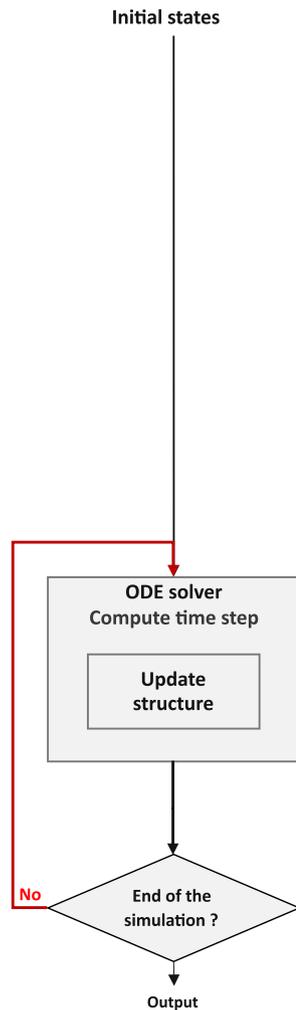
$$\frac{dC_x}{dt} = \mu_{total} \cdot C_x - \frac{F_s}{V} \cdot C_x$$

$$\frac{dC_s}{dt} = - (q_{s(red)} + q_{s(ox)}) \cdot C_x - \frac{F_s}{V} \cdot C_s + \frac{F_s}{V} \cdot C_{s,in}$$

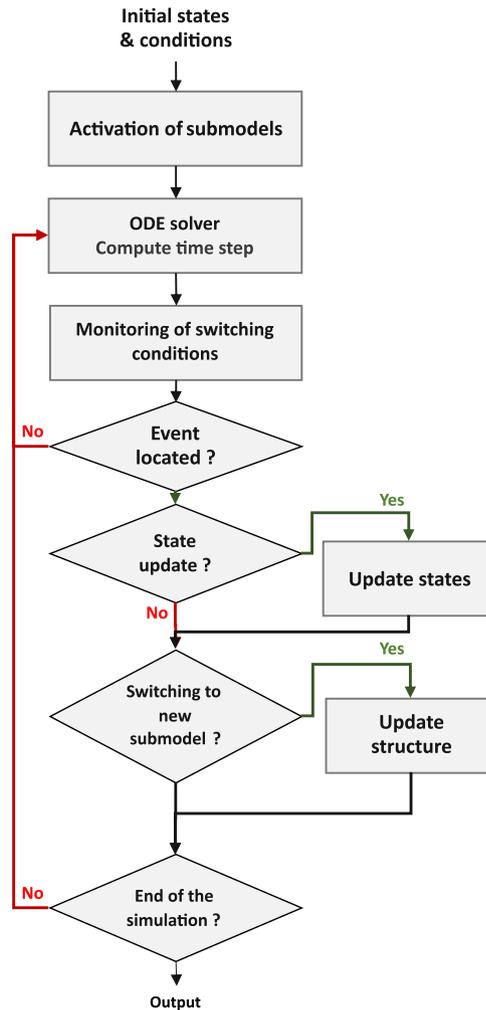
$$\frac{dC_e}{dt} = (Y_{e/s} \cdot q_{s(red)} - q_{e(ox)}) \cdot C_x - \frac{F_s}{V} \cdot C_e$$

$$\frac{dV}{dt} = F_s + F_{Base} + F_{Acid} - F_{sampling} \quad (2.18)$$

## Standard TSM



## Proposed EDM



**Fig. 3.** Computational schemes of the standard time stepping method (TSM) and the proposed event driven method (EDM). The EDM scheme accounts explicitly for any discontinuities in the model by monitoring conditions and switching to the corresponding submodels (or updating state values).

Based on the derived stoichiometric coefficients in eqs. (2.5), (2.6) and (2.7), produced carbon dioxide (CER) and oxygen uptake (OUR) can be added as additional model outputs:

$$OUR = \left( q_{s(ox)} \cdot \alpha_1 \frac{M_{W_{O_2}}}{M_{W_s}} + q_{e(ox)} \cdot \alpha_9 \frac{M_{W_{O_2}}}{M_{W_e}} \right) C_x \cdot V \quad (2.19)$$

$$CER = \left( q_{s(ox)} \cdot \alpha_3 \frac{M_{W_{CO_2}}}{M_{W_s}} + q_{e(ox)} \cdot \alpha_{11} \frac{M_{W_{CO_2}}}{M_{W_e}} + q_{s(red)} \cdot \alpha_6 \frac{M_{W_{CO_2}}}{M_{W_s}} \right) C_x \cdot V \quad (2.20)$$

The Respiratory Quotient (RQ), which is considered a valid indicator for different metabolic pathways [59], is defined as:

$$RQ = \frac{CER}{OUR} \quad (2.21)$$

A RQ greater than one indicates that the *S. cerevisiae* is producing ethanol by Crabtree effect (oxidoreductive growth). A RQ close to one indicates that glucose is mostly oxidatively consumed. RQ values around 2/3 indicate ethanol oxidative consumption. Summing up, the ODE model includes biomass, substrate and ethanol concentrations, and volume as states  $x(t)$ . Glucose, acid and base addition are considered as time-dependent inputs  $u(t)$ . Initial estimates of the model parameters ( $\theta$ )

are taken from [55,57] and given in Table 4. Table 1 shows all symbols and meanings.

### 3. Results

#### 3.1. Growth model implementations

Section 3.1 discusses the implementation of the kinetic growth model of *S. cerevisiae* (Section 2.5) which describes growth on three different pathways. Two methods for the implementation of the ODE system are used (see Section 2.1 for details):

(I) time stepping method (TSM): uses the routines of the *MATLAB* *ODESUITE* package without event detection option.

(II) event driven method (EDM): uses the routines of the *MATLAB* *ODESUITE* package with an event detection option. This means that, during the solution of the ODEs, conditions are monitored, and corresponding events are detected using *MATLAB*'s ODE event location algorithm. If an event is detected, the integration is terminated, and the model is switched. The integration is then restarted with the new submodel and/or adapted initial conditions.

Fig. 3 shows in detail the computational schemes of both methods.

3.1.1. Model implementation following the time stepping method

For TSM, the implementation of different submodels is straightforward using the standard programming language expressions. For the conditions equations (2.12) and (2.14), extrema functions {MIN, MAX}, or alternatively conditional statements {if, switch, while, match...} [63–65] are used. The generic syntax for the equations (2.12) and (2.14) are written for example as:

$$q_{s(ox)} = \text{MIN}(q_s, q_{s(ox)}^{\max}) \quad | \quad q_{e(ox)} = \text{MIN}(q_e, q_{e(ox)}^{\max}) \quad (3.1)$$

or:

$$\begin{array}{l|l} \text{If } (q_s \leq q_{s(ox)}^{\max}) & \text{If } (q_e \leq q_{e(ox)}^{\max}) \\ q_{s(ox)} = q_s & q_{e(ox)} = q_e \\ \text{else} & \text{else} \\ q_{s(ox)} = q_{s(ox)}^{\max} & q_{e(ox)} = q_{e(ox)}^{\max} \\ \text{end} & \text{end} \end{array} \quad (3.2)$$

It is noted that for eqs. (3.1) and (3.2), the only difference is the syntax; the execution of both statements gives the exact same results.

Sampling volumes are incorporated in the model simply by considering “sampling” flow rate  $F_{\text{sampling}}$  (i.e., flow negative pulse signal) in eq. (2.18). As the changes in the overall volume over time is calculated as the difference between input and output flow rates [12], it is widely accepted to consider sampling volume  $F_{\text{sampling}}$  as a part of the flows that are leaving the reactor [65–67].

In the TSM method, the conditional statements are computed at each evaluation time point of the ODE solver. While the time steps are adaptively chosen based on the integration error estimate, there is no active control on the location of events. The solver is therefore likely to miss the exact time point when a switch between metabolic pathways occurs, or the samples are taken. In addition, in the TSM implementation positivity of the ODE solution is not enforced.

3.1.2. Model implementation following the event driven method

The kinetic growth model by Sonnleitner and Käppli [55] describes how *Saccharomyces cerevisiae* grows based on different metabolic pathways, i.e., regimes with different substrate uptake. Switches between these pathways are triggered by events. They are actively located by monitoring of the conditions (2.12) and (2.14), whose threshold is given by equations (2.11) and (2.13), respectively.

Adopting the event driven method (given in Section 2.1.2) requires the model to be separated into: a) conditions and b) submodels.

The growth model including the three metabolic pathways (as three submodels) can be represented by the general mass balances equations (2.18) written in matrix form as:

where the conversion matrix containing the yield coefficients is multiplied by the reaction vector containing the current reaction rates  $q$  ( $t$ ).  $\mathcal{A} \in \mathbb{R}^{N_q \times N_q}$  is the activation matrix, whose elements {1,0} are used to activate/deactivate submodels. Table 3 shows all possible submodels, the corresponding diagonal elements of  $\mathcal{A}$ , and the selection criteria.

$$\frac{d}{dt} \begin{bmatrix} C_x \\ C_s \\ C_e \end{bmatrix} = \begin{bmatrix} Y_{x/s(ox)} & Y_{x/s(ox)} & Y_{x/s(red)} & Y_{x/e(ox)} & Y_{x/e(ox)} \\ -1 & -1 & -1 & 0 & 0 \\ 0 & 0 & Y_{e/s} & -1 & -1 \end{bmatrix} \cdot \mathcal{A} \cdot \begin{bmatrix} q_s \\ q_{s(ox)}^{\max} \\ q_s^{red} \\ q_e \\ q_{e(ox)}^{\max} \end{bmatrix} C_x - D \begin{bmatrix} C_x \\ C_s \\ C_e \end{bmatrix} + \frac{F_s}{V} \begin{bmatrix} 0 \\ C_{in} \\ 0 \end{bmatrix} \quad (3.3)$$

The selection process is illustrated as a decision tree in Appendix C. The Boolean trigger function  $\mathcal{E}_1(t)$  considering condition (2.12) reads:

$$\mathcal{E}_1(t) := \begin{cases} 1, & \text{if } q_s(x(t), u(t), \theta, t) - q_{s(ox)}^{\max}(x(t), u(t), \theta, t) \geq 0 \\ 0, & \text{if } q_s(x(t), u(t), \theta, t) - q_{s(ox)}^{\max}(x(t), u(t), \theta, t) < 0 \end{cases} \quad (3.4)$$

The Boolean trigger function  $\mathcal{E}_2(t)$  considering condition (2.14) reads:

$$\mathcal{E}_2(t) := \begin{cases} 1, & \text{if } q_e(x(t), u(t), \theta, t) - q_{e(ox)}^{\max}(x(t), u(t), \theta, t) \geq 0 \\ 0, & \text{if } q_e(x(t), u(t), \theta, t) - q_{e(ox)}^{\max}(x(t), u(t), \theta, t) < 0 \end{cases} \quad (3.5)$$

3.1.2.1. Accounting for additional discontinuities and non-physical solutions in the model. Non-physical solutions and highly nonlinear kinetic terms: Nonlinear kinetic models such as Monod-type growth models can exhibit stiff behaviors, especially when the affinity of the microorganisms to the used substrates is high [12], and the affinity constant ( $K$ ) is small (roughly  $< 0.2$ ). While these models predict a maximum growth rate for most substrate concentrations, rate limitation occurs in a very limited substrate range close to 0 and the reaction is stopped at zero concentration. Accordingly, the growth rate curve exhibits a very steep slope for low substrate concentrations whereas it is almost constant elsewhere. For changing substrate concentrations, from low to high, or vice-versa, the steepness of the response appears to change suddenly, the model shows an “almost-discontinuous” behavior. In addition, while Monod is defined for positive substrate concentrations, for negative concentrations it gives non-physical solutions, i.e., positive rates below  $-K$  and negative rates between  $-K$  and 0. Because of this, an often-encountered issue in the numerical solution of ODEs with Monod-type models are the negative substrate concentrations. The consequences range from inaccuracies in the computed model predictions, to instability of the ODE model which might lead to simulation failure. These non-physical solutions can be avoided by following the event driven implementation [68]. It is noted that some ODE solvers, such as in the *SUNDIALS ODE suite* [47], directly provide an option for

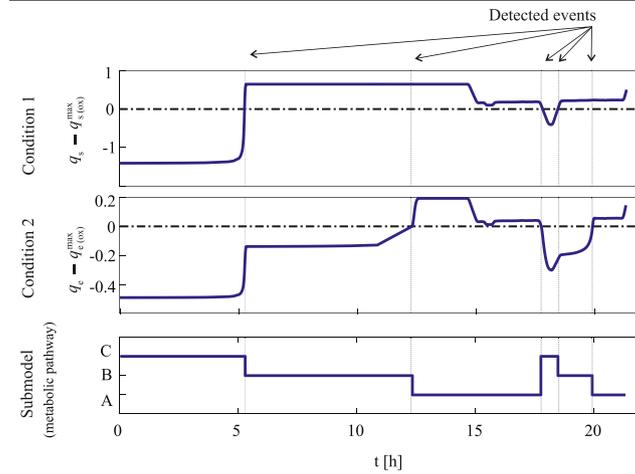
**Table 3**  
Selection of submodels according to the state of the Boolean trigger functions (True or False) and corresponding values of the activation matrix  $\mathcal{A}$  in equation (3.3).

Metabolic pathway (activated sub-model)	Trigger function		Activation Matrix $\mathcal{A}(\mathcal{E}(t))$
	$\mathcal{E}_1(t)$	$\mathcal{E}_2(t)$	
A - Glucose and Ethanol oxidation (The sum of fluxes is less or equals the maximum oxidative uptake rate)	TRUE	TRUE	$\mathcal{A} = \text{diag}[1 \ 0 \ 0 \ 1 \ 0]$
B - Glucose and Ethanol oxidation (The sum of fluxes exceeds the maximum oxidative uptake rate. Ethanol uptake is limited to the maximum oxidative capacity)	TRUE	FALSE	$\mathcal{A} = \text{diag}[1 \ 0 \ 0 \ 0 \ 1]$
C- Crabtree effect (oxidoreductive pathway)	FALSE	TRUE	$\mathcal{A} = \text{diag}[0 \ 1 \ 1 \ 0 \ 0]$

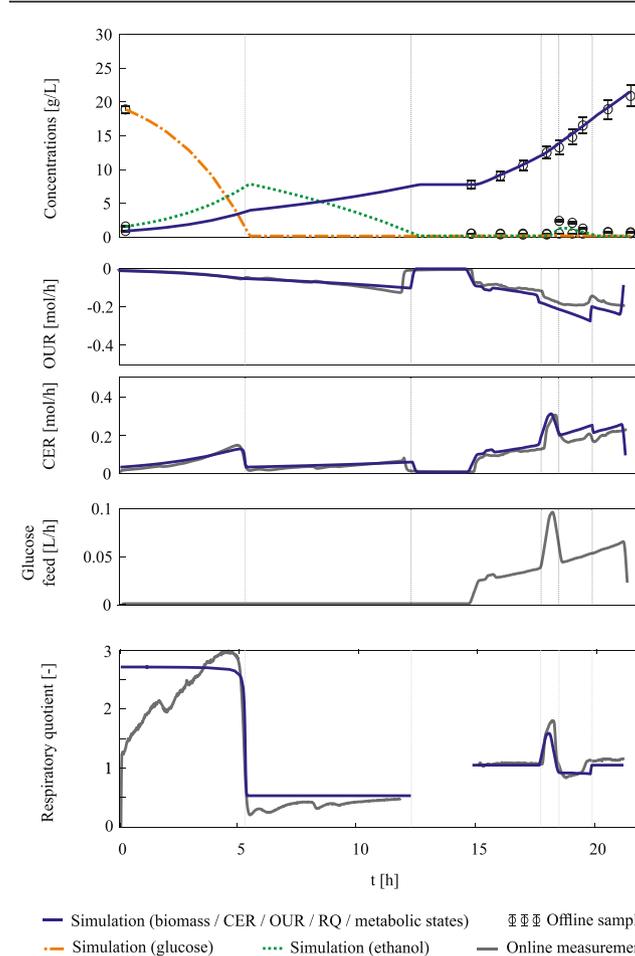
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar. The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.



Conditions & submodels



States



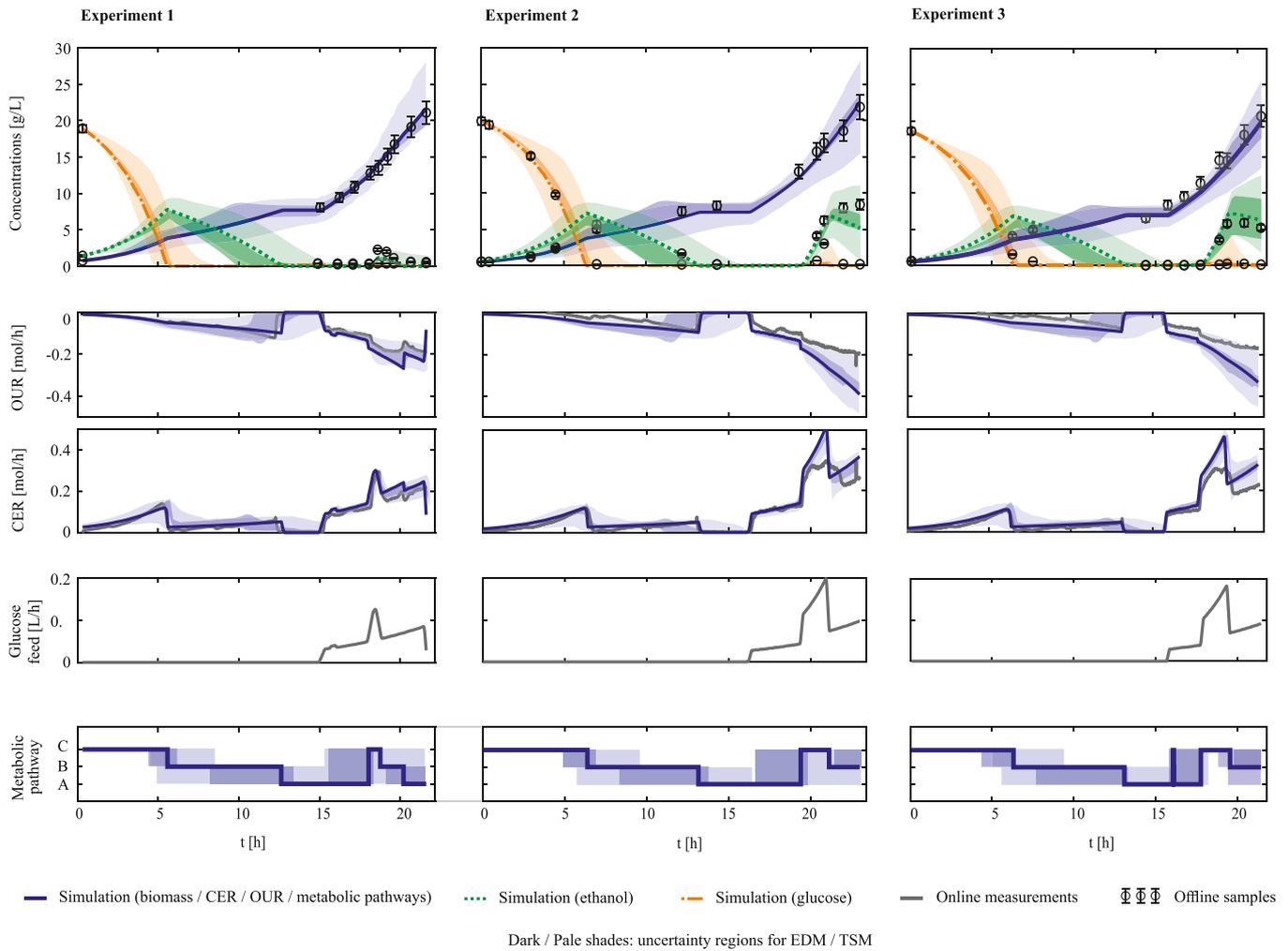
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar. The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.



**Fig. 4.** Fitting the model following the EDM implementation to data from experiment 1. Above: Trigger functions (blue) and condition thresholds (dotted line). Detected events (blue line crosses dotted line) trigger a switch between submodels. Submodel C indicates a metabolism described by the Crabtree effect, submodel B indicates oxidative growth on glucose and limited uptake rate of ethanol, submodel A indicates normal oxidative growth on ethanol and glucose. Below: Simulated and measured liquid and off-gas concentrations, feed signal and RQ signal. Because the metabolic activities between ~13 – 15 [h] stop, OUR and CER are almost zeros, and the ratio RQ is unreliable. Therefore, RQ is not shown in this time-window.

the computation of non-negative solutions. However, some *MATLAB*'s ODE solvers for stiff and nonlinear problems such as *ODE23S* and *ODE15i*, do not provide this option. Therefore, corresponding zero crossing conditions (state events) were considered in the EDM implementation.

*Feeds and samples:* Volume changes of an ideal stirred reactor are usually modeled by mass balance (differential) equations. In fed-batch fermentation, the measured flow rates (except  $F_{\text{Sampling}}$ ) usually show comparatively smooth curves. These curves are represented by discrete signals and can be transformed to smooth functions with relatively little



**Fig. 5.** Fitting the model to data from experiment 1, 2, and 3. The black solid lines and markers represent measurements, the colored lines represent the best fit (reference solution obtained by EDM). Uncertainties in the model predictions and in the location of switches between submodels are depicted by shaded areas for TSM (pale) and EDM (dark). The prediction uncertainty is obtained by a sampling considering the uncertainty in the parameter estimates, see section 2.2 for details.

effort, e.g., by applying a smoothing filter and by subsequent interpolation using piecewise spline interpolation. Avoiding discontinuities on the right-hand-side of eq. (2.18) can significantly improve the efficiency and accuracy of its solution [69], [70]. Hence, the volume changes due to sampling are modeled as instantaneous changes using time events and switches in the volume  $V_{\text{new}} = V_{\text{old}} - \Delta V_{\text{sampling}}$ , and the term  $F_{\text{sampling}}$  in eq. (2.18) therefore is omitted.

All conditions for metabolic switches, non-physical solutions and sampling are combined in one vector of switching conditions which are monitored for any root.

### 3.2. Model fitting adopting the event driven method

Fig. 4 shows results of the first *S. cerevisiae* fed-batch cultivation. The model was fitted simultaneously to the data of all three experiments using EDM. The model parameters are given in Table 4. Overall, five events were detected (Fig. 4 upper part). This means that after initialization of the simulation with submodel C, the following submodels were activated sequentially: B, A, C, B, A.

From the simulated and measured data shown in Fig. 4 (and later in Fig. 5 for the three experiments) the following conclusions can be drawn, during the batch phase (roughly between 0 and 15 h), three different phases were recognized:

- I) From the beginning of the fermentation with glucose concentration being at maximum until the point of glucose depletion that limits the glucose inflow, cells metabolize glucose both oxidatively and reductively, leading to the so-called “Crabtree effect” (metabolic pathway C) also indicated by a high RQ.
- II) Directly after glucose depletion, the previously produced ethanol is oxidized (metabolic pathway B). The metabolic transition between these first two phases is associated with an instant drop in CER whereas OUR remains unchanged due to the usage of the full oxidative capacity of the cells and leads to a RQ below 1.

Upon ethanol depletion, metabolic activity stops (OUR and CER  $\sim 0$ ) and the model changes to glucose oxidation (metabolic pathway A), which is set as the default. After the end of the batch phase, different feeding phases were started (15–23 h):

- III) After feed start ( $\sim 15$  h) subcritical glucose flux (metabolic pathway A) is aimed by a small exponential ramp. An RQ of  $\sim 1$  indicates that glucose consumption is purely oxidative.
- IV) After that, an increased glucose feed leads to ethanol formation through the “Crabtree effect” (metabolic pathway C) similar to time window (I) but for a shorter time.
- V) After sensing significant ethanol accumulation, the feed is changed to subcritical glucose fluxes and co-utilization of provided glucose and the produced ethanol can be observed (RQ  $< 1$

**Table 4**

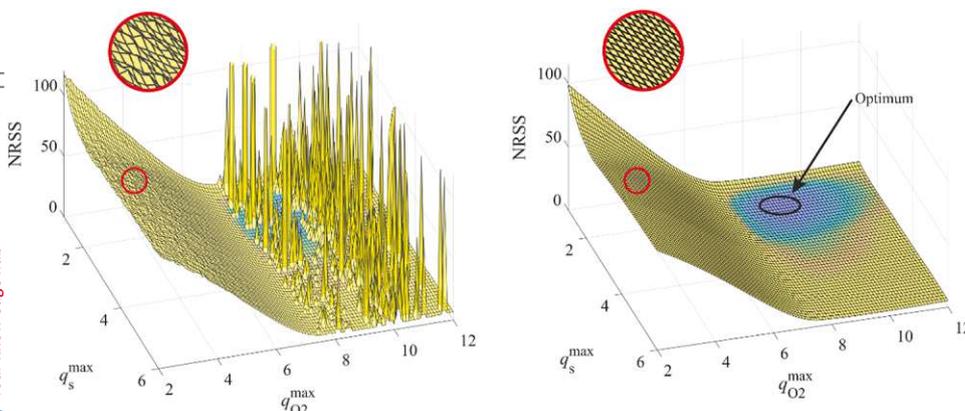
Parameter identifiability analysis and parameter uncertainty quantification considering data from experiment 1, 2, 3 and following the event driven method for model implementation. The subset selection method selects the identifiable parameter subspace. Non-identifiable parameters are fixed to their initial guess values and are not considered in the fitting problem. The lower bound of the confidence interval (LB-CI) and upper bound of the confidence interval (UB-CI) of the parameters are presented for 95% confidence interval ( $\pm 2\sigma$ ). The symbol % $\sigma$  represents the relative standard deviation of the estimated parameters.

	Initial guess	Identifiability analysis Parameter subset Selection	Uncertainty quantification			
			Estimated value $\mu$	% $\sigma$	LB-CI	UB-CI
$q_s^{max}$	3.5	Identifiable, Active	1.68	3.52	1.56	1.80
$q_{O_2}^{max}$	7.5	Identifiable, Active	8.70	2.94	8.24	9.26
$Y_{x/s(red)}$	0.05	Identifiable, Active	0.10	25.74	0.048	0.15
$Y_{x/s(ox)}$	0.5	Identifiable, Active	0.53	5.94	0.46	0.59
$Y_{x/e(ox)}$	0.72	Identifiable, Active	0.4	3.62	0.37	0.43
$q_e^{max}$	0.24	Not identifiable, Non-active	0.24	-		
$K_s$	0.1	Not identifiable, Non-active	0.105	-		
$K_o$	0.1	Not identifiable, Non-active	0.105	-		
$K_i$	0.1	Not identifiable, Non-active	0.1	-		

& metabolic pathway B). After depletion of ethanol only the added glucose is metabolized (RQ  $\sim 1$  & metabolic pathway A).

The model predictions are well aligned with measured concentrations and the off-gas signals with an overall normalized root-mean-square error below 4.5%. The location of switches and the identified submodels are in good accordance with the indications by the computed RQ, compare the switches between the identified submodels (Fig. 4 above) with the computed RQ values (Fig. 4 below).

When RQ is greater than 1, cells are, in parallel to the oxidative route, also consuming glucose in a fermentative regime causing ethanol production (metabolic pathway C). When RQ  $\approx 1$ , this is a clear indicator of purely glucose consumption in an oxidative regime (corresponding to metabolic pathway A at higher glucose concentrations, and B at lower ones). RQ  $\approx 2/3$  is a clear indicator of purely ethanol consumption in an oxidative regime (metabolic pathway B at higher ethanol concentrations, and A at lower ones).



**Fig. 6.** NRSS (normalized residual sum of squared errors) optimization surfaces of a reduced two-dimensional parameter estimation problem for the model implementation following TSM (left) and EDM (right). The surfaces are constructed by evaluations (repeated simulations) for a grid of parameters. The optimal solution is around  $q_{s(ox)}^{max} = 1.7 g \cdot h^{-1} \cdot g^{-1}$  and  $q_{O_2}^{max} = 8.7 g \cdot h^{-1} \cdot g^{-1}$ . (Note: that oxygen uptake is limiting and therefore maximum glucose uptake rate is a non-sensitive parameter). The TSM produces stochastic errors which produce a noisy surface. This noise is not static but dynamically changes throughout the evaluation procedure and impedes the optimization algorithm to converge to the minimum. Note that for TSM, the large peaks result mostly from “non-physical solutions”, see section 3.1.2.1.

The simulation results imply an immediate change in cell metabolism after each event. This is due to the assumption of Sonnleitner and Käppli's that cells can instantaneously change between metabolic pathways. Although some authors prefer to consider adaptation times after metabolic changes [57], it was reported that cells remain biochemically active during these times but cell division is highly affected [71]. Therefore, we restrict ourselves to the Sonnleitner and Käppli's assumption, as this discussion goes beyond the paper's purpose.

The results from the parameter identifiability analysis and uncertainty quantification of the three experiments are shown in Table 4. Nine parameters are considered and ordered starting with the most identifiable parameters. Five parameters are selected as identifiable. It is noted that the same ranking and selection was also found for the individual fitting of the two different experiments considered in the following subsection. An interesting finding is that  $q_e^{max}$  is not identifiable although it plays a role in one of the conditions (see eq. (2.13)), the reason might be the direct correlation between  $q_e^{max}$  and  $Y_{x/e(ox)}$  in eq. (2.17). Hence, only one of both parameters is uniquely identifiable.

A detailed analysis of the impact of uncertainties in the parameter estimates is given in section 3.3.3.

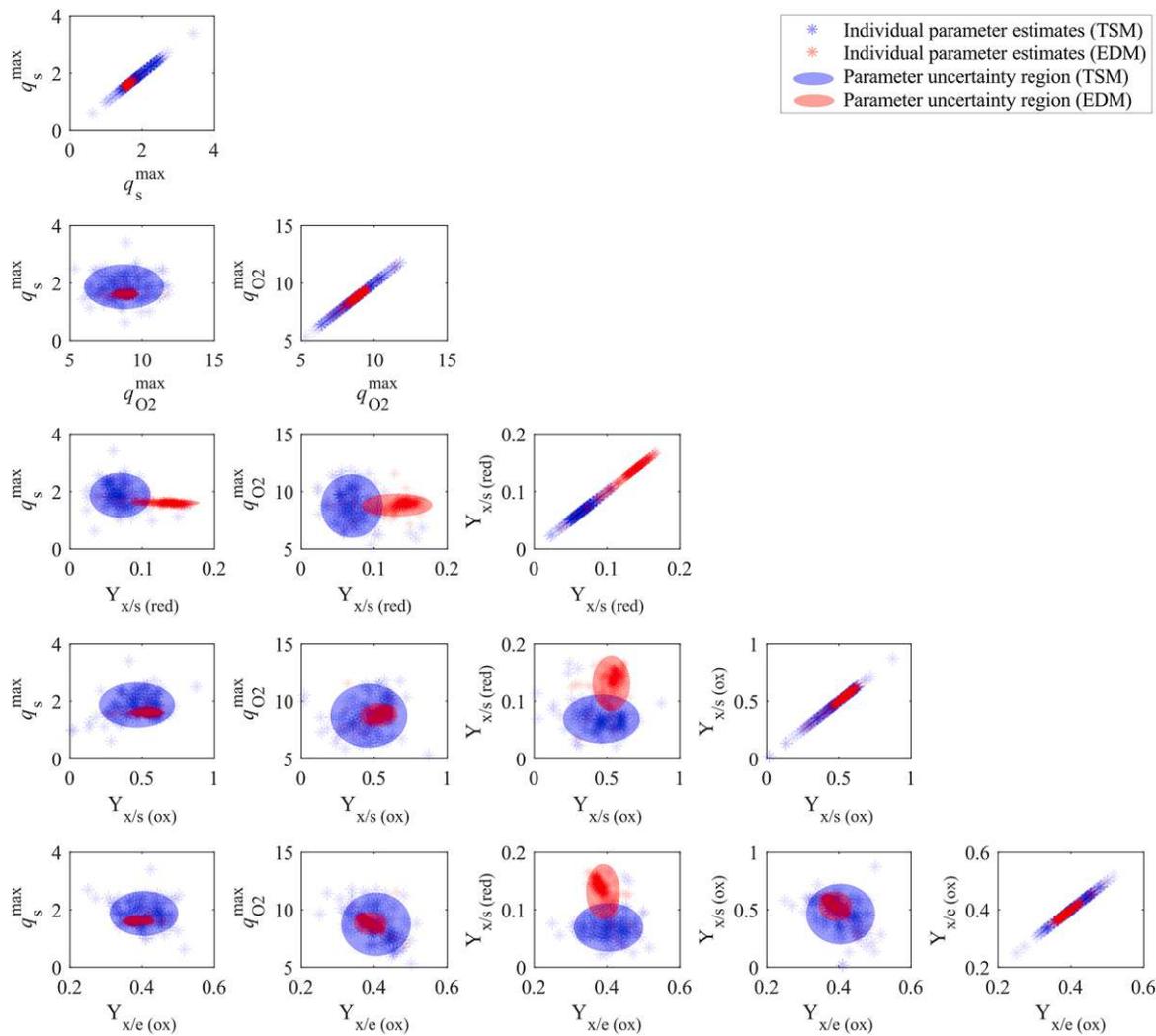
### 3.3. The advantages of using the event driven method over the time stepping method

Section 3.3 presents results for all three experiments. Note that the best fitting results obtained from the EDM are here referred to as the reference solution. Corresponding parameter estimates are reported in Table 4.

**Table 5**

Relative error (% $\sigma$ ) of the estimated parameters at different perturbation levels of the initial parameter guess. Low errors indicate a good parameter identification, whereas high errors indicate a poor identification.

	Perturbation of the initial parameter guess						
	0%	10%		30%		50%	
	Reference % $\sigma$	TSM % $\sigma$	EDM % $\sigma$	TSM % $\sigma$	EDM % $\sigma$	TSM % $\sigma$	EDM % $\sigma$
$q_s^{max}$	3.52	8.5	6.72	20.84	3.7	30.22	6.36
$q_{O_2}^{max}$	2.94	7.26	4.16	15.41	5.71	22.04	4.16
$Y_{x/s(red)}$	25.74	16.10	26.43	30.68	18.56	41.33	26.71
$Y_{x/s(ox)}$	5.94	12.53	8.55	28.54	12.27	38.99	8.5
$Y_{x/e(ox)}$	3.62	6.14	5.14	12.04	5.94	17.34	5.14



**Fig. 7.** Pairwise parameter estimates obtained by fitting 500 simulated data sets (generated by resampling the data from all experiments). For each fit the initial parameter guess was perturbed (perturbation levels 30%) from the best fitting parameter values in Table 4 (reference solution). The scatter plot shows the individual parameter estimates together with an approximation of the parameter confidence regions using ellipsoids. The confidence regions obtained by TSM are significantly larger when compared with results from EDM.

### 3.3.1. Prediction uncertainty of the identified models implemented using EDM and TSM

Fig. 5 shows results from fitting the model to the three experiments. The solid lines represent the best fit, i.e., the reference solution. The shaded areas represent the 95% confidence interval of the EDM (dark) and the TSM (pale) implementation. Overall, the solid line obtained by EDM describes nicely the discrete measurements of biomass, ethanol and glucose as well as the continuous CER and OUR measurements with an average normalized root-mean-square error below 5%. Towards the end of the fermentations the off-gas signals show slightly higher model mismatches which could be due to slower mass transfer by higher cell concentrations and/or sensor saturation or decreased sensitivities by higher CO<sub>2</sub> and lower O<sub>2</sub> concentrations in the off-gas stream.

The prediction uncertainty (see Section 2.2 for details) is a measure for the reliability of the model predictions and critically depends on the uncertainties in the parameter estimates. These uncertainties result from poor parameter sensitivities, parameter correlations and measurement errors (EDM and TSM). Using TSM, these uncertainties might be additionally increased by inaccuracies in the event detection and location of model switches. Note that these inaccuracies are also referred to as numerical noise (see Section 3.3.2).

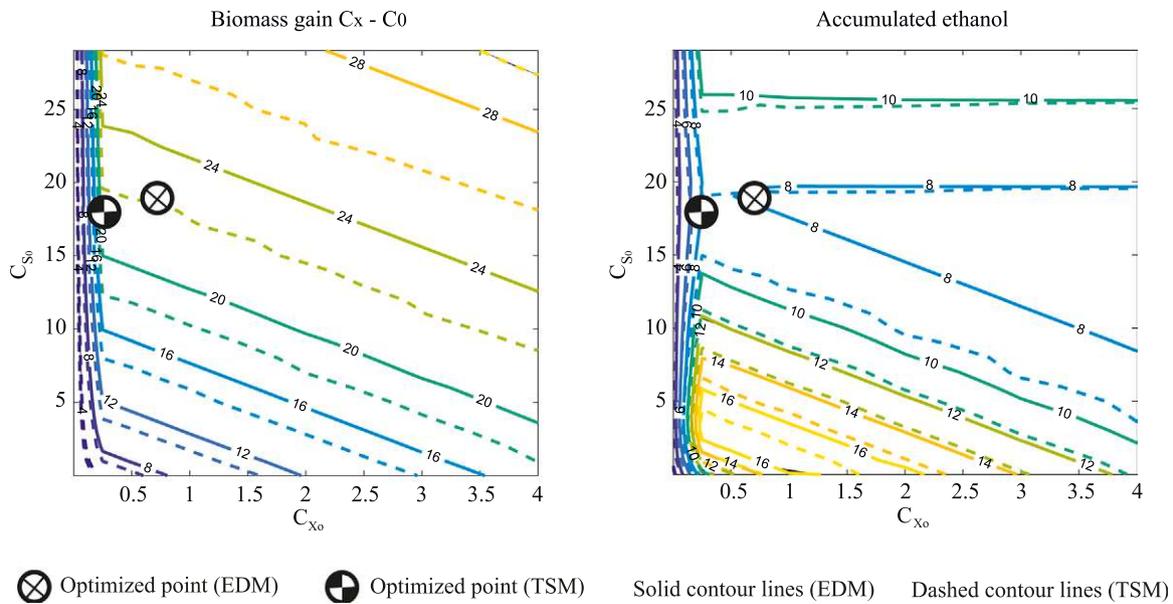
Overall, it seems that the TSM implementation has a lower predictive power compared to the EDM implementation. This can be attributed to

the effects of numerical noise. Thus, it is not surprising to see that for all three experiments in Fig. 5 the prediction uncertainties are higher for TSM compared to EDM. These results give a first indication that EDM produces more robust (reproducible) model predictions for liquid, gas and metabolic pathways.

### 3.3.2. Numerical noise and its implications for fitting the model to the measurements

Inaccurate and possibly non-physical solutions are highly undesirable for numerical analysis. Using TSM the errors in the event location are not controlled, the solution of the model can potentially be corrupted by noise. This affects the computed states as well as any quantity derived from them, such as the residuals (in a parameter estimation problem). In this situation, the objective function is a “noisy function”. This is also a problem for the computation of sensitivities and gradient information during optimization iterations, e.g., in the perturbation gradient estimation methods (using finite difference approximations), gradients are computed by evaluating the objective function in several points in the neighborhood of the current guess, using finite step sizes. Stochastic errors (noise) in the objective function values leads to errors in the computed gradients. This can be problematic for the solution of the optimization problem.

The consequences of these errors in the simulation are illustrated in



**Fig. 8.** Demonstration of the usage of the model for effective model-based DoE to optimize experiment 2. Right) The maximum accumulated ethanol, Left) The gain of biomass concentration during the time course of the experiment. Both are plotted against initial biomass and glucose concentrations. Lines: -solid line (EDM), -dashed line (TSM). All units are [g/L]. Models implemented with TSM produces shifted and curvy isolines compared to the ones implemented with EDM, when the model is evaluated at different initial concentrations.

**Fig. 6.** For TSM and EDM, the objective function surfaces are evaluated considering a simplified parameter estimation problem with two unknown parameters, the maximum oxygen ( $q_{O_2}^{max}$ ) and substrate ( $q_s^{max}$ ) uptake rates. EDM produces a continuous and differentiable surface, whereas the surface produced by TSM is noisy, discontinuous and non-differentiable. This reduces the effectiveness of the gradient-based optimization.

A quantitative analysis is given in Appendix D where the model is fitted to 500 simulated data sets (fitting problem in Section 3.3.1 considering all three experiments and all active parameters given in Table 4). Each parameter estimation problem was initialized with a perturbed parameter initial guess. The distribution of the NRSS values at the solution was used to assess the convergence of the parameter estimation, i.e., low values, close to the reference NRSS, indicate good convergence. It turns out that the results obtained by EDM are very similar indicating a robust convergence to the optimum. In contrast, the results of the TSM are clearly affected by the numerical noise produced by the inaccuracies in the location of events. Using TSM, the optimizer often gets stuck (does not converge) which results in solutions with very high residual values (NRSS).

The performance of EDM and TSM is described in a more general perspective in Appendix E where the results obtained by TSM and EDM are compared in terms of the fitting error and computation times for different ODE solver types and their error tolerances. Again, the results indicate a lower fitting error when EDM is used. However, this comes at the price of a relatively small increase in the computation times.

### 3.3.3. Numerical noise and its implications for the identifiability of parameters

Table 5 shows the relative errors of the estimated parameter which were selected by the parameter identifiability analysis, see section 2.2. The first column shows results for the reference solution (column "Reference" in Table 5 is taken from Table 4). All columns show results obtained by fitting 500 simulated data sets. The fitting was done for EDM and TSM for perturbed parameter initial guesses (and at increasing perturbation levels). It can be noted that the results for the EDM are not significantly affected by the perturbation in the initial parameter guess. The reported parameter errors are similar to the errors obtained for the

best reference solution. These results again proof the robust convergence of the optimization algorithm. In contrast, for TSM, with higher perturbation levels, the errors in the parameter estimates increase significantly.

These observations are confirmed by the scatterplots in Fig. 7 which shows the parameter confidence regions obtained by TSM and EDM for a perturbation of the initial parameter guesses by 30%. As to be expected the confidence regions obtained by TSM are significantly larger when compared with results from EDM. This can be seen for example for the pair  $q_s^{max}$  and  $q_{O_2}^{max}$ , which are also important triggers to switch between the different growth pathways (see equations (2.12) and (2.14)). Interestingly enough, the confidence ellipsoids obtained by TSM are not only larger, but their location is also different (e.g.,  $Y_{x/s(red)}$  and  $Y_{x/s(ox)}$ ). This means that besides a larger parameter uncertainty, the usage of TSM also leads to different estimates.

The inflated confidence regions found by TSM mean in practical terms, that the TSM is not able to accurately determine the key physiological characteristics of the cell, namely maximum rates and conversion yields. This could extend by collinearity to influence other important parameters. Hence, it is not feasible for bioprocessing engineers who rely on TSM modeling to identify a reliable set of model parameters for such a process.

### 3.3.4. Numerical noise and its implications for the uncertainty in model predictions

In this section, different initial concentrations are evaluated to predict the potential process behavior.

Fig. 8 shows the isolines (contours) of two different objectives: the biomass gain and the reached ethanol concentration. Both quantities are related to the initial biomass and glucose concentrations which define the two-dimensional design space.

The combined goal is to avoid excessive ethanol formation (oxidoreductive pathway) while at the same time maximizing the cell growth (biomass concentration). Considering the isoline based on EDM, the optimal operating point is around: 1 [g/L] initial biomass and 19 [g/L] initial glucose concentrations. The isolines of the TSM model are shifted with an optimum at 0.5 [g/L] initial biomass and 17 [g/L] glucose. Compared to EDM, this yields a reduced process performance with 25%

lower biomass. Besides that, another effect can be seen within Fig. 8. Compared to the EDM, the TSM model shows a curvy behavior (dashed isolines), i.e., repeated evaluation of the model with small differences in the initial concentrations produces significant shifts in the isolines which makes them harder to interpret. It can be concluded that predictions based on TSM implementation are unreliable and therefore cannot be recommended for simulation-based optimization.

As in many other biotechnological upstream processes which aim mostly for maximum cell yield, this process must run close-to-optimal conditions, e.g., oxidative metabolism close to the boundaries of oxidative capacity of the cell, which implies running very close to the metabolic (boundaries) switches, e.g., bottleneck kinetics, and causes the simulation to be highly sensitive to inaccurate event location.

EDM by explicitly accounting for these switches, ensures the simulation to run without any deviations and therefore keeping the predictions on track.

#### 4. Discussion and conclusions

Although successfully employed in other fields, the explicit consideration of events and switches in bioprocess modeling seems underestimated and still not sufficiently exploited. Different sources of discontinuities still limit the usage of process models in biotechnological processes. This includes operational discontinuities, such as instantaneous feed addition or offline sampling as well as metabolic changes triggered by inducer addition or internal process dynamics during the batch and fed-batch operations. Besides the need to explicitly account for these sudden changes, a sound implementation allows for more reliable and generically applicable models which can be used for process design, monitoring and targeted control of cell metabolism in an industrial context.

A robust modeling approach has been developed for the respiro-fermentative growth of *S. cerevisiae*. This has been achieved by the consideration of metabolic switches as events in the framework of an earlier established model [55]. The comparative analysis of the proposed EDM for model implementation, and the simpler and often used TSM underlines that models implemented with EDM deliver more

#### Appendices

##### Appendix A. Numerical implementation of the event driven method

###### Boolean trigger functions

Events are triggered when the sign of the condition in (eq. 2.3), changes, i.e., a zero crossing in  $c(\cdot)$  is detected. It is noted that [38] use a more general definition of state conditions which represent logical propositions. These logical propositions may contain a number of relational expressions and sets of connectives (e.g., NOT, AND, OR). However, in this contribution Boolean trigger functions  $\mathcal{C}(t)$  are used [27]. For  $\mathcal{C}(t)$ , output values true are mapped to positive values and output values false are mapped to negative values such that the corresponding state condition  $c(\cdot)$  has a root at every change of the corresponding Boolean values:

$$\mathcal{C}(t) := \begin{cases} 1, & \text{if } c(x(t), u(t), \theta, t) \geq 0 \\ 0, & \text{if } c(x(t), u(t), \theta, t) < 0 \end{cases} \quad (\text{A.1})$$

Moreover, in this contribution, a decision tree is formulated out of the values of the Boolean functions which links different conditions. This approach allows for a more straightforward implementation where the monitoring of conditions and detection of events is decoupled from the evaluation of logical operations. Thus, once, one or more events are detected, the new active submodel is selected based on the evaluation of the decision tree.

###### Discontinuity locking

In the event driven method, the system of equations for each subinterval is locked throughout the solution. This means that the system of equations cannot change even if one or more state conditions are satisfied [38]. The state conditions are monitored continuously, and if any of them are satisfied, the exact time of occurrence is then located, equations are switched, new initial states might be calculated, and the integration is restarted. This approach is efficient and correct provided that the system of equations employed before the state event is mathematically well behaved in a small interval following the state event (even if the solution is not physically meaningful) [77].

accurate location of metabolic switches, lower prediction error and lower parameter uncertainty.

The results of the presented case study encourage further investigations using EDM modeling with other interesting discontinuous behaviors. The *S. cerevisiae* growth model could be adapted to multi-substrate mixtures by considering additional pathways and potential interactions such as diauxic growth. Switches in the reaction routes (similar to the switches in this contribution) are conceivable for the consideration of overflow metabolism for *E. coli* [65], [72], or for Crabtree-positive *P. pastoris* [73], [74]. External and auto-induced production switches in recombinant protein production in *S. cerevisiae*, *E. coli* or *P. pastoris* and other organisms imply critical changes in cell metabolism. The induction itself might be externally triggered (e.g., [73], [74]), and formulated as a process related switch, or, in case of auto-induction (e.g., phosphate starvation for *E. coli* and *P. pastoris* [75], [76]), an implicit switch could be formulated (e.g., depending on available substrate), which triggers a change to a new reaction route for product formation.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgments

This work was partially funded by the Austrian Research Funding Association (FFG) within both: the program Bridge 1 in the project "AdaMo" (No. 864705), and the COMET Centre CHASE (No. 868615), funded within the COMET – Competence Centers for Excellent Technologies programme by the BMK, the BMDW and the Federal Provinces of Upper Austria and Vienna.

The authors thank Boehringer Ingelheim RCV GmbH & Co KG for their participation in the project and for the fruitful discussions, and acknowledge TU Wien Bibliothek for the financial support through its Open Access Funding Programme.

### Chattering control

When there are many discontinuity points in a small-time interval, the system is said to have a chattering behavior. In this situation, the use of an event location routine can lead to an expensive procedure [36]. Chattering can be observed for solutions which produce a sliding along the critical threshold of a certain condition without a clear threshold crossing. Measured noisy signals as time dependent inputs to the model, e.g., a measured feed, might also produce chattering. In this contribution, chattering was found for signals with high frequent noise when the monitored conditions are close to a critical threshold.

In order to reduce chattering, in this contribution, a hysteresis band is defined for the threshold in each condition as follows:

$$\begin{aligned} c(x(t), u(t), \theta, t) &= -\epsilon & \text{if } \mathcal{E}(t) &= 1 \\ c(x(t), u(t), \theta, t) &= +\epsilon & \text{if } \mathcal{E}(t) &= 0 \end{aligned} \quad (\text{A.2})$$

where the magnitude of  $\epsilon$  defines the magnitude of the hysteresis band. It can be seen that the sign of  $\epsilon$  depends on the current state of the Boolean trigger function  $\mathcal{E}(t)$ . Alternatively, the hysteresis band could also be defined based on the sign of the rate of change of  $c(\cdot)$ . Here, for positive rates,  $dc/dt \geq 0$ , a positive  $+\epsilon$  is used, while for negative rates,  $dc/dt < 0$ , a negative  $-\epsilon$  is used.  $\epsilon$  is a tuning parameter, its value needs to be chosen individually for the specific problem, keeping in mind that using small values avoids any delay effects in the event detection.

### Appendix B. Experimental design and calculation of stoichiometric coefficients calculations

Yeast fermentation media and process parameters are shown in Tables B1-B2.

The calculations of the stoichiometric coefficients for known molecular weights of HX, OX, NX, can be done by solving the linear system in eq. (B.1).

$$\begin{array}{c} \mathbf{V} \\ \mathbf{x} \\ \mathbf{s} \\ \mathbf{e} \\ \mathbf{CO}_2 \\ \mathbf{o}_2 \\ \mathbf{NH}_3 \\ \mathbf{H}_2\mathbf{O} \end{array} \begin{array}{ccc} \mathbf{r1} & \mathbf{r2} & \mathbf{r3} \\ \left[ \begin{array}{ccc} 0 & 0 & 0 \\ \alpha_2 & \alpha_5 & \alpha_{10} \\ 1 & 1 & 0 \\ 0 & \alpha_8 & -1 \\ \alpha_3 & \alpha_6 & \alpha_{11} \\ -\alpha_1 & 0 & -\alpha_9 \\ -\alpha_2 & -\alpha_5 & -\alpha_{10} \\ \alpha_4 & \alpha_8 & \alpha_{12} \end{array} \right] \end{array} \cdot \begin{array}{c} \mathbf{V} \\ \mathbf{x} \\ \mathbf{s} \\ \mathbf{e} \\ \mathbf{CO}_2 \\ \mathbf{o}_2 \\ \mathbf{NH}_3 \\ \mathbf{H}_2\mathbf{O} \end{array} \begin{array}{c} \left[ \begin{array}{cccccccc} 0 & 1 & 6 & 2 & 1 & 0 & 0 & 0 \\ 0 & \text{HX} & 12 & 6 & 0 & 0 & 3 & 2 \\ 0 & \text{OX} & 6 & 1 & 2 & 2 & 0 & 1 \\ 0 & \text{NX} & 0 & 0 & 0 & 0 & 1 & 0 \end{array} \right] \end{array} \begin{array}{c} \mathbf{C} \\ \mathbf{H} = 0 \\ \mathbf{O} \\ \mathbf{N} \end{array} \quad (\text{B.1})$$

**Table B.1**  
Yeast fermentation media.

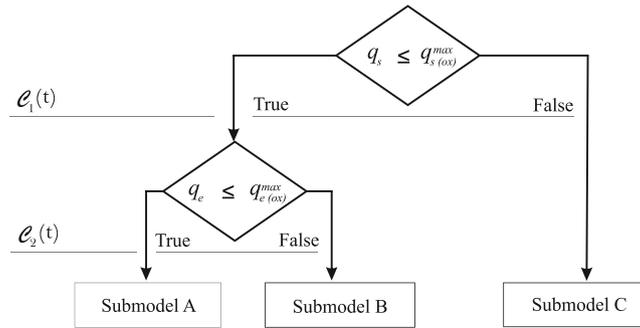
	Batch (1,5 L)	Fed-Batch (1 L)
Glucose monohydrate	33 g	220/200/198 g
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	7,5 g	5 g
KH <sub>2</sub> PO <sub>4</sub>	4,5 g	3 g
MgSO <sub>4</sub> * 7H <sub>2</sub> O	0,75 g	0,5 g
Struktol J 650	0,1 mL	0,1 mL
Trace Elements 750 x	1995 mL	1,33 mL
Vitamins* 750 x	1995 mL	1,33 mL

**Table B.2**  
Process parameters.

Culture	Saccharomyces cerevisiae, CBS 8340, Wild type
pH setpoint	4,8
Temperature set	30 °C
Agitator Speed	1000 rpm
Air flow	2,25 L/min (1,5 vvm)
Base	2 M NaOH
Base density (2 M NaOH)	1080 g/L
Feed density	1078 g/L

Appendix C. Decision tree

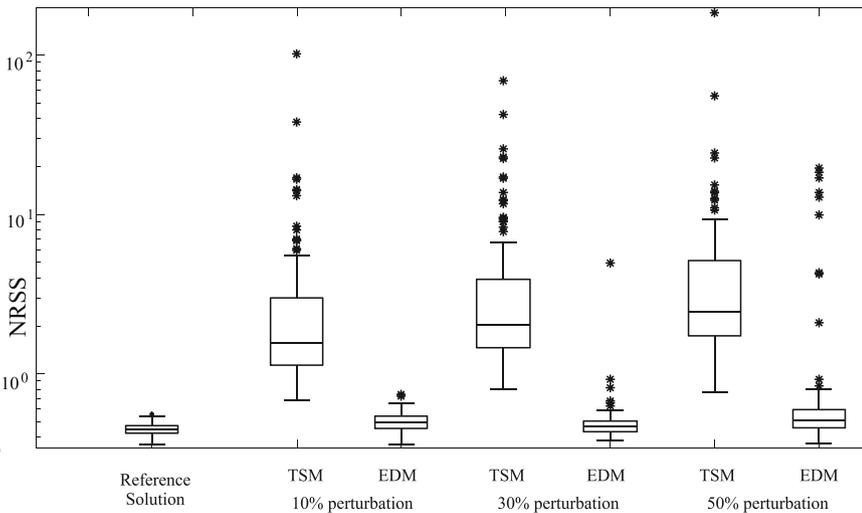
See Fig. C1.



**Fig. C1.** Decision tree with two Boolean trigger functions  $\mathcal{C}_1(t)$ ,  $\mathcal{C}_2(t)$  and three submodels for the *Saccharomyces cerevisiae* fermentation model. Switches are made based on the metabolic flux capacity “bottleneck concept”. The decision tree is part of EDM computational scheme. EDM monitors the switching conditions, when an event is located, the submodel is switched.

Appendix D. Numerical noise and its implications for the convergence of the optimization algorithm

Fig. D1 shows the normalized residual sum of squared errors (NRSS) for models implemented by TSM and EDM, and obtained from fitting 500 simulated data sets at different perturbation levels of the initial parameter guesses of the identifiable parameter subset. These initial guesses were perturbed at three levels, 10%, 30% and 50%, as described in Section 2.2. While the NRSS obtained by EDM is very similar for all perturbation levels, the results of the TSM clearly increase for increased perturbation levels.



**Fig. D1.** Normalized residual sum of squares (NRSS) box plots obtained from fitting 500 simulated data sets (generated by resampling the data from all experiments) and starting with initial parameter values at different perturbation levels (10%,30% and 50%) from the best fitting parameter values in Table 4. The reference solution was obtained by EDM using the best fitting parameter values as initial parameter guess. Each box plot shows the interquartile range (IQR), lower and upper 1.5 \*IQR whiskers, median and outliers results.

Appendix E. Fitting errors versus computation times

Table E1 shows comparative results of fitting errors and computation times using TSM and EDM, and using different ODE solvers and error tolerances for their numerical solution. The analysis is based on the results obtained by fitting 50 simulated data sets (due to the very long computation times needed at lower error tolerances). The perturbation level of the parameter guesses is set to 50%. The performance is defined by the ratio of the means of the prediction error as:  $\Delta J_{EDM/TSM} = \frac{\mu(NRSS(\theta)_{EDM})}{\mu(NRSS(\theta)_{TSM}}$ , and the ratio in the simulation time as:  $\Delta t_{EDM/TSM} = \frac{\mu(t_{EDM})}{\mu(t_{TSM})}$ .

The results always show a better chance for EDM to get smaller fitting errors. However, this comes at the cost of increased computation times, here usually between 14% and 54%.

The table also shows that the fitting error difference for non-stiff solvers such as ODE45 and ODE23 is less than for their counterpart stiff solvers. This is because of the very small step size adopted by non-stiff solvers when applied to stiff problems. Using TSM, a smaller step size means more accurate detection of switches, less numerical noise. This improves the convergence and therefore the fitting error. For ODE23s, the problem was not solvable at many initializations with TSM. Here the solver “runs forever” without giving any results.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar. The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.



Table E1

Fitting errors and ( $\Delta J_{EDM/TSM}$ ) and computation times ( $\Delta t_{EDM/TSM}$ ) for EDM and TSM.

RelTol/Solver	ODE15S		ODE23S		ODE45		ODE23	
	$\Delta J_{EDM/TSM}$	$\Delta t_{EDM/TSM}$	$\Delta J_{EDM/TSM}$	$\Delta t_{EDM/TSM}$	$\Delta J_{EDM/TSM}$	$\Delta t_{EDM/TSM}$	$\Delta J_{EDM/TSM}$	$\Delta t_{EDM/TSM}$
$10^{-3}$	0.20	1.27	Not solvable at many initializations		0.10	1.14	0.58	1.19
$10^{-6}$	0.31	1.19	Not solvable at many initializations		0.76	1.43	0.71	1.30
$10^{-9}$	0.43	1.22	Not solvable at many initializations		0.65	1.29	0.81	1.54

## References

- [1] I. Gosling, Process simulation and modeling for industrial bioprocessing: tools and techniques, *Ind. Biotechnol.* 1 (2) (2005) 106–109, <https://doi.org/10.1089/ind.2005.1.106>.
- [2] A. Toumi, C. Jürgens, C. Jungo, B.A. Maier, V. Papavasileiou, D. Petrides, Design and optimization of a large scale biopharmaceutical facility using process simulation and scheduling tools, *Pharm. Eng.* 30 (2010) 34–36.
- [3] J.E. Jiménez-Hornero, I.M. Santos-Dueñas, I. García-García, Optimization of biotechnological processes. The acetic acid fermentation. Part I: The proposed model, *Biochem. Eng. J.* 45 (1) (2009) 1–6, <https://doi.org/10.1016/j.bej.2009.01.009>.
- [4] V. Abt, T. Barz, M.N. Cruz-Bournazou, C. Herwig, P. Kroll, J. Möller, R. Pörtner, R. Schenkendorf, Model-based tools for optimal experiments in bioprocess engineering, *Curr. Opin. Chem. Eng.* 22 (2018) 244–252, <https://doi.org/10.1016/j.coche.2018.11.007>.
- [5] M.N. Cruz Bournazou, T. Barz, D.B. Nickel, D.C. Lopez Cárdenas, F. Glauche, A. Knepper, P. Neubauer, Online optimal experimental re-design in robotic parallel fed-batch cultivation facilities, *Biotechnol. Bioeng.* 114 (3) (2017) 610–619, <https://doi.org/10.1002/bit.26192>.
- [6] H. Narayanan, et al., Bioprocessing in the Digital Age: The Role of Process Models, *Biotechnol. J.* 15 (2020), <https://doi.org/10.1002/biot.201900172>.
- [7] D. Dochain, State and parameter estimation in chemical and biochemical processes: a tutorial, *J. Process Control* 13 (8) (2003) 801–818, [https://doi.org/10.1016/S0959-1524\(03\)00026-X](https://doi.org/10.1016/S0959-1524(03)00026-X).
- [8] C. Dietzsch, O. Spadiut, C. Herwig, On-line multiple component analysis for efficient quantitative bioprocess development, *J. Biotechnol.* 163 (4) (2013) 362–370, <https://doi.org/10.1016/j.jbiotec.2012.03.010>.
- [9] S. Natarajan, J.H. Lee, Repetitive model predictive control applied to a simulated moving bed chromatography system, *Comput. Chem. Eng.* 24 (2–7) (2000) 1127–1133, [https://doi.org/10.1016/S0098-1354\(00\)00493-2](https://doi.org/10.1016/S0098-1354(00)00493-2).
- [10] R. Singh, K.V. Gernaey, R. Gani, Model-based computer-aided framework for design of process monitoring and analysis systems, *Comput. Chem. Eng.* 33 (1) (2009) 22–42, <https://doi.org/10.1016/j.compchemeng.2008.06.002>.
- [11] M. Degerman, K. Westerberg, B. Nilsson, Determining critical process parameters and process robustness in preparative chromatography - A model-based approach, *Chem. Eng. Technol.* 32 (6) (2009) 903–911, <https://doi.org/10.1002/ceat.200900019>.
- [12] P.M. Doran, *Bioprocess Engineering Principles: Second edition*, Academic Press, 2012.
- [13] A.S. Rathore, H. Winkle, Quality by design for biopharmaceuticals, *Nat. Biotechnol.* 27 (2009) 26–34.
- [14] J. Nielsen, *Physiological Engineering Aspects of Penicillium Chrysogenum*, WORLD SCIENTIFIC, 1997.
- [15] K.V. Gernaey, A.E. Lantz, P. Tufvesson, J.M. Woodley, G. Sin, Application of mechanistic models to fermentation and biocatalysis for next-generation processes, *Trends Biotechnol.* 28 (7) (2010) 346–354, <https://doi.org/10.1016/j.tibtech.2010.03.006>.
- [16] A. Meitz, P. Sagmeister, W. Lubitz, C. Herwig, T. Langemann, Fed-batch production of bacterial ghosts using dielectric spectroscopy for dynamic process control, *Microorganisms* 4 (2) (2016), <https://doi.org/10.3390/microorganisms4020018>.
- [17] J.J. DiStefano, *Dynamic Systems Biology Modeling and Simulation*, Academic Press, 2015.
- [18] A.K. Gombert, J. Nielsen, Mathematical modelling of metabolism, *Curr. Opin. Biotechnol.* 11 (2) (2000) 180–186, [https://doi.org/10.1016/S0958-1669\(00\)00079-3](https://doi.org/10.1016/S0958-1669(00)00079-3).
- [19] O.D. Kim, M. Rocha, P. Maia, A review of dynamic modeling approaches and their application in computational strain optimization for metabolic engineering, *Front. Microbiol.* 9 (2018) 1690, <https://doi.org/10.3389/fmicb.2018.01690>.
- [20] A. Tsopanoglou, I. Jiménez del Val, Moving towards an era of hybrid modelling: advantages and challenges of coupling mechanistic and data-driven models for upstream pharmaceutical bioprocesses, *Curr. Opin. Chem. Eng.* 32 (2021), 100691, <https://doi.org/10.1016/J.COACHE.2021.100691>.
- [21] N.J. Stanford, T. Lubitz, K. Smallbone, E. Klipp, P. Mendes, W. Liebermeister, Systematic construction of kinetic models from genome-scale metabolic networks, *PLOS One* 8 (11) (2013), e79195, <https://doi.org/10.1371/journal.pone.0079195>.
- [22] M.L. Shuler, J.D. Varner, *Cell growth dynamics*. Comprehensive Biotechnology, Pergamon, 2011, pp. 32–38.
- [23] S. Ulonska, D. Waldschitz, J. Kager, C. Herwig, Model predictive control in comparison to elemental balance control in an E. coli fed-batch, *Chem. Eng. Sci.* 191 (2018) 459–467, <https://doi.org/10.1016/J.CES.2018.06.074>.
- [24] M.N. Cruz Bournazou, T. Barz, D.B. Nickel, D.C. Lopez Cárdenas, F. Glauche, A. Knepper, P. Neubauer, Online optimal experimental re-design in robotic parallel fed-batch cultivation facilities, *Biotechnol. Bioeng.* 114 (3) (2017) 610–619, <https://doi.org/10.1002/BIT.26192>.
- [25] E.A. del Rio-Chanona, D. Zhang, V.S. Vassiliadis, Model-based real-time optimisation of a fed-batch cyanobacterial hydrogen production process using economic model predictive control strategy, *Chem. Eng. Sci.* 142 (2016) 289–298, <https://doi.org/10.1016/J.CES.2015.11.043>.
- [26] P. Noll, M. Henkel, History and evolution of modeling in biotechnology: modeling & simulation, application and hardware performance, *Comput. Struct. Biotechnol. J.* 18 (2020) 3309–3323, <https://doi.org/10.1016/J.CSBJ.2020.10.018>.
- [27] F. Fröhlich, F.J. Theis, J.O. Rädler, J. Hasenauer, Parameter estimation for dynamical systems with discrete events and logical operations, *Bioinformatics* 33 (7) (2017) 1049–1056, <https://doi.org/10.1093/bioinformatics/btw764>.
- [28] M. Cox, D.L. Nelson, *Lehninger Principles of Biochemistry*, Palgrave Macmillan, 2008.
- [29] H.G. Crabtree, Observations on the carbohydrate metabolism of tumours, *Biochem. J.* 23 (3) (1929) 536–545, <https://doi.org/10.1042/bj0230536>.
- [30] M. Huang, J. Bao, J. Nielsen, Biopharmaceutical protein production by *Saccharomyces cerevisiae*: current state and future prospects, *Pharm. Bioprocess* 2 (2) (2014) 167–182, <https://doi.org/10.4155/pbp.14.8>.
- [31] T. Gao, Y. Ren, S. Li, X. Lu, H. Lei, Immune response induced by oral administration with a *Saccharomyces cerevisiae*-based SARS-CoV-2 vaccine in mice, *Microb. Cell Fact.* 20 (1) (2021) 95, <https://doi.org/10.1186/s12934-021-01584-5>.
- [32] K.M. Yang, N.R. Lee, J.M. Woo, W. Choi, M. Zimmermann, L.M. Blank, J.B. Park, Ethanol reduces mitochondrial membrane integrity and thereby impacts carbon metabolism of *Saccharomyces cerevisiae*, *FEMS Yeast Res* 12 (6) (2012) 675–684, <https://doi.org/10.1111/j.1567-1364.2012.00818.x>.
- [33] P.I. Barton, C.C. Pantelides, Modeling of combined discrete/continuous processes, *AIChE J.* 40 (6) (1994) 966–979, <https://doi.org/10.1002/aic.690400608>.
- [34] C.C. Pantelides, P.I. Barton, Equation-oriented dynamic simulation current status and future perspectives, *Comput. Chem. Eng.* 17 (1993) S263–S285, [https://doi.org/10.1016/0098-1354\(93\)80240-N](https://doi.org/10.1016/0098-1354(93)80240-N).
- [35] P.I. Barton, C.K. Lee, Modeling, simulation, sensitivity analysis, and optimization of hybrid systems, *ACM Trans. Model. Comput. Simul.* 12 (4) (2002) 256–289, <https://doi.org/10.1145/643120.643122>.
- [36] L. Dieci, L. Lopez, A survey of numerical methods for IVPs of ODEs with discontinuous right-hand side, *J. Comput. Appl. Math.* 236 (16) (2012) 3967–3991, <https://doi.org/10.1016/j.cam.2012.02.011>.
- [37] R.L. Burden and J.D. Faires, *Numerical Analysis*, Ninth Edit. Boston, Richard Stratton, 2011.
- [38] T. Park, P.I. Barton, State event location in differential-algebraic models, *ACM Trans. Model. Comput. Simul.* 6 (2) (1996) 137–165, <https://doi.org/10.1145/232807.232809>.
- [39] E. Fredriksson, C. Andersson, and J. Åkesson, “Discontinuities handled with events in Assimulo,” In: Proceedings of the 10th International Modelling Conference. March 10–12, 2014, Lund, Sweden, vol. 96, pp. 827–836, 2014, doi: 10.3384/ecp14096827.
- [40] P.T. Piironen, Y.A. Kuznetsov, An event-driven method to simulate Filippov systems with accurate computing of sliding motions, *ACM Trans. Math. Softw.* 34 (3) (2008) 1–24, <https://doi.org/10.1145/1356052.1356054>.
- [41] E. Kofman, A third order discrete event method for continuous system simulation, *Lat. Am. Appl. Res.* 36 (2) (2006) 101–108.
- [42] E. Kofman, Discrete event simulation of hybrid systems, *SIAM J. Sci. Comput.* 25 (5) (2004) 1771–1797, <https://doi.org/10.1137/S1064827502418379>.
- [43] S. Engell, S. Kowalewski, C. Schulz, O. Stursberg, Continuous-discrete interactions in chemical processing plants, *Proc. IEEE* 88 (7) (2000) 1050–1068, <https://doi.org/10.1109/5.871308>.
- [44] B. Furenes, B. Lie, Using event location in finite-difference methods for phase-change problems, *Numer. Heat. Transf. Part B Fundam.* 50 (2) (2006) 143–155, <https://doi.org/10.1080/10407790500459338>.
- [45] S.P. Corwin, S. Thompson, S.M. White, Solving ODEs and DDEs with impulses, *J. Numer. Anal. Ind. Appl. Math.* 3 (1–2) (2008) 139–149.
- [46] L.F. Shampine, M.W. Reichelt, Ode Suite, *J. Sci. Comput.* 18 (1997) 1–22, <https://doi.org/10.1137/S1064827594276424>.
- [47] A.C. Hindmarsh, P.N. Brown, K.E. Grant, S.L. Lee, R. Serban, D.E. Shumaker, C. S. Woodward, SUNDIALS: suite of nonlinear and differential/algebraic equation solvers, *ACM Trans. Math. Softw.* 31 (3) (2005) 363–396, <https://doi.org/10.1145/1089014.1089020>.
- [48] L.F. Shampine, S. Thompson, Event location for ordinary differential equations, *Comput. Math. Appl.* 39 (5–6) (2000) 43–54, [https://doi.org/10.1016/S0898-1221\(00\)00045-6](https://doi.org/10.1016/S0898-1221(00)00045-6).

- [49] I. Gladwell, L.F. Shampine, S. Thompson, *Solving ODEs with MATLAB*, Cambridge University Press, 2003.
- [50] K. Soetaert, T. Petzoldt, D. Germany, and R.W. Setzer, Package deSolve: Solving Initial Value Differential Equations in R.
- [51] M. Sofroniou and R. Knapp, "Advanced Numerical Differential Equation Solving in Mathematica," ... Inc. URL <http://reference.wolfram.com/mathematica/> ..., pp. 1–372, 2008, Accessed: Apr. 05, 2020. [Online]. Available: [www.wolfram.com/services/customerservice](http://www.wolfram.com/services/customerservice).
- [52] S. Kowalewski, et al., *A Case Study in Tool-Aided Analysis of Discretely Controlled Continuous Systems: The Two Tanks Problem*, Springer, Berlin, Heidelberg, 1999, pp. 163–185.
- [53] P.J. Mosterman, "An overview of hybrid simulation phenomena and their support by simulation packages," Lect. Notes Comput. Sci. (Incl. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinforma. ) vol. 1569 (1999) 165–177, [https://doi.org/10.1007/3-540-48983-5\\_17](https://doi.org/10.1007/3-540-48983-5_17).
- [54] H. Lundvall, P. Fritzon, and B. Bachmann, Event Handling in the OpenModelica Compiler and Runtime System.
- [55] B. Sonnleitner, O. Käppli, Growth of *Saccharomyces cerevisiae* is controlled by its limited respiratory capacity: formulation and verification of a hypothesis, *Biotechnol. Bioeng.* 28 (6) (1986) 927–937, <https://doi.org/10.1002/bit.260280620>.
- [56] D.C. López, T. Barz, M. Peñuela, A. Villegas, S. Ochoa, G. Wozny, Model-based identifiable parameter determination applied to a simultaneous saccharification and fermentation process model for bio-ethanol production, *Biotechnol. Prog.* 29 (4) (2013) 1064–1082, <https://doi.org/10.1002/btpr.1753>.
- [57] J. Scheiblaue, S. Scheiner, M. Joks, B. Kavsek, Fermentation of *Saccharomyces cerevisiae* – combining kinetic modeling and optimization techniques points out avenues to effective process design, *J. Theor. Biol.* 453 (2018) 125–135, <https://doi.org/10.1016/j.jtbi.2018.05.016>.
- [58] D.C. López, C. T. S. Barz, Körkel, G. Wozny, Nonlinear ill-posed problem analysis in model-based parameter estimation and experimental design, *Comput. Chem. Eng.* 77 (2015) 24–42, <https://doi.org/10.1016/J.COMPHEMENG.2015.03.002>.
- [59] B. Efron, R. Tibshirani, *An Introduction to the Bootstrap*, CRC Press, 1994.
- [60] R.C.M. Brekelmans, L.T. Driessen, H.J.M. Hamers, D. Den Hertog, Gradient estimation schemes for noisy functions, *J. Optim. Theory Appl.* 126 (3) (2005) 529–551, <https://doi.org/10.1007/s10957-005-5496-2>.
- [61] M. Aehle, K. Bork, S. Schaep, A. Kuprijanov, R. Horstkorte, R. Simutis, A. Lübbert, Increasing batch-to-batch reproducibility of CHO-cell cultures using a model predictive control approach, *Cytotechnology* 64 (6) (2012) 623–634, <https://doi.org/10.1007/s10616-012-9438-1>.
- [62] J.G.J. Dekkers, H.E. de Kok, J.A. Roels, Energetics of *Saccharomyces cerevisiae* CBS 426: comparison of anaerobic and aerobic glucose limitation, *Biotechnol. Bioeng.* 23 (5) (1981) 1023–1035, <https://doi.org/10.1002/BIT.260230510>.
- [63] K.J. Flynn, Do we need complex mechanistic photoacclimation models for phytoplankton? *Limnol. Oceanogr.* 48 (6) (2003) 2243–2249, <https://doi.org/10.4319/lo.2003.48.6.2243>.
- [64] H.Y. Lin, B. Mathisizik, B. Xu, S.-O. Enfors, P. Neubauer, Determination of the maximum specific uptake capacities for glucose and oxygen in glucose-limited fed-batch cultivations of *Escherichia coli*, *Biotechnol. Bioeng.* 73 (5) (2001) 347–357, <https://doi.org/10.1002/bit.1068>.
- [65] I.C. Rocha, Model-Based Strategies for Computer-Aided Operation of a Recombinant *E. Coli* Fermentation, Escola de Engenharia Universidade do Minho, Braga, 2003.
- [66] R. Callewaert, L. De Vuyst, Bacteriocin production with *Lactobacillus amylovorus* DCE 471 is improved and stabilized by fed-batch fermentation, *Appl. Environ. Microbiol.* 66 (2) (2000) 606–613, <https://doi.org/10.1128/AEM.66.2.606-613.2000>.
- [67] J. Kager, A. Tuveri, S. Ulonska, P. Kroll, C. Herwig, Experimental verification and comparison of model predictive, PID and model inversion control in a *Penicillium chrysogenum* fed-batch process, *Process Biochem* 90 (2020) 1–11, <https://doi.org/10.1016/j.procbio.2019.11.023>.
- [68] L.F. Shampine, S. Thompson, J.A. Kierzenka, G.D. Byrne, Non-negative solutions of ODEs, *Appl. Math. Comput.* 170 (1) (2005) 556–569, <https://doi.org/10.1016/j.amc.2004.12.011>.
- [69] T.M. Alsoudani, Discontinuities in Mathematical Modelling: Origin, Detection and Resolution, no. March. 2016.
- [70] T. Barz, M. Adnan Jouned, Julian Kager, Christoph Herwig, Event driven analysis to enhance model calibration of experiments with high offline sampling rates 463 (2021), <https://doi.org/10.1016/B978-0-323-88506-5.50073-5>.
- [71] B.R. Gibson, S.J. Lawrence, J.P.R. Leclair, C.D. Powell, K.A. Smart, Yeast responses to stresses associated with industrial brewery handling, *FEMS Microbiol. Rev.* 31 (2007) 535–569, <https://doi.org/10.1111/j.1574-6976.2007.00076.x>.
- [72] E. Anane, D.C. López, C. Neubauer P., M.N. Cruz Bournazou, Modelling overflow metabolism in *Escherichia coli* by acetate cycling, *Biochem. Eng. J.* 125 (2017) 23–30, [10.1016/J.BEJ.2017.05.013](https://doi.org/10.1016/J.BEJ.2017.05.013).
- [73] C. Canales, C. Altamirano, J. Berrios, The growth of *Pichia pastoris* Mut+ on methanol-glycerol mixtures fits to interactive dual-limited kinetics: model development and application to optimised fed-batch operation for heterologous protein production, *Bioprocess Biosyst. Eng.* 41 (12) (2018) 1827–1838, <https://doi.org/10.1007/s00449-018-2005-1>.
- [74] E. Çelik, P. Çalik, S.G. Oliver, A structured kinetic model for recombinant protein production by Mut+ strain of *Pichia pastoris*, *Chem. Eng. Sci.* 64 (23) (2009) 5028–5035, <https://doi.org/10.1016/j.ces.2009.08.009>.
- [75] H. Song, J. Jiang, X. Wang, J. Zhang, High purity recombinant human growth hormone (rhGH) expression in *Escherichia coli* under *phoA* promoter, *Bioengineered* 8 (2) (2017) 147–153, <https://doi.org/10.1080/21655979.2016.1212137>.
- [76] J. Ahn, J. Hong, M. Park, H. Lee, E. Lee, C. Kim, J. Lee, E.S. Choi, J.K. Jung, H. Lee, Phosphate-responsive promoter of a *Pichia pastoris* sodium phosphate symporter, *Appl. Environ. Microbiol.* 75 (11) (2009) 3528–3534, <https://doi.org/10.1128/AEM.02913-08>.
- [77] F.E. Cellier, Combined continuous/discrete system simulation by use of digital computers techniques and tools, 1979.

## 2.2.4 Robust modelling of *S. cerevisiae* fed-batch cultures by proper handling of model discontinuities

## Robust modelling of *S. cerevisiae* fed-batch cultures by proper handling of model discontinuities

M. Adnan JOUNED<sup>a,b</sup>, Julian KAGER<sup>c</sup>, Christoph HERWIG<sup>a,d</sup>, Tilman BARZ<sup>b</sup>

<sup>a</sup> ICEBE, TU Wien, Gumpendorfer Straße 1a 166/4, 1060 Vienna, Austria

<sup>b</sup> AIT Austrian Institute of Technology GmbH, Giefinggasse 2, 1210 Vienna, Austria

<sup>c</sup> Department of Chemical and Biochemical Engineering, Technical University of Denmark, Building 228A, 2800 Kgs. Lyngby, Denmark

<sup>d</sup> Competence Center CHASE GmbH, Altenbergerstraße 69, 4040 Linz, Austria

### ABSTRACT

Mechanistic ordinary differential equation (ODE) models are important tools for the characterization, monitoring and control of yeast cultivations. However, a simplified representation of the cell metabolism and process dynamics can cause mathematical discontinuities in the models which can have crucial implications on the model's output.

A solution to this is to formulate discontinuous (piecewise) growth models and to incorporate metabolic and process-related switches expressed as logical operations.

This contribution discusses the application of a piecewise kinetic growth model for targeted *Saccharomyces cerevisiae* lab scale cultivations that were conducted with different glucose and ethanol fluxes designed to trigger switches between metabolic pathways. The so-called Event Driven Method (EDM) is proposed to properly handle metabolic and process-related discontinuities, in addition to nonlinear kinetic model terms that exhibit discontinuous behavior.

The results show that compared with a standard implementation without active event location, the proposed approach leads to more accurate identification of the switches and model parameters and thus, less predictions uncertainties.

**Keywords:** event driven modelling, piecewise kinetic growth models, yeast cultivation, metabolic discontinuities, process-related discontinuities.

# 1 INTRODUCTION

Mechanistic models play an important role for effective experimental design [1], [2], real-time monitoring and predictive control [3]. These models represent the knowledge of the underlying physical characteristics of the process and the physiological behavior of the organisms using mathematical expressions and model parameters [4].

Mechanistic kinetic growth models use stoichiometric information, nonlinear reaction rates and mass and concentration balances [4], and are usually written as a set of deterministic and continuous Ordinary Differential Equations (*ODEs*). Unstructured models do not incorporate a detailed metabolic and physiological description of the organism. Internal reactions are often lumped together and represented as one overall metabolic pathway. In contrast, structured models like metabolic flux analysis provide a more detailed mathematical description of the intracellular metabolism but due to their complexity and the difficulty to measure cell internal reactions, their application in practice is still very limited [4], [5]. Therefore, in the context of industrial biotechnology, unstructured or simplified structured models are commonly used [4]. These models primarily focus on the description of simplified pathways and simplified biological transitions. However, in reality, transitions are mostly continuous, highly nonlinear, and dependent on intracellular mechanisms, but as dynamics happen in very different timescales, most of the transitions are simplified into discontinuous behaviors. Therefore, switches expressed as logical operations often need to be incorporated in the model [6]. The result is a (discontinuous) piecewise growth model. Examples of that are models that describe the Crabtree effects [7], or models that consider sudden external changes such as pulse feeding and sampling.

Discontinuous (piecewise) growth models can be mathematically expressed as a combination of a set of continuous differential (and algebraic) equations with discontinuous right-hand side, and a set of time-dependent and/or state-dependent conditions, also referred to as event functions. If a condition is fulfilled, an event is triggered, and the model is switched or updated. There are mainly two possible approaches to deal with ODEs with discontinuous right-hand sides [8]: the *time stepping method (TSM)* and the *event driven method (EDM)*.

The TSM simply ignores discontinuities and uses solvers for continuous initial value problems (IVPs). In discontinuous regions the solution might violate the crucial assumption of smoothness [8]. In contrast, the EDM locates discontinuities (defined as events [9]) using event functions. The solution is a result of a sequence of IVPs, described by differential equations and interspersed by instantaneous events that cause a discrete change to the initial value problem currently being solved.

This paper shows parameter estimation and model prediction results of a *Saccharomyces cerevisiae* cultivation model [10]. The model is fitted to data from targeted lab scale experiments where ethanol and glucose fluxes were manipulated by the substrate feed rate to trigger switches between different metabolic pathways. The paper highlights the consequences when not explicitly accounting for discontinuities during model implementation and its numerical solution. If discontinuities are not properly handled during model implementation, the solution will be influenced by noise which affects the accuracy of parameters estimation and increases the model prediction uncertainty.

## 2 METHODS

### 2.1 Experimental setup, growth model and simulation environment

The model of *Sonnleitner and Käppeli* [10] in which the authors describe how *Saccharomyces cerevisiae* grows using different metabolic pathways is used. Three reactions (metabolic pathways) can be distinguished: A- Glucose and ethanol oxidation (where sum of fluxes is below maximum oxidative capacity), B- Glucose and ethanol oxidation (where sum of fluxes is higher than maximum oxidative capacity), C- Crabtree effect (the glucose flux exceeds the maximum oxidative capacity). *MATLAB* “*ODE suite*” is used to solve the ODEs. Three *Saccharomyces cerevisiae* (wildtype, *CBS8340*) experiments were conducted in 3 L aerated and stirred glass-reactors. For online analytics, CO<sub>2</sub> and O<sub>2</sub> content in the off-gas was measured by a gas analyzer. For offline analytics, glucose, ethanol and biomass concentration were measured. Samples were taken both manually and by an autosampler at irregular basis and the sample volume (ranging from ~4-20 mL) was logged. More details on experimental setup, model structure, parameters and units can be found in [11].

### 2.2 Model parameters and predictions uncertainties quantification

The model is fitted to the experimental data considering the normalized residual sum of squares (NRSS) of the measured and predicted liquid and gas concentrations. For  $N_L$  liquid concentrations which were measured in  $M_L$ , and for  $N_G$  gas concentrations which were continuously monitored and evaluated at  $M_G$  time points, and with  $\theta$  as unknown parameter vector, the minimization problem reads:

$$\min_{\theta} \phi^{NRSS}(\theta) \quad \text{with}$$
$$\phi^{NRSS}(\theta) = \frac{1}{M_L} \sum_{i=1}^{N_L} \sum_{j=1}^{M_L} (Y_{i,j}(\theta) - Y_{i,j}^m)^2 + \frac{1}{M_G} \sum_{k=1}^{N_G} \sum_{l=1}^{M_G} (Y_{k,l}(\theta) - Y_{k,l}^m)^2$$

Out of whole model parameters, five parameters were selected as identifiable:  $q_s^{max}$ ,  $q_{O_2}^{max}$ ,  $Y_{x/s(red)}$ ,  $Y_{x/s(ox)}$  and  $Y_{x/e(ox)}$  based on the identifiability analysis described in [12].

The uncertainty of the parameters and model predictions was analyzed using bootstrapping [13], which involves resampling of the experimental data and re-estimation of the parameters. For each experiment 500 *Monte Carlo* (MC) datasets were generated based on the measurement's uncertainty, and perturbations at 10%, 30% and 50% of initial guesses of the selected parameters.

### 2.3 Growth model implementations

For the implementation of the growth model of *S. cerevisiae*, the following methods are used:

- time stepping method (TSM): uses the routines of the *MATLAB ODESUITE* package without event detection option.
- event driven method (EDM): uses the routines of the *MATLAB ODESUITE* package with an event detection option. During the solution, the conditions are monitored, and corresponding events are detected using *MATLAB*'s ODE event location algorithm. If

an event is detected, the integration is terminated, and the model is switched. The integration is then restarted with the new submodel and/or adapted initial conditions.

The following discontinuities are hereby handled in the EDM implementation:

A- *Metabolic discontinuities*: Switches that happen between the metabolic pathways A, B and C (see 2.1). These are actively located by monitoring the metabolic conditions (i.e., by solving the corresponding algebraic condition equations outside the ODE system).

B- *Process-related discontinuities*: In our experimental setup, the measured in/out-flow rates usually show comparatively smooth curves, except for sampling, where samples are taken by syringes in a very short time. Therefore, the volume changes due to sampling are modelled with events, where the reactor volume is updated (outside the ODE system) directly after each sample, see [14] for more details.

C- *Discontinuities caused by highly nonlinear kinetic terms (Non-physical solutions)*: Nonlinear kinetic models such as Monod-type growth models can exhibit stiff behaviors, especially when the affinity of the organism to the used substrates of glucose and ethanol is high [4], indicated by a small affinity constant ( $K$  roughly  $< 0.2$ ). For that, corresponding zero crossing conditions were considered.

Figure 1 shows in detail the computational schemes of both methods.

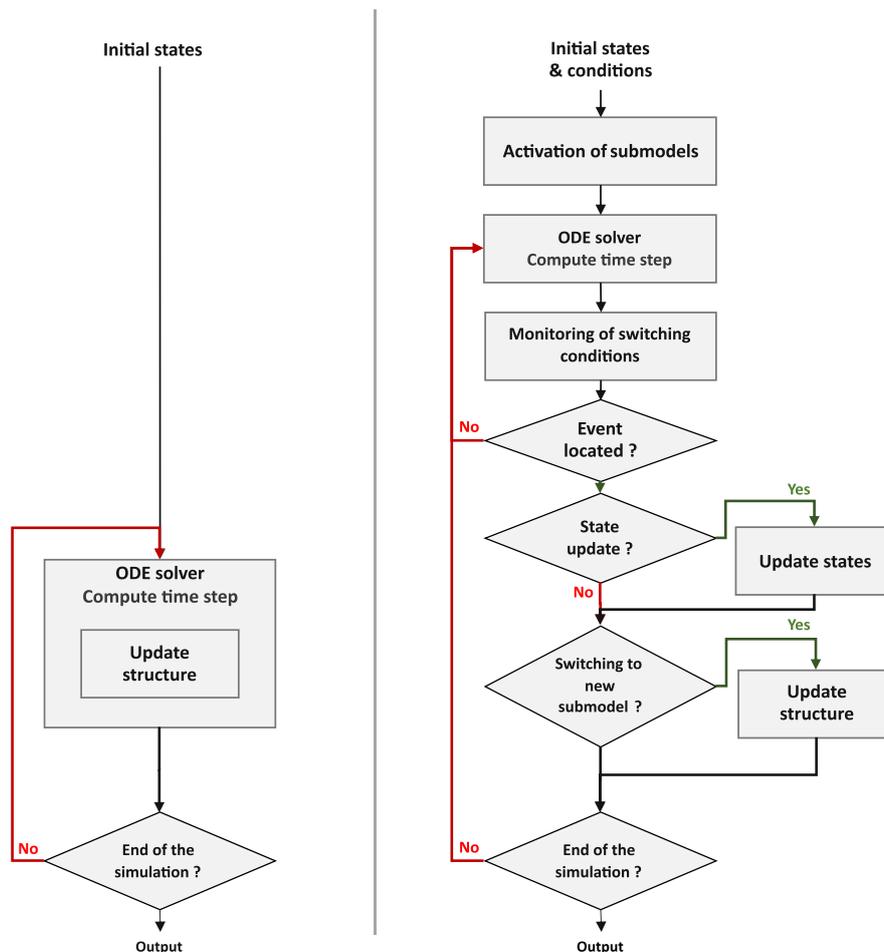


Figure 1: Computational schemes of the standard time stepping method (TSM) and the proposed event driven method (EDM). The EDM scheme accounts explicitly for any discontinuities in the model by monitoring conditions and switching to the corresponding submodels or updating state values, The figure is adapted from [11].

### 3 RESULTS

Using TSM, the events are not accurately located, therefore, the solution of the model is affected by calculation inaccuracies, leading to noisy outputs. This is a problem for the computation of sensitivities and gradient information during optimization iterations, because in many optimization methods, gradients are computed by evaluating the objective function in several points in the neighborhood of the current guess, using finite step sizes. Stochastic errors (noise) in the objective function values leads to errors in the computed gradients.

A consequence of the calculation inaccuracies during parameter identification (See 2.2) is illustrated in Figure 2. It shows the normalized residual sum of squares (NRSS) for models implemented by TSM and EDM and obtained from 500 data sets fittings at different perturbation levels of the initial parameters (10%, 30% and 50%). Overall EDM leads to lower NRSS values throughout all perturbation levels indicating a better convergence to the optimal parameters, while the NRSS obtained by TSM is higher, with a broader distribution and a visible dependence on the initial parameter perturbation levels.

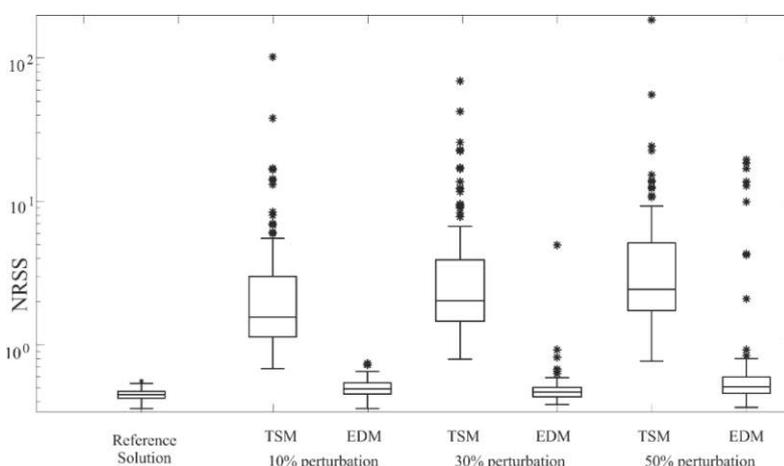


Figure 2: NRSS box plots obtained from fitting 500 simulated data sets (generated by resampling the data from all experiments) and starting with initial parameter values at different perturbation levels (10%,30% and 50%).

Table 1 shows the relative errors of the estimated parameter which were selected by the parameter identifiability analysis. The first column shows results for the reference solution. All columns show results obtained by 500 data sets fittings with different perturbation levels of the initial parameters. It can be noted that the results for the EDM are not significantly affected by the perturbation in the initial parameter guess. The reported parameter errors are similar to the errors obtained for the best reference solution. In contrast, for TSM, with higher perturbation levels, the errors in the parameter estimates increase significantly.

Figure 3 shows how EDM produces a continuous and differentiable optimization surface, whereas TSM surface is noisy, discontinuous and non-differentiable (top figures). Model simulations for some neighbored values of a selected parameter reveals how a gradual change of a parameter value can cause unsystematic deviations in TSM model predictions, while this is not observed for EDM (bottom figures). The unsystematic results of TSM are a consequence of the inaccurate events location.

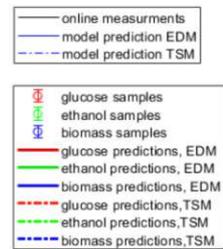
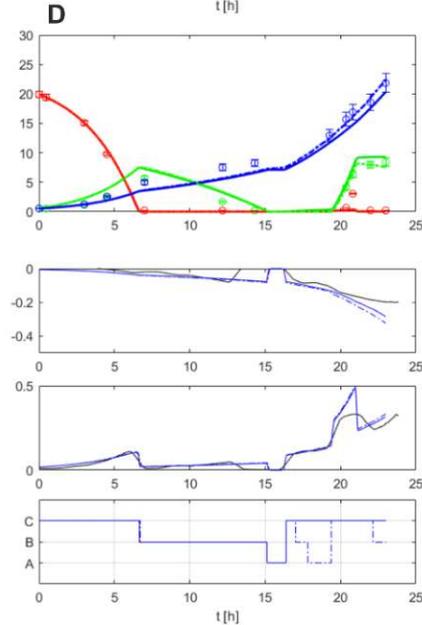
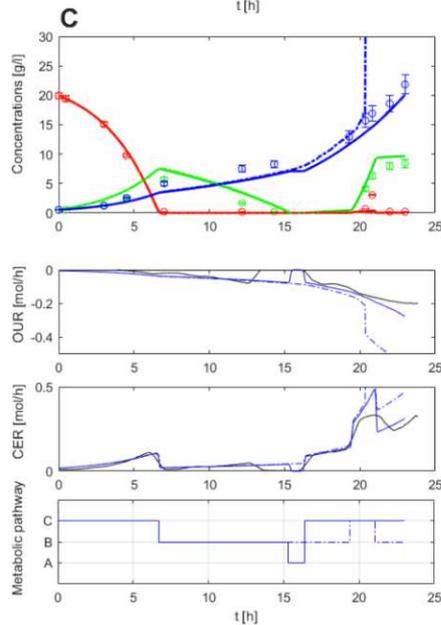
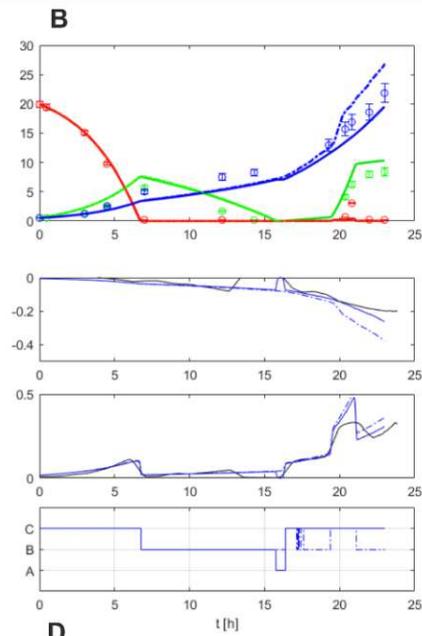
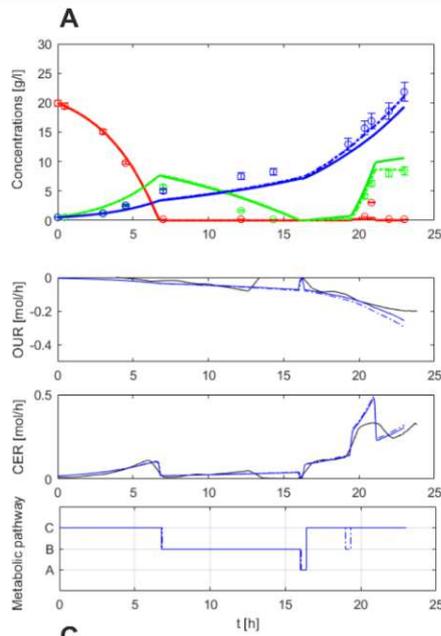
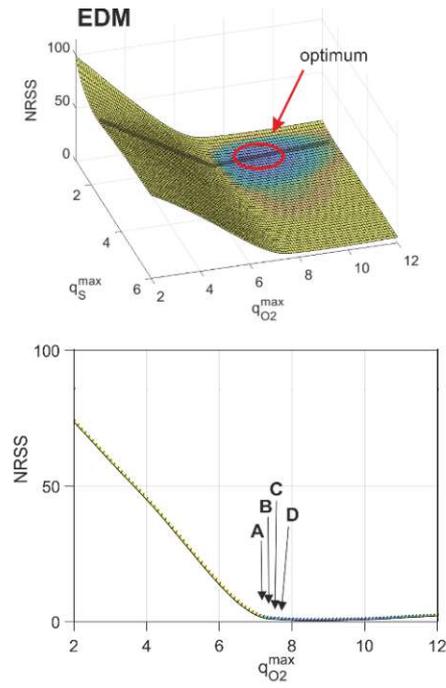
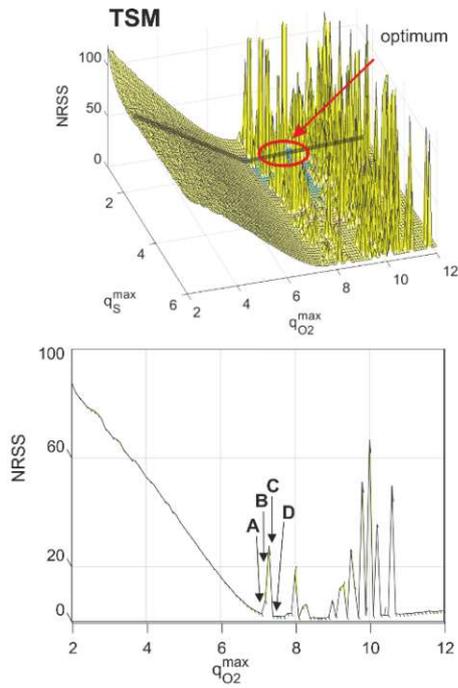


Figure 3: Top) NRSS optimization surfaces of a reduced two-dimensional parameter estimation problem for the model implementation following TSM and EDM for one cultivation. The surfaces are constructed by evaluations (repeated simulations) for a grid of parameters. The optimal solution is around  $q_{s(ox)}^{max} = 1.7 \text{ g} \cdot \text{h}^{-1} \cdot \text{g}^{-1}$  and  $q_{O_2}^{max} = 8.7 \text{ g} \cdot \text{h}^{-1} \cdot \text{g}^{-1}$ . The TSM produces stochastic errors which produce a noisy surface that impedes the optimization algorithm to converge to the minimum. Note that for TSM, the large peaks result mostly from “non-physical solutions”. Bottom) Simulation results at selected values of  $q_{O_2}^{max}$ : maximum oxygen uptake rate (A-  $q_{O_2}^{max} = 7.1$ , B-  $q_{O_2}^{max} = 7.3$ , C-  $q_{O_2}^{max} = 7.5$ , D-  $q_{O_2}^{max} = 7.6$ ), rest of the parameters are set to constants. Dashed lines refer to TSM results, and solid ones refer to EDM.

Table 1: Relative error (% $\sigma$ ) of the estimated parameters at different perturbation levels of the initial parameter guess. Low errors indicate a good parameter identification, whereas high errors indicate a poor identification.

	Perturbation of the initial parameter guess						
	0%	10 %		30 %		50 %	
	<b>Reference</b> <b>%<math>\sigma</math></b>	<b>TSM</b> <b>%<math>\sigma</math></b>	<b>EDM</b> <b>%<math>\sigma</math></b>	<b>TSM</b> <b>%<math>\sigma</math></b>	<b>EDM</b> <b>%<math>\sigma</math></b>	<b>TSM</b> <b>%<math>\sigma</math></b>	<b>EDM</b> <b>%<math>\sigma</math></b>
$q_s^{max}$	<b>3.52</b>	8.5	6.72	20.84	3.7	30.22	6.36
$q_{O_2}^{max}$	<b>2.94</b>	7.26	4.16	15.41	5.71	22.04	4.16
$Y_{x/s(red)}$	<b>25.74</b>	16.10	26.43	30.68	18.56	41.33	26.71
$Y_{x/s(ox)}$	<b>5.94</b>	12.53	8.55	28.54	12.27	38.99	8.5
$Y_{x/e(ox)}$	<b>3.62</b>	6.14	5.14	12.04	5.94	17.34	5.14

## 4 CONCLUSIONS

The comparative analysis of the proposed EDM method versus the commonly used TSM method for model implementation, underlines that (discontinues) mechanistic ODE models implemented with EDM deliver more accurate location of metabolic switches, lower prediction error and lower parameter uncertainty.

## 5 References

- [1] I. Gosling, “Process simulation and modeling for industrial bioprocessing: Tools and techniques,” *Ind. Biotechnol.*, vol. 1, no. 2, pp. 106–109, Jun. 2005, doi: 10.1089/ind.2005.1.106.
- [2] A. Toumi, C. Jürgens, C. Jungo, B. A. Maier, V. Papavasileiou, and D. Petrides, “Design and optimization of a large scale biopharmaceutical facility using process simulation and scheduling tools,” *Pharm. Eng.*, vol. 30, Mar. 2010.
- [3] D. Dochain, “State and parameter estimation in chemical and biochemical processes: A tutorial,” *J. Process Control*, vol. 13, no. 8, pp. 801–818, Dec. 2003, doi: 10.1016/S0959-1524(03)00026-X.
- [4] P. M. Doran, *Bioprocess engineering principles: Second edition*, vol. 9780080917. Academic Press, 2012.
- [5] M. L. Shuler and J. D. Varner, “Cell growth dynamics,” in *Comprehensive Biotechnology*, Pergamon, 2011, pp. 32–38. doi: 10.1016/B978-0-444-64046-8.00065-3.
- [6] F. Fröhlich, F. J. Theis, J. O. Rädler, and J. Hasenauer, “Parameter estimation for dynamical systems with discrete events and logical operations,” *Bioinformatics*, vol. 33, no. 7, pp. 1049–1056, Dec. 2017, doi: 10.1093/bioinformatics/btw764.
- [7] H. G. Crabtree, “Observations on the carbohydrate metabolism of tumours,” *Biochem. J.*, vol. 23, no. 3, pp. 536–545, Jan. 1929, doi: 10.1042/bj0230536.
- [8] L. Dieci and L. Lopez, “A survey of numerical methods for IVPs of ODEs with discontinuous right-hand side,” *J. Comput. Appl. Math.*, vol. 236, no. 16, pp. 3967–3991, 2012, doi: 10.1016/j.cam.2012.02.011.
- [9] T. Park and P. I. Barton, “State event location in differential-algebraic models,” *ACM Trans. Model. Comput. Simul.*, vol. 6, no. 2, pp. 137–165, 1996, doi: 10.1145/232807.232809.
- [10] B. Sonnleitner and O. Käppeli, “Growth of *Saccharomyces cerevisiae* is controlled by its limited respiratory capacity: Formulation and verification of a hypothesis,” *Biotechnol. Bioeng.*, vol. 28, no. 6, pp. 927–937, 1986, doi: 10.1002/bit.260280620.
- [11] M. A. Jouned, J. Kager, C. Herwig, and T. Barz, “Event driven modeling for the accurate identification of metabolic switches in fed-batch culture of *S. cerevisiae*,” *Biochem. Eng. J.*, vol. 180, p. 108345, Mar. 2022, doi: 10.1016/j.bej.2022.108345.
- [12] D. C. López, T. Barz, M. Peñuela, A. Villegas, S. Ochoa, and G. Wozny, “Model-based identifiable parameter determination applied to a simultaneous saccharification and fermentation process model for bio-ethanol production,” *Biotechnol. Prog.*, vol. 29, no. 4, pp. 1064–1082, 2013, doi: 10.1002/btpr.1753.
- [13] B. Efron and R. Tibshirani, *An introduction to the bootstrap*. CRC Press, 1994.
- [14] T. B. 463 M. Adnan JOUNED, Julian KAGER, Christoph HERWIG, “Event driven analysis to enhance model calibration of experiments with high offline sampling rates,” 2021, p. 463.

### 2.2.5 A unique response behavior in the dissolved oxygen tension signal of *E. coli* in minibioreactor system equipped with intermittent bolus feeding

Article

# A unique response behavior in the dissolved oxygen tension signal of *E. coli* in minibioreactor system equipped with intermittent bolus feeding

M. Adnan Jouned<sup>1</sup>, Julian Kager<sup>2</sup>, Vignesh Rajamanickam<sup>3</sup>, Christoph Herwig<sup>1</sup>, and Tilman Barz<sup>4,\*</sup>

<sup>1</sup> ICEBE, TU Wien, Gumpendorfer Straße 1a 166/4, 1060 Vienna, Austria.

<sup>2</sup> Department of Chemical and Biochemical Engineering, Technical University of Denmark, Building 228A, 2800 Kgs. Lyngby, Denmark.

<sup>3</sup> Boehringer Ingelheim RCV GmbH & Co KG, Biopharmaceuticals Austria, Dr. Boehringer Gasse 5-11, A-1121 Vienna, Austria.

<sup>4</sup> Center for Energy, AIT Austrian Institute of Technology GmbH, Giefinggasse 2, 1210 Vienna, Austria.

\* Correspondence: Tilman.barz@ait.ac.at

**Abstract:** Intermittent bolus feeding strategies for *E. coli* cultivations in minibioreactor systems (MBRs) are known to have a profound effect on cell metabolism. Pulsed feeding results in temporal substrate surplus and transient oxygen limitation which leads to the production of inhibitory by-products. For each substrate pulse, the dissolved oxygen tension (DOT) signal exhibits a negative pulse that consists of two segments: a consistent decline and then an increase to the saturation value. A unique response behavior in some DOT pulses in *E. coli* cultivations in automated milliliter bioreactor systems equipped with intermittent bolus feeding is observed in this contribution. These pulses don't only show the previously mentioned two segments, rather they show four segments: first a partial decline followed by a small increase or a flattened curve and then another decline with a different slope, and after that, a return to the saturation value. This response seemed to appear at a dilution ratio higher than a certain threshold, and to become more pronounced with higher dilution ratio. This contribution attempts to provide a systematic analysis of the observed phenomenon. Our hypothesis is that the pattern of responses in particular DOT segments is linked to specific metabolic states. The analysis highlights a plausible relationship between a metabolic adaptation behavior and newly observed, and not reported before in literature, DOT signal segment. The quantitative analysis, and mechanistic model simulations support this hypothesis and show the possibility to obtain key growth parameters from the DOT signal. For our analysis, the estimation of model parameters involved in the overflow switching condition was possible using only DOT signal and biomass samples.

**Keywords:** dissolved oxygen tension; *E. coli* cultivation; mechanistic model; data-driven analysis; signal analysis;

**Citation:** To be added by editorial staff during production.

Academic Editor: Firstname Last-name

Received: date

Revised: date

Accepted: date

Published: date

**Copyright:** © 2023 by the authors.

Submitted for possible open access

publication under the terms and

conditions of the Creative Commons

Attribution (CC BY) license

<https://creativecommons.org/licenses/by/4.0/>.

<https://creativecommons.org/licenses/by/4.0/>.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.  
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

TU  
WIEN  
BIBLIOTHEK  
Your knowledge hub



## 1. Introduction

The development cycle of biopharmaceutical processes goes through three stages: in the first stage, the screening and characterization of the organisms take place. Second, the reaction conditions (e.g., medium and process variables) are optimized. In the last stage, the scale up to pilot and production scale takes place [1]. During these phases, a vast number of development cultivations is usually required [2].

High throughput technology (HTP) is widely used today to accelerate process development [3], [4]. HTP relies on real time monitoring and control, full automation of the systems and reduction of culture volume to achieve that [3].

Many HTP platforms with miniaturized bioreactors have been commercialized in the recent decade, [3] provided a nice review on that. The miniaturized bioreactors can be categorized into [5]: I) sub milliliter category: usually called microbioreactors [5], [6], and II) 1–10 milliliter category: usually called minibioreactors [5], [7]. Bioreactors with a volume of more than 10 ml; usually in the range of 10–100 ml are called small-scale bioreactors [5].

Miniaturized stirred bioreactor systems (MSBRs), sometimes simply called (MBRs) [8]–[10], are closely designed after the conventional stirred lab-scale bioreactor systems, and have been developed as an alternative to shake flasks, and microtiter plates (MTPs)[2], [11].

Bioprocesses development for MBR cultivations have been introduced for different organisms such *E. coli*, *S. cerevisiae* and *Bacillus subtilis* [1], [5], [8], [11]

*E. coli* bacterial cell is a suitable candidate for miniaturized systems due to its low susceptibility to shear damage. This allows for higher agitation rates [2].

*E. coli* cultivations in MBRs usually rely on intermittent bolus feeding strategies with relatively high frequencies and make use of the automatic pipetting robotic systems to deliver feeding pulses and to obtain at-line and off-line samples.

Manual off-line samples are usually very limited or not possible [2] as this might endanger the stability of the cultivation. The reactors have stirring and gassing elements. Stirring is achievable with different techniques such as a mechanical impeller [12] or an inductive system [8]. The temperature is usually regulated by a thermostat and a water-cooling system. Dissolved oxygen and pH are monitored almost in real-time using fluorometric optical sensors.

Feeding strategy is known to have a strong effect on protein expression in *E. coli* [1], [13]. A completely continuous feeding strategy in MBRs is hard to realize [1].

On the other hand, intermittent feeding results in drastic changes in the substrate concentration in the medium before, during, and after each feeding pulse. [14] proposed a concept of “hunger” and “starvation” states, and [15] recently proposed a “feast-famine” concept, both depending on substrate availability. Both contributions reported on the changes in the physiological and metabolic responses. [15] showed that the changes happen at the short and long-time scale.

Generally, If the substrate pulse injection time is very short, the cell consumes glucose oxidatively at the maximum oxidative capacity, and the excess sugar is metabolized into acetate in the overflow regime [15], [16]. The accumulated acetate can be later consumed oxidatively by the cell.

Hence, the intermittent feeding leads to frequent changes in the metabolic states [1], and to transnational oxygen limitation. This can have negative effects on cell physiology and growth [1], [14].

The most important process condition for aerobic cultivations in MBRs is the oxygen supply [17]. The insufficient oxygen transfer to the liquid phase to satisfy cell oxygen needs is a known issue for MBRs [5], [18]–[20]. The availability of oxygen in the medium greatly affects the performance of the cells, leading to drastic changes in the cultivation kinetics [21]. For intermittent feeding strategies, each feeding pulse triggers a temporary disturbance in the dissolved oxygen in the medium. This can be a challenging task for the reactor oxygen controlling system.

Further, the low number of possible samples in MBRs limits the chances of effective monitoring and control.

*E. coli* MBRs cultivation dynamics happen on different timescales [15]. Dynamics such as the metabolic changes after a substrate pulse occur in a short timescale range, possibly seconds timescale, while dynamics such as cell growth happens on a longer timescale, usually hours timescale. Modelling and analysis methods should cover both scales to avoid overlooking possibly important dynamics.

Dissolved oxygen tension (DOT) is a commonly obtained online signal in *E. coli* cultivations. It contains important information on cell metabolism. For example, [22], [23] showed the possibility to control the inhibitory by-products production in *E. coli* by avoiding the anaerobic metabolism using information derived from DOT sensors.

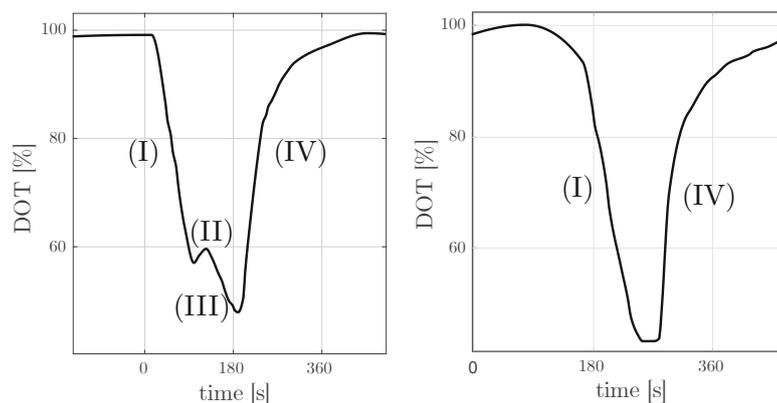
However, the encoded metabolic activities in the DOT signal are ambiguous, and the signal has high and low frequency details, hence, the separation of the useful characteristics from the background noise can be difficult. Also, a combination of sensor time delay and high substrate affinity of *E. coli* shifts the metabolic activities nonlinearly along the DOT signal [10].

DOT signal is influenced by two opposing components [13]: cell oxygen demand defined by oxygen uptake rate (OUR) and the oxygen delivery to the medium by reactor aeration and stirring systems defined by the oxygen transfer rate (OTR). If the stirring and aeration parameters are set constant and no control over the dissolved oxygen level in the medium is applied, the OTR component will have a constant trend, and the metabolic activities described by OUR are clearly revealed.

Many contributions reported on the response behavior of the DOT signal, and the possible relationship between metabolic activities of *E. coli* and certain segments of the DOT signal [16], [24]–[28]. *E. coli* switches almost immediately to the overflow metabolism after adding a substrate to the medium, this is usually associated with drastic decrease in the DOT value. [16] reported on the difference in the slopes of the DOT signal with different substrate types. Cells have different oxygen demand for glucose and acetate. [16] showed no difference if the acetate is introduced from outside the reactor or if it is produced locally by the overflow metabolism. The return of the DOT signal to the saturation value starts after the end of the metabolic activities. The increase of DOT values happens under the influence of only the OTR component, since the OUR component is zero.

In this contribution, an additional DOT response behavior is observed. The response is linked to a segment that appears after the end of the segment that exists after substrate addition. The newly-observed segment remains for a short time. It is recognized by a significant change in the DOT slope, and it is especially noticed when a large volume of substrate is added.

Figure 1 shows two examples of DOT pulses in an actual *E. coli* cultivation in a mini-bioreactor system, in which the following DOT segments can be distinguished: (I) A response occurs after substrate addition and is associated with a direct decline. (II) A response occurs after the end of the segment (I) and characterized by a small increase, a flattened curve, or very slow decrease. (III) A response aligned with a decline but with a different slope to the segment (I). (IV) A response aligned with a return to the saturation value.



**Figure 1.** Examples of two DOT pulses from real experimental data. The pulses show different response behavior. Left) DOT pulse with two segments, Right) DOT pulse with 4 segments.

This contribution presents a systematic experimental study of the DOT signal response to the intermittent bolus feeding in MBRs for *E. coli* cultivations. The feeding plans are designed to have a systematic variation in the pulses' frequency and amplitude.

The novelty of the paper lies in the detailed analysis of the DOT signal. The hypotheses inferred from the analysis were checked through: I) a quantitative analysis of growth model parameters and a comparison to literature values, and II) mechanistic model simulations.

Interestingly, that promotes a hypothesis on the relationship between a metabolic adaptation behavior and the newly-observed DOT segment, not reported before in literature.

The proposed analysis and modelling approaches provide means to better understand the intermittent bolus feeding effect on *E. coli* cultivations in MBRs and help to address oxygen supply issues.

The paper is arranged as the following: section 2 demonstrates the experimental setup.

Section 3 shows the experimental results and the DOT segmentation and analysis results. The inferred hypotheses, quantitative analysis and mechanistic modelling results are presented in section 4. Discussions on the analysis, hypotheses, modelling results, industrial relevance and future outlook are present in section 5. Finally, the conclusions are summarized in section 6.

## 2. Materials and Methods

### 2.1. Minibioreactor system and media

A block of eight bioreactors (bioREACTOR8; 2mag AG) equipped with pH and dissolved oxygen (DO) sensors (Mini-Bioreactors HTBD LG1-PSt3-Hg; PreSens GmbH, Regensburg, Germany) and fluorescence readers (MCR-LG1-v2; PreSens GmbH) were used for the *E. coli* cultivations. Temperature control and head-space cooling of the bioreactor blocks achieved by VersaCool™ Refrigerated Circulating Bath (Thermo Fisher Scientific GmbH, Schwerte, Germany). Gassing and mixing of the culture vessels is provided by the gas inducing, and inductive stirring elements.

For optical density measurements, an at-line microplate spectrophotometer (SPECTRAMax PLUS384; Molecular Devices Corporation, San Jose, USA) was used. For the off-line samples, a robotic arm (Robotic Manipulator Arm (RoMa), Tecan Trading AG) is used to transfer the samples to a deep freezer storage unit used to store the samples before the HPLC analysis. Glucose and ethanol concentrations of the filtered supernatant were

analyzed by HPLC (*Thermo Fischer, USA*) with a Supelco gel C-610 H ion exchange column (*Sigma-Aldrich, USA*) and a refractive index detector (*Thermo Fischer, USA*). The mobile phase was 0.1 %  $\text{H}_3\text{PO}_4$  with a constant flow rate of 0.5 ml/min at a temperature of 4 °C. The average sample volume was 300  $[\mu\text{L}]$ . The headspace of the bioreactors blocks was cooled to 4 °C to reduce the evaporation.

Experimental runs were conducted with *E. coli BL21* strain, a strain that can normally oxidize glucose in the overflow metabolism.

## 2.2. Experimental design

To calibrate the lower limit of the DO sensors at 0 [%], all reactors were gassed with 250  $[\text{mLh}^{-1}]$  nitrogen for 20 min and the stirring speed was set to 2800 [rpm]. To calibrate the upper limit of the DO sensors at 100 [%], all reactors were gassed with 250  $[\text{mLh}^{-1}]$  air for 20 min.

After calibration, the stirring speed is set to 1900 [rpm], and the gassing was set to 62.5  $[\text{mLh}^{-1}]$  air for the batch phase, and then the stirrer speed increased to 2800 [rpm] when glucose pulses are added. The pH is controlled at 6.8. The dissolved oxygen was intentionally uncontrolled.

Each experimental run initiated with 8 [ml] of medium and 5.7  $[\text{mg.mL}^{-1}]$  of dry cell weight. The batch phase lasted for almost 13 hours. After that, the fed-batch started with high frequent glucose pulses. The concentration of the fed glucose was 600  $[\text{mg.mL}^{-1}]$ . For two hours at the beginning of the fed-batch a ramp in the pH from 6.8 to 7.2 was considered to facilitate the induction of the culture. After these two hours the culture is induced with IPTG (76  $[\mu\text{L}]$ ). This time window is neglected in the contribution to avoid the dependency of cell metabolism on pH changes. The analysis time window of interest starts after 15 hours and ends by the end of the cultivation during which all process variables except feeding are kept constant.

The increasing feeding plan for all reactors is shown in Table 1. The feeding plan is designed so that injected substrate volumes and frequency results in full consumption of the substrate in between pulses, so no substrate accumulation in between pulses can be assumed. The reference feeding plan [8] meant to deliver a 78  $[\mu\text{L.h}^{-1}]$  of glucose on average to the medium. This is in principle supposed to be achieved by glucose pulses with fixed amplitude and frequency. However, as the pulses are entirely controlled by the reactor system software, the amplitude and the interval of the pulses can deviate due to conflicts with other internally scheduled tasks like sampling or pH titration. More on the experimental setup and reference feeding plan can be found in [8]. A picture of the system and setup is shown in Figure 2.

**Table 1.** Intermittent bolus feeding plan for each experimental run. The reference feeding plan meant to deliver a 78  $[\mu\text{L.h}^{-1}]$  of glucose on average to the medium. For all runs, the feeding concentration was 600  $[\text{mg.mL}^{-1}]$ .

Reactor Nr.	Total feeding volume compared to the reference plan [%]	Individual Feeding pulse volume $[\mu\text{L}]$	Feeding pulses time interval $[\text{min}]$	Average dilution ratio per feeding pulse $[\mu\text{L}/\text{mL}]$
A	25%	5	30	0.625
B	50%	5.5	20	0.688
C	75%	6.5	12	0.813
D	100%	6.5	9	0.813
E	100%	6.5	9	0.813
F	125%	8.5	9	1.063
G	150%	6	4	0.750
H	175%	7.5	4	0.938



**Figure 2.** The used mini-bioreactor system (bioREACTOR8; 2mag AG), and the experimental setup (Boehringer Ingelheim RCV GmbH & Co KG).

### 2.3. Computing platform

The analysis computations were carried out in *MATLAB* R2022a. “*ODE suite*”, mainly *ODE15s* solver, is used to solve the mechanistic model.

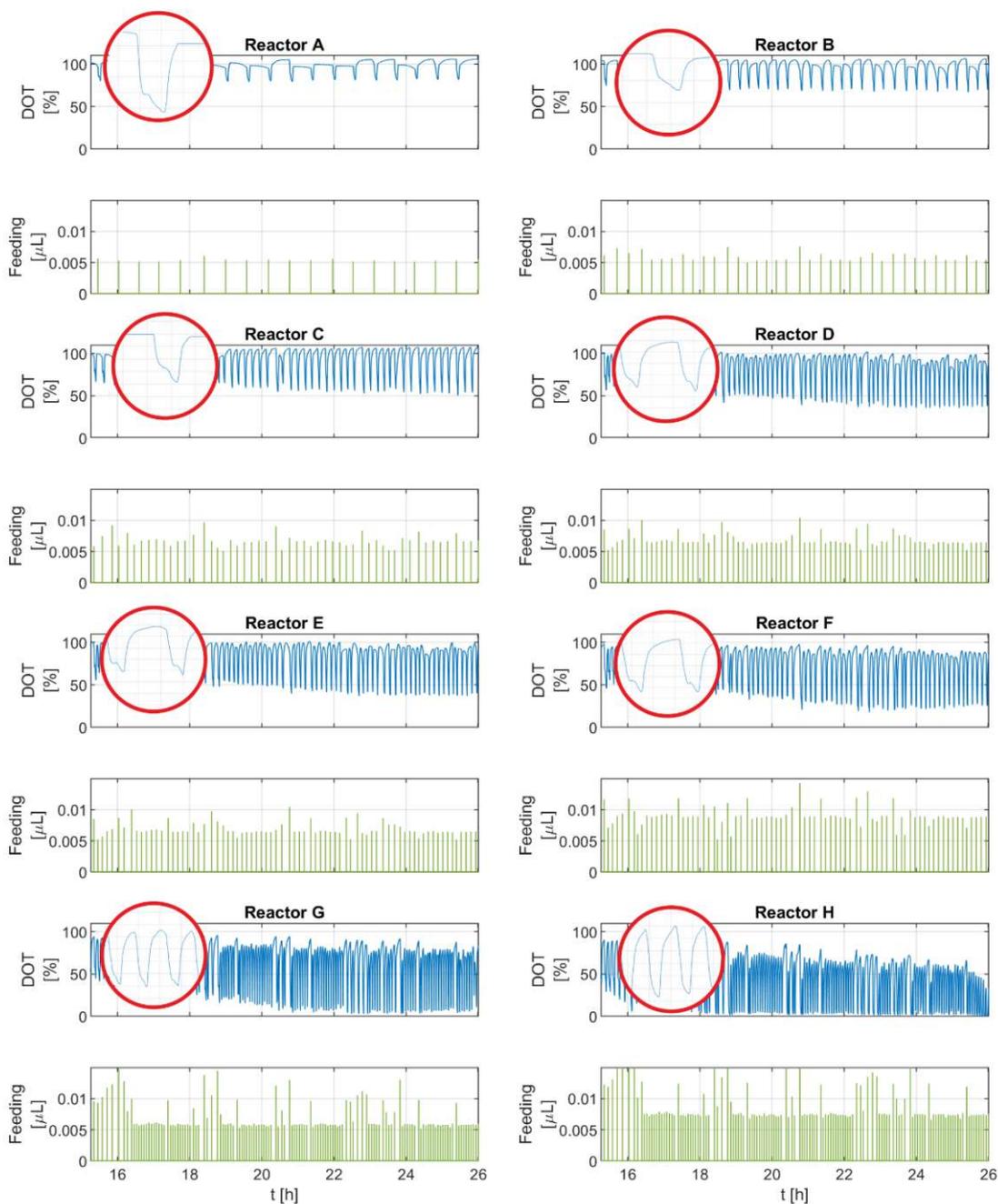
## 3. Experimental results and signal analysis

### 3.1. Experimental runs results

Figure 3 shows the DOT signal and feeding pulses for the eight experimental runs. The DOT pulses seem consistent for all runs. For each feeding pulse a DOT pulse is produced. Despite some deviations, in general the glucose feeding pulses seem to be equidistant with similar amplitudes for each individual run.

The visual inspection reveals that DOT pulses frequency increases with the increasing feeding pulses frequency. The amplitude of the DOT signal also seems to be correlated with the amount of the glucose added at each pulse. For runs (D, E,..H), the upper and lower boundaries of the DOT signal drift downward towards the end of the runs. This trend became more pronounced in the experimental runs with the largest feeding volumes, mostly runs G and H.

A closer look on the individual DOT pulses reveals that some pulses have a different profile to the others. These pulses don't only show two segments: a straight decline and then an increase to the saturation value (as commonly described in literature), rather they show four segments: first a partial decline followed by a small increase, flattened curve or very slow decrease and then another decline, after that, a return to the saturation value follows. Both types of pulses are observed in all runs.



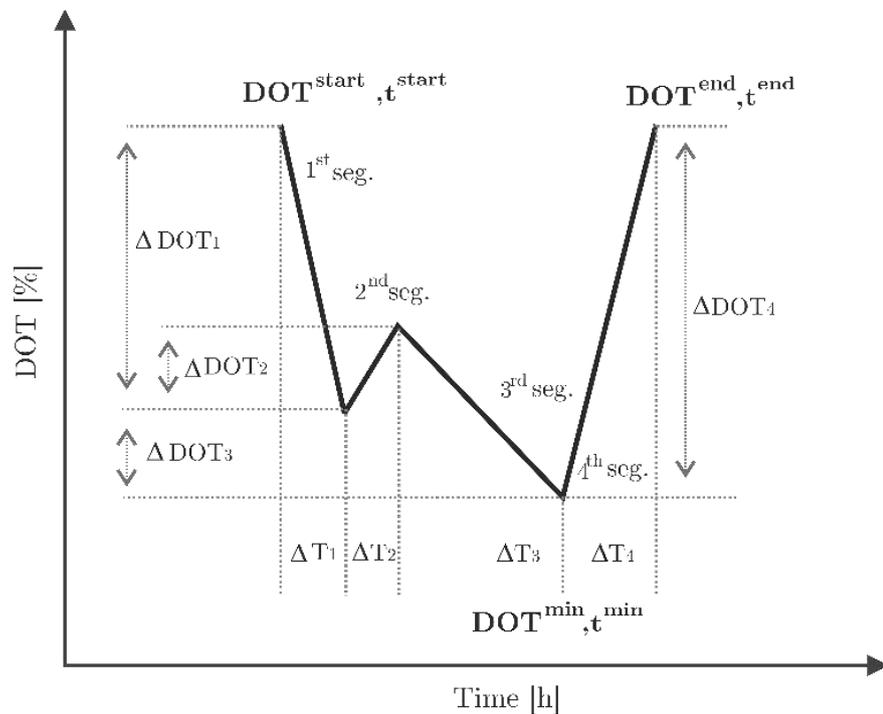
**Figure 3.** DOT signals and intermittent feeding pulses of all experimental runs.

The biomass, glucose and acetate samples values are shown in appendix B - Figure 12. The volume changes are shown in appendix B - Figure 13. Biomass samples indicate an increasing biomass growth with more feeding. On the other side, the measured acetate and glucose concentrations are very low and barely measurable. Generally, the acetate and glucose concentrations are always less than 1 [mg/ml] for the whole time of the cultivations. The limited reactors working volumes prevented acquiring more samples.

### 3.2. DOT signal analysis

Each DOT pulse is assumed to have possibly but not necessarily four segments, each segment represents a unique response behavior similar to the responses shown in Figure 1. This assumption is made based on the literature observations (see the introduction section) and an additional response observed by the authors. For the sake of the analysis, a segmentation algorithm (described in detail in appendix 7.1) is built to detect the segments.

Segment metrics (descriptive features) are extracted for each segment. The possible relationships between the metrics on one hand and process analytics of: feeding volume, reactor volume, dilution ratio, glucose concentration, acetate concentration and biomass concentration on the other hand are inspected visually and analyzed by regression analysis to detect any interesting patterns. Only data with a clear and a meaningful pattern are shown and discussed in the results and discussions sections Figure 4 shows an illustration of the possible segments of a DOT pulse and the related metrics.



**Figure 4.** DOT signal segments and relevant extractable metrics. Each DOT pulse is assumed to have possibly but not necessarily four segments.

The following metrics are defined for each segment  $i$ :

A. Segment time length: is defined as

$$\Delta T_i = t_i^{end} - t_i^{start} \quad (1)$$

B. Segment slope: is defined as

$$\frac{\Delta DOT_i}{\Delta T_i} = \frac{DOT_i^{end} - DOT_i^{start}}{t_i^{end} - t_i^{start}} \quad (2)$$

C. Segment area: the area of DOT curve

$$\int_{t_{start}}^{t_{end}} DOT_i \cdot dt \quad (3)$$

Similarly, the area of the OTR is calculated, once the  $K_{La}$  value is known, as  $\int_{t_{start}}^{t_{end}} OTR_i \cdot dt$ .

Table 2 shows the metrics of the segments. (X) refers to the presented results. Rest of the metrics did not result in meaningful patterns for the current data.

Table 2. Signal analysis metrics of each DOT pulse, (X) refers to the presented combinations.

#	Metric / Segment	Segments				
		1st segment	2nd segment	3rd segment	4th segment	All segments
A	Time length		x	x		
B	Slope	x	x	x	x	
C	Area					x

### Segmentation results

The segmentation results are shown exemplarily for one experimental run (reactor E) in Figure 5. The top subfigure shows the DOT raw (interpolated) signal. The subplots below show the segmentation results for each individual pulse. It can be noticed that pulses with 4 segments are generally aligned with feeding pulses with high amplitude.

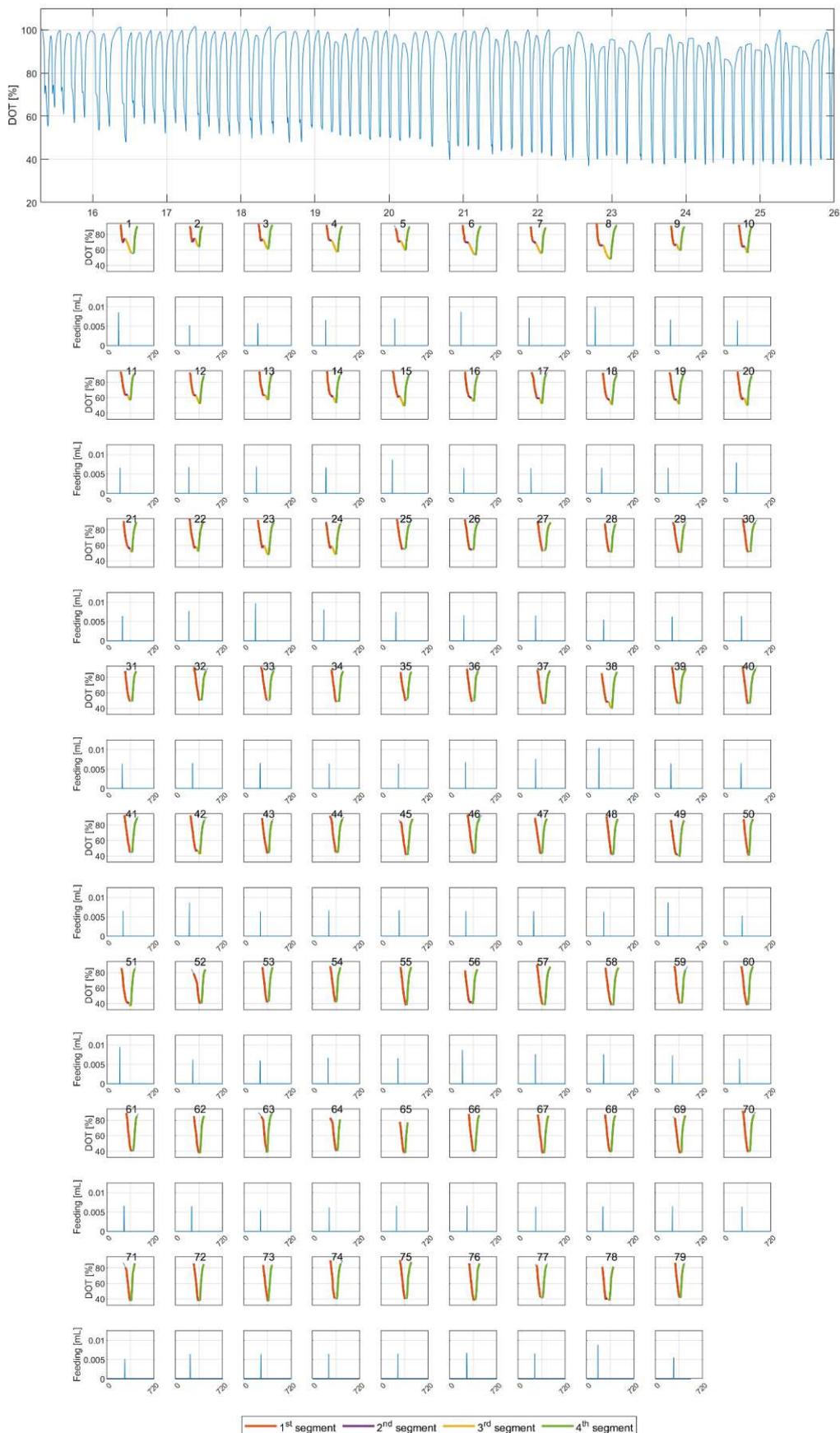


Figure 5. Top) measured DOT signal for reactor E. Bottom) the results of the segmentation algorithm. X axis represents the pulse time in seconds [s].

### Metric A: segment's time length

Figure 6 shows the time length of the 2<sup>nd</sup> segment for all experimental runs along the time course of the cultivations. The time length seems to be high, for all runs, at the beginning, and lower towards the end of the cultivation. The 2<sup>nd</sup> segment doesn't appear for all glucose pulses. The mean of the reported values is 31 seconds, with a range of 6.5-55.5 seconds for  $\pm 2$  standard deviations.

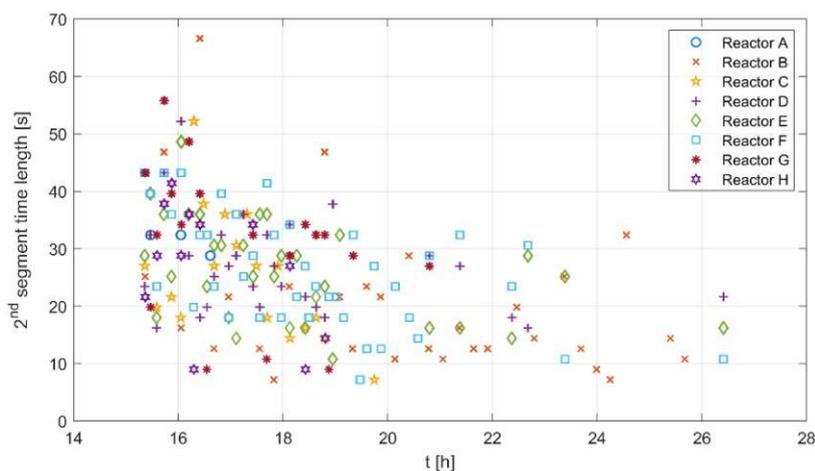


Figure 6. 2<sup>nd</sup> segment time length for all experimental runs along the cultivations time course.

Figure 7 shows the 3<sup>rd</sup> segment time length against the dilution ratio, biomass concentration and dilution per biomass unit. The points in the middle and right subplots are calculated in the neighborhood of biomass samples where the biomass concentration change can be neglectable.

The figure shows almost no correlation between 3<sup>rd</sup> segment time length and dilution ratio, and a weak relationship with biomass. This is because the effect of biomass and dilution ratio are alternately overlooked from many points in both figures. For example, a high dilution ratio with a large biomass concentration results in the same time length of a low dilution ratio with a low biomass concentration. However, there seems to be a correlation with the dilution ratio per biomass unit.

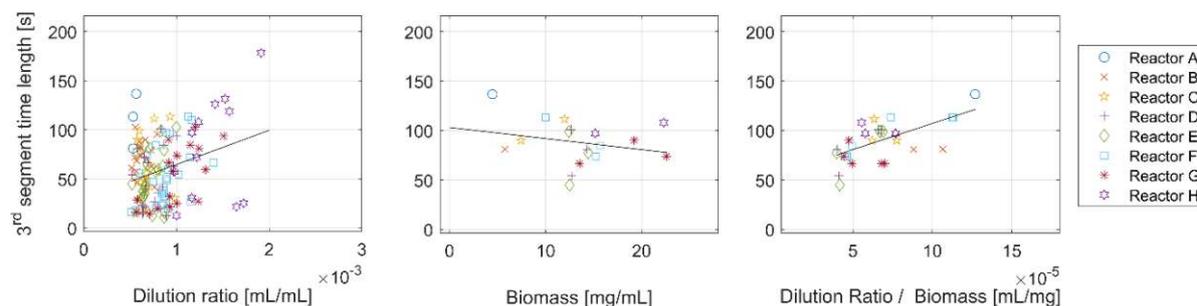
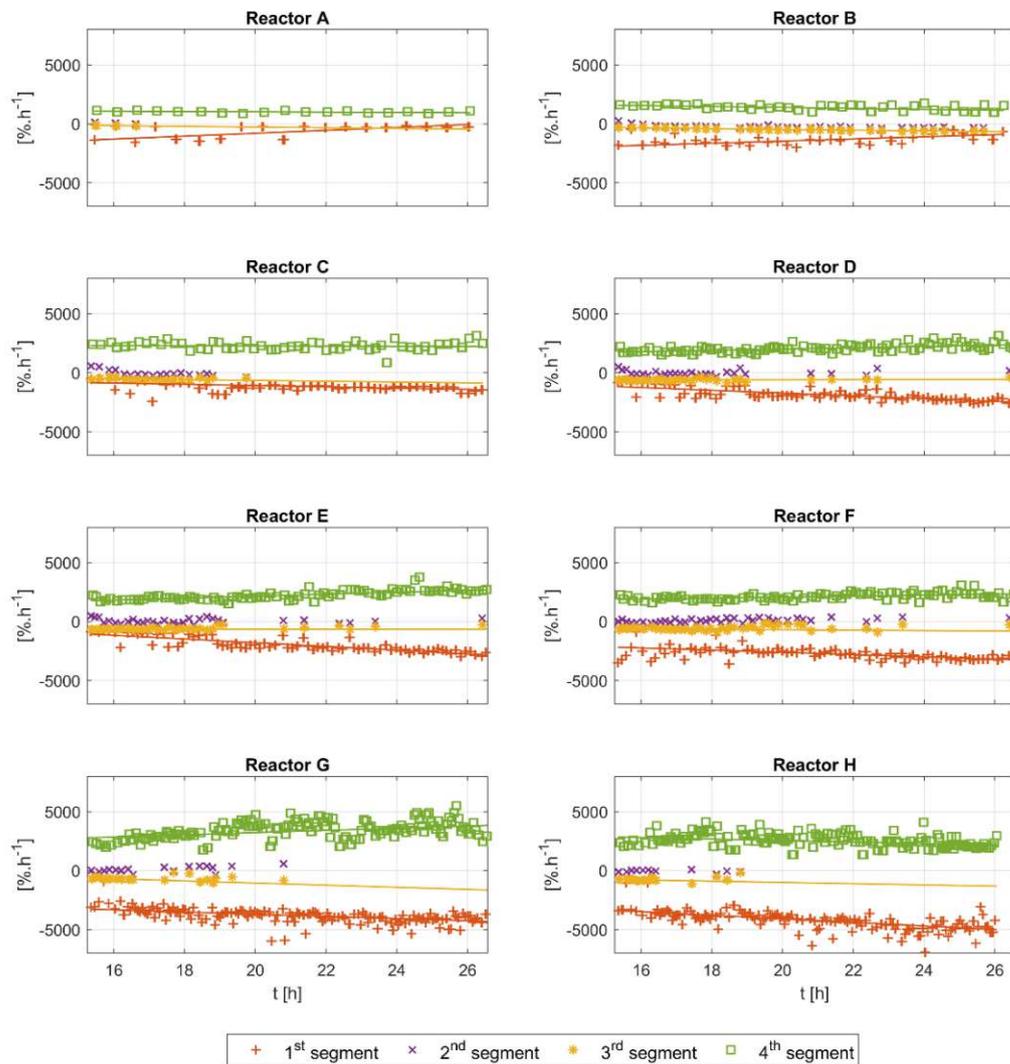


Figure 7. Time length of the 3<sup>rd</sup> segment against dilution ratio, biomass concentration and dilution ratio per biomass concentration.

### Metrics B: slope of the segments

The analysis of the slopes of the detected segments is shown in Figure 8. All segments' slopes appear to have a relatively similar and constant trend in all runs. The slopes of the 1<sup>st</sup> and 3<sup>rd</sup> segments show negative values with a visible difference between them. The slope of the 4<sup>th</sup> segment is always positive. For the 2<sup>nd</sup> segment, positive values are detected at the beginning of the runs, then the values become lower, closer to zero or

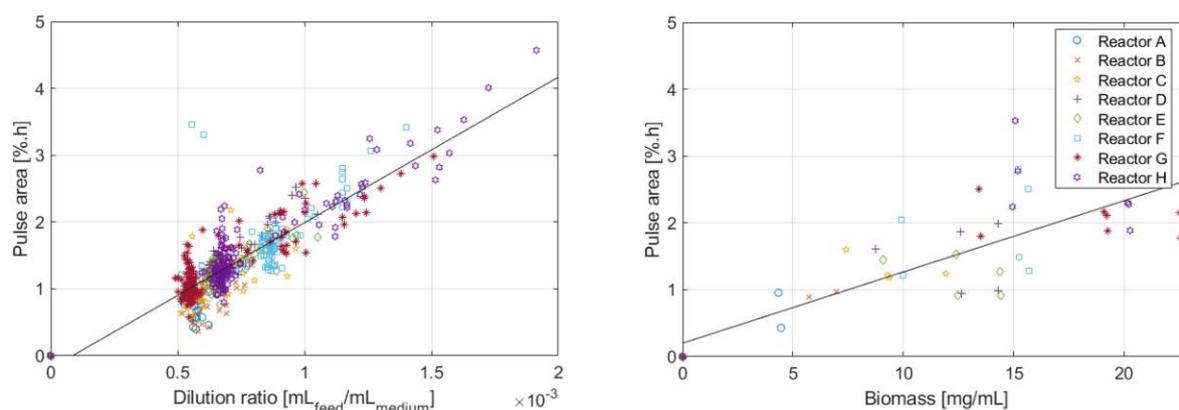
slightly negative. Again, the 2<sup>nd</sup> and 3<sup>rd</sup> segments don't appear for all glucose pulses. The slopes of 1<sup>st</sup>, 3<sup>rd</sup> and 4<sup>th</sup> segments drift slightly towards the end of the cultivations.



**Figure 8.** Slope analysis of the 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>, and 4<sup>th</sup> segments of all experimental runs. The slopes of the segments tend to be relatively constant along the time course of each cultivation. The 2<sup>nd</sup> and 3<sup>rd</sup> segments don't appear for each DOT pulse.

### Metric C: The area of the pulse

Figure 9 shows the results of the analysis of the DOT individual pulses area plotted against the biomass concentration and the dilution ratio. Both figures suggest a possible correlation between the area of the DOT pulse with both the dilution ratio and the biomass concentration. The results on the right figure are plotted only in the neighborhood of biomass samples, where the biomass concentration change can be negotiable.



**Figure 9.** The analysis of the individual DOT pulses areas plotted against the dilution ratio (left) and biomass concentration (right).

#### 4. Mechanistic modelling and model-based analysis

##### 4.1. Model-based analysis of the segments

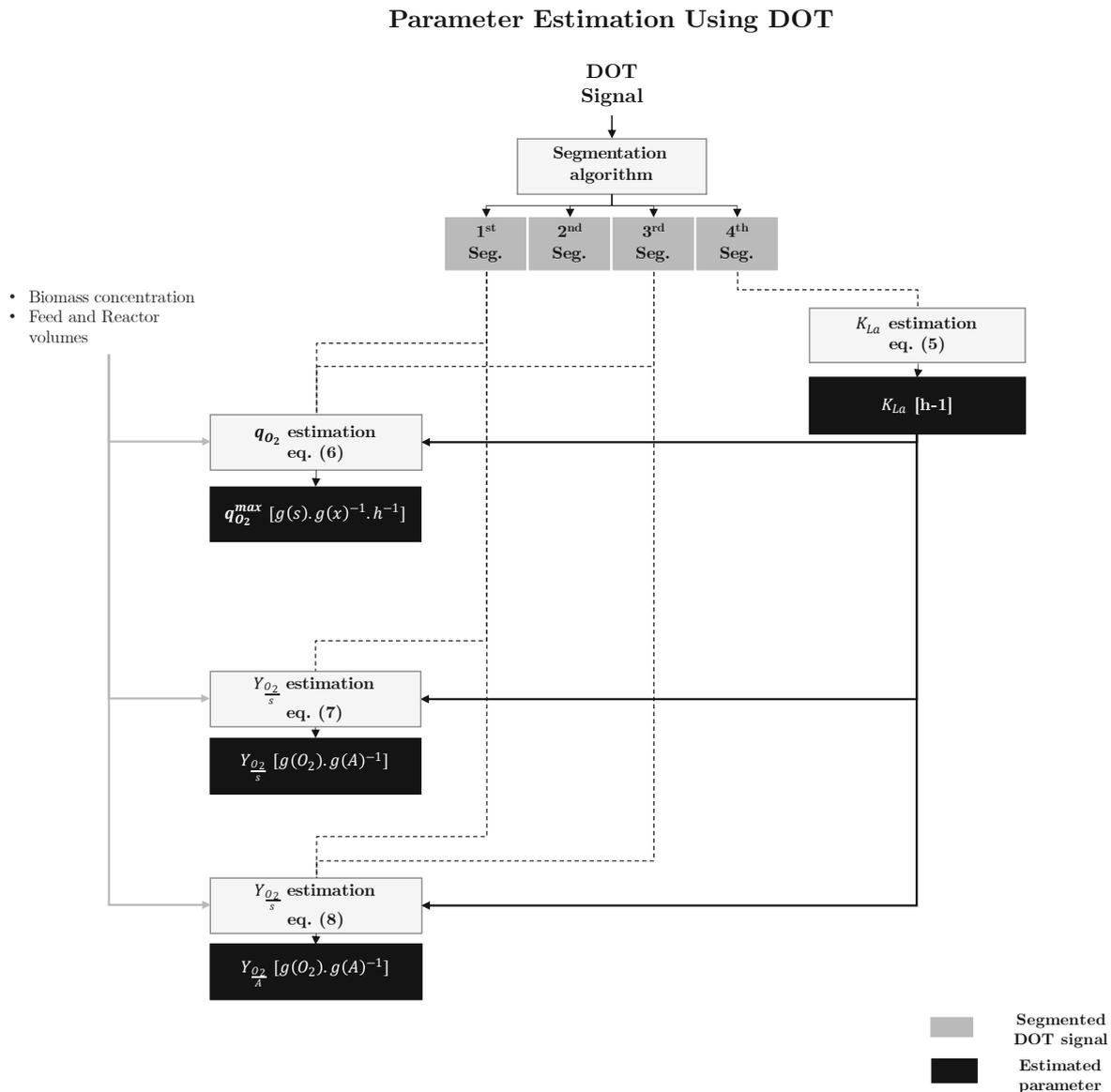
##### 4.1.1. Model hypothesis

From the visual inspection of the figures in section 3.2, the following hypotheses on the segments are made:

1. 1<sup>st</sup> segment: starts after the addition of the glucose to the medium with a short delay of roughly 2-4 [s]. The slope of this segment is always negative and smaller than the 3<sup>rd</sup> segment slope. Cells in this segment oxidize glucose (for the whole or most of the time window) in the overflow regime. Expectedly, the amount of accumulated acetate in the medium depends on the volume of the glucose added.
2. 2<sup>nd</sup> segment: the length and the slope of this segment differs between the experiments and along each experiment. This segment can be noticed as: an increase to higher DOT value, a flat, or a very slow decrease of the DOT signal, which indicates a transition phase from the 1<sup>st</sup> segment to the 3<sup>rd</sup> segment. The 2<sup>nd</sup> segment appears at dilution ratio higher than a certain threshold.
3. 3<sup>rd</sup> segment: starts after the end of the 2<sup>nd</sup> segment. The slope is always negative but less steep than the 1<sup>st</sup> segment slope. Cells are assumed to oxidize the accumulated acetate from the 1<sup>st</sup> segment.
4. 4<sup>th</sup> segment: starts when the DOT pulse reaches the minimal value, and ends when DOT reaches its starting point. This segment features a return to higher DOT values, mostly but not necessarily to the saturation value. The metabolism in the whole 4<sup>th</sup> segment is assumed to be inactive.

##### 4.1.2. Quantitative Analysis

Figure 10 shows a proposed workflow to extract model parameters using metrics and segments information defined in section 3.2 and the model hypothesis in section 4.1.1.



**Figure 10.** Parameter estimation workflow using DOT signal segmentation. The workflow can estimate the overflow condition parameters in *E. coli* using biomass samples, given the feeding and reactor working volumes are known.

To account for sensor delay in the signal, the actual dissolved oxygen signal  $DOT$  can be obtained from the measured dissolved oxygen signal  $DOT_m$ , following:

$$DOT = \tau \cdot \frac{dDOT_m}{dt} + DOT_m \quad (4)$$

$K_{La}$  value is calculated by the help of the segments time length using equation (5):

$$K_{La} = - \frac{\log \left( \frac{DOT^* - DOT^{end}}{DOT^* - DOT^{min}} \right)}{\Delta T_4} \quad (5)$$

The segments slopes can give information on the actual specific glucose uptake rate  $q_{s(ox)}$ .

In the first segment, the cell is assumed to consume glucose in the overflow regime, therefore  $q_{s(ox)} = q_{s(ox)}^{max} = \frac{q_{O_2}}{Y_{O_2/S}}$ .

In the third segment, the cell is assumed to oxidate only acetate where  $q_{A(ox)} = q_{A(ox)}^{max} = \frac{q_{O_2}}{Y_{O_2/A}}$ .

In the 1<sup>st</sup> segment, the specific oxygen uptake rate  $q_{O_2}$  for a DOT pulse in the neighborhood of a biomass sample with a concentration  $C_x$  is calculated following the equation (6):

$$q_{O_2} = \frac{\left. \frac{dDOT}{dt} \right|_{[\Delta T_1]} + (DOT^* - DOT_1^{end}) \cdot K_{La}}{H \cdot C_x} \quad (6)$$

Where  $H$  is Henry derived constant.  $q_{O_2}$  can be similarly calculated from the 3<sup>rd</sup> segment.

Model parameter  $q_{O_2}^{max}$  can be calculated, under the assumption that the instantaneous glucose addition causes a maximum oxygen uptake rate, at least at the beginning, simply as  $q_{O_2}^{max} = MAX(q_{O_2})$ .

The previous assumption holds true for the 3<sup>rd</sup> segment, only if the accumulated acetate concentration is high enough to cause maximal uptake in the cell. Therefore, it is better to calculate  $q_{O_2}^{max}$  using the 1<sup>st</sup> segment.

The amount of oxygen needed to oxidize a certain amount of glucose, is determined by the stoichiometric yield coefficient  $Y_{O_2/S}$ . This value can be either calculated from the stoichiometric matrix or estimated empirically as a model parameter. However, in a DOT pulse with only 1<sup>st</sup> and 4<sup>th</sup> segments and an equal start and end values, the amount of oxygen delivered to the cell is known  $O_2 \text{ mass } [mg(O_2)] = V \cdot \int OUR \cdot dt = V \cdot \int OTR \cdot dt = V \cdot \int (DOT^* - DOT) \cdot K_{La} \cdot dt$ . The amount of glucose delivered to the cell during this time window is known, therefore the yield can be calculated by the help of the segments area, as the equation (7) reads:

$$Y_{O_2/S} = \frac{V \cdot \int_{t_{start}}^{t_{end}} (DOT^* - DOT) \cdot K_{La} \cdot dt}{F_s \cdot C_{s,in}} \quad (7)$$

Where  $V$  is reactor working volume,  $K_{La}$  is the estimated value from eq. (5). Similarly, the amount of oxygen needed to oxidize a certain amount of acetate, is determined usually by the stoichiometric yield coefficient  $Y_{O_2/A}$ . The amount of accumulated acetate can be determined by knowing the glucose flux that exceeds the maximum oxidative capacity. This can be written as:  $Acetate \text{ mass } [mg(A)] = (F_s \cdot C_{s,in} - \int q_{s(ox)}^{max} \cdot dt \cdot C_x) \cdot Y_{A/S}$ .

The amount of oxygen that goes to oxidize acetate can be extracted by integrating the oxygen uptake rate along the 3<sup>rd</sup> segment and substituting for the DOT difference between  $t_3^{start}$  and  $t_3^{end}$ . This can be easily calculated by extrapolating DOT from  $t_3^{end}$  to  $t_3^{start}$  where the DOT value equals DOT value at  $t_3^{start}$ . The yield can be therefore written, by the help of the 1<sup>st</sup> and 3<sup>rd</sup> segments areas, as:

$$Y_{O_2/A} = \frac{V \cdot \left( \int_{t_3^{start}}^{t_3^{end}} (DOT^* - DOT) \cdot K_{La} \cdot dt \right)}{\left( F_s \cdot C_{s,in} - \int_{t_1^{start}}^{t_1^{end}} q_{s(ox)}^{max} \cdot dt \cdot C_x \right) \cdot Y_{A/S}} \quad (8)$$

The results of the parameters estimation workflow and relevant values are listed in Table 3.

**Table 3.** Estimated parameters and relevant values.

Reactor Nr.	Estimated parameters								Estimated relevant values			
	$K_{La}$		$q_{O_2}^{max}$		$Y_{O_2/S}$		$Y_{O_2/A}$		$OUR$		$q_o$	
	$-\sigma$	$+\sigma$	$-\sigma$	$+\sigma$	$-\sigma$	$+\sigma$	$-\sigma$	$+\sigma$	$-\sigma$	$+\sigma$	$-\sigma$	$+\sigma$
A	180	250	0.09	0.10	0.05	0.08	0.07	0.15	0.5	0.6	0.05	0.12
B	200	250	0.10	0.15	0.05	0.07	0.07	0.15	0.6	0.8	0.07	0.15
C	220	320	0.10	0.20	0.06	0.09	0.11	0.25	1.0	2.0	0.10	0.20
D	200	300	0.10	0.20	0.05	0.1	0.10	0.21	1.0	2.0	0.07	0.20
E	200	330	0.09	0.23	0.05	0.1	0.09	0.20	1.0	2.0	0.07	0.20
F	190	220	0.08	0.15	0.05	0.1	0.10	0.20	1.0	1.7	0.05	0.15
G	180	250	0.08	0.15	0.06	0.11	0.10	0.20	1.0	3.0	0.05	0.15
H	160	185	0.06	0.14	0.05	0.08	0.08	0.19	1.0	2.0	0.04	0.12

#### 4.2. Combined modelling of DOT response behavior

The discrete growth model of [29] is considered. The metabolic switches between the metabolic states are implemented using the Event Driven Method (EDM) similar to [30]. This approach seeks an accurate location of the metabolic events, distinctly different from other approaches focusing on formulating a continuous metabolic transition, e.g., [31]. A discontinuous (piecewise) modeling approach to incorporate metabolic switches and states can help to acquire accurate results [30], especially when the changes happen in a short timescale.

The actual dissolved oxygen signal ( $DOT$ ) is measured with a first order delay  $\tau = 36[s]$  caused by the response time of the sensor, therefore an additional equation for the measured dissolved oxygen signal ( $DOT_m$ ) is also considered, the actual dissolved oxygen equation reads:

$$\frac{dDOT}{dt} = OUR - OTR = (DOT^* - DOT) \cdot K_{La} - (Y_{O_2/S} \cdot q_{s(ox)} + Y_{O_2/A} \cdot q_{A(ox)}) \cdot H \cdot C_x \quad (9)$$

After considering the probe response time, the measured dissolved oxygen reads:

$$\frac{dDOT_m}{dt} = \frac{1}{\tau} \cdot (DOT - DOT_m) \quad (10)$$

More on the model and the nomenclature is found appendix 0.

Following the hypotheses made in section 4.1.1. The following submodels are considered:

- I. *Submodel I: Overflow metabolism:* the cells consume glucose at the maximum oxidative uptake rate. The excess glucose goes to form acetate. This state is active during the 1<sup>st</sup> segment.

- II. *Submodel II*: Adaptation state: in which the metabolism is paused or attenuated. This state is active always after the end of the overflow metabolism (*submodel I*) when the acetate accumulation exceeds a certain threshold (assumed to be 0.1 [mg/ml], similar to the literature values [16], [32]). The length of this state is linked to the concentration of the accumulated acetate. This state is only active for the time window of the 2<sup>nd</sup> segment.
- III. *Submodel III*: Acetate oxidation: the acetate is exclusively oxidized. This state is active during the 3<sup>rd</sup> segment.
- IV. *Submodel IV*: Static state: no active metabolic activities. This state is active the whole time except when there is a glucose pulse. It is also active during the 4<sup>th</sup> segment.

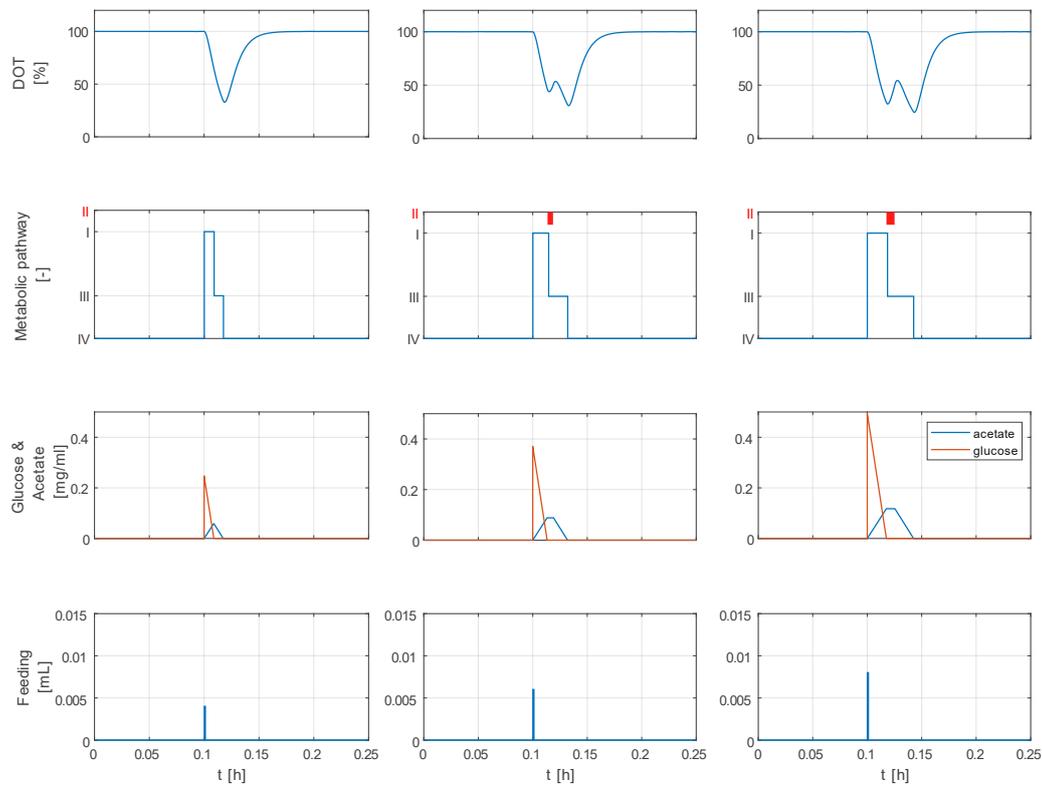
### Adaptation state as a model extension

The time length  $\Delta t$  of the adaptation state is influenced by the concentration of acetate, therefore it is proposed to be defined as  $\Delta t = t_{adap}^{max} (1 - \frac{C_A^{max} - C_A}{C_A^{max}})$ .  $C_A^{max}$  is the maximum acetate concentration observed in the medium,  $t_{adap}^{max}$  is the maximum adaptation time observed in the analysis. These values can be empirically identified. In the adaptation state, all model rates  $q$  including  $q_{A(ox)}$  and  $q_{S(ox)}$  are set to reduced values  $q_{adap}$  by a reduction factor  $R(t)$ . A reduction factor of  $R(t) = 100\%$  means the cell stops fully to uptake the substrates. For this contribution a complete reduction of metabolic activities is assumed. Model rates then can be written as:

$$q_{adap} = q \cdot R(t) \quad (11)$$

The biomass change during this short time window can be negligible. The relevant model parameters are taken from the analysis results made earlier in Table 3 (for reactor E), rest of the model parameters are listed in appendix 0.

The simulation results in Figure 11 are shown for three increasing glucose pulse volumes (4, 6 and 8 [ $\mu$ L]). The simulation results show that when the adaptation state (eq. (11)) is considered in the model, DOT pulses exhibit a response behavior similar to the pulses reported in Figure 5. These DOT pulses show four segments: a decline, a small increase, then another decline with a different slope, and finally a return to the saturation value. The simulated 2<sup>nd</sup> and 3<sup>rd</sup> segments emerge when the glucose pulse volume exceeds the defined threshold. The length and depth of the 2<sup>nd</sup> segment is related to the added glucose volume. A glucose pulse under the defined threshold results in a DOT pulse with two segments: a straight decline followed by return to the saturation value.



**Figure 11.** Mechanistic model simulations of a DOT pulse after adding one glucose feeding pulse per time. The relevant parameters of the simulation are calculated from DOT signal analysis in section 4.1.2 and shown in Table 3 (experimental run E). Second row shows the active submodel: I) overflow metabolism, II) adaptation state: metabolism is paused, III) acetate oxidation, and IV) static: no active substrate metabolism. The three columns show different feeding pulses. Middle and right columns show how feeding pulses exceeding a certain threshold triggers an adaptation state in the cell. The adaptation state can be observed by the existence of the 2<sup>nd</sup> segment in the DOT signal.

## 5. Discussion

### 5.1. Hypothesis verification & quantitative analysis

The analysis hypothesis on the relationships between the metabolic activities and DOT signal segments is supported by the results of the quantitative analysis and the mechanistic model simulations. Additionally, the following observations can be noticed:

- I. 1<sup>st</sup> segment: the observed delay time (2–4 [s]), after which the cell starts to actively metabolize glucose seems within the range of  $\tau_4$  reported by [24]. The authors referred to this delay as “Light-off phenomenon”. The assumption about cell metabolizing the glucose in the overflow metabolism is aligned with literature findings [15], [16]. Since the pulse injection time is very short (around 1 [s]), a sudden increase in glucose concentration in the medium is expected. This triggers the overflow metabolism. The relatively low  $q_{O_2}^{max}$  in Table 3 supports this assumption.
- II. 2<sup>nd</sup> segment: the change of the slope of this segment seems to be correlated with the amount of accumulated acetate. This segment appears only after a certain dilution ratio threshold as exemplarily shown in Figure 5.

- III. 3<sup>rd</sup> segment: this segment appears only after the 2<sup>nd</sup> segment. Figure 7 shows a likely positive correlation between the segment length and the dilution ratio per biomass unit. A plausible explanation is that with a high enough dilution ratio per biomass concentration, the acetate production under the overflow metabolism in the 1<sup>st</sup> segment is triggered. In the 3<sup>rd</sup> segment, the cells consume the accumulated acetate. The time needed for that is correlated with the amount of acetate produced, and by that, the time length is correlated with dilution ratio per biomass unit. Figure 12 shows no acetate accumulations in the neighborhood of the DOT pulses. This further supports the notion of transient production of acetate in the 1<sup>st</sup> segment, and the transient and full consumption of acetate in the 3<sup>rd</sup> segment. Additionally, for this segment, it is assumed to be negligible to no glucose concentration in the medium for this time window. Figure 12 shows no considerable glucose concentration at all for all runs. However, due to the sparsity and lack of enough glucose and acetate samples, this can't be thoroughly verified.
- IV. 4<sup>th</sup> segment: the assumption on inactive metabolism in this segment matches literature findings.

The pulse area analysis in Figure 9 shows relatively linear trends, suggesting a possible relationship with two factors: the amount of glucose added to the medium and the biomass concentration. The exact relationship between the areas and these factors is hard to be estimated due to the low number of biomass samples. The area of the pulses is directly linked to the amount of oxygen deposited in the medium along the time span of the DOT pulse. Equation (7) shows one possible mathematical description of this observation. The absolute amount of glucose added is known, and the absolute amount of oxygen consumed per unit of biomass concentration can be calculated by integrating the oxygen uptake rate (OUR) over the time window of the pulse, hence, the cell oxygen to glucose yield  $Y_{O_2/S}$  can be calculated. Once  $Y_{O_2/S}$  is calculated, oxygen to acetate yield  $Y_{O_2/A}$  can be also calculated by the help of glucose to acetate yield  $Y_{A/S}$  as described in the equation (8).

The results in section 4.1.2 shows systematic steps to quantitatively assess model parameters and relevant terms. Table 3 shows the extracted values, and Table 4 shows the corresponding literature values.

**Table 4.** Literature values of the volumetric mass transfer coefficient  $K_{La}$ , oxygen uptake rate  $OUR$ , oxygen specific uptake rate  $q_o$ , and the working volume  $V$  for *E. coli* cultivations for MBRs systems.

#	$K_{La}$ [ $h^{-1}$ ]	$OUR$ [ $g \cdot L^{-1} \cdot h^{-1}$ ]	$q_o$ [ $g \cdot g^{-1} \cdot h^{-1}$ ]	$V$ [ $mL$ ]	Notes	Literature
1	20-75	-	-	0.15	Impeller speed up to 200-800 [ $rpm$ ], OD (600 nm) up to 6 [-]	[33], [34]
2	58-90	~0.5	-	1	Dry cell weight up to 0.33 [ $g \cdot L^{-1}$ ]	[18]
3	90-400	-	-	6	Impeller speed 1300-1850 [ $rpm$ ], dry cell weight up to 1.8 [ $g \cdot L^{-1}$ ]	[35]
4	400-1440	-	-	10	Impeller speed up to 2800 or 4000 [ $rpm$ ], dry cell weight up to 16.5 [ $g \cdot L^{-1}$ ]	[7]
5	180-720	1-3.6	0.3-0.5	8-12	Impeller speed 1080-2400 [ $rpm$ ], dry cell weight up to 20.5 [ $g \cdot L^{-1}$ ]	[17]

6	Up to 1440	-	-	8-14	Impeller speed 3000 [rpm], dry cell weight up to 13 [g.L <sup>-1</sup> ]	[19]
7	216-396	-	-	10-100	Impeller speed up to 100-7000 [rpm], dry cell weight up to 10 [g.L <sup>-1</sup> ]	[36], [37]

The value of  $OUR$ , and  $K_{La}$  seem comparable to literature values of experiments with similar volume and biomass concentration.  $q_o$  the oxygen specific uptake rate value seems slightly lower than the value reported by [17].

The yields  $Y_{O_2/S}$  and  $Y_{O_2/A}$  values seem to be lower than the values reported in literature. [16] reported values of almost 1 [g.g<sup>-1</sup>] for both yields. [31] reported yields of 1.56 and 0.54 [g.g<sup>-1</sup>] and later they reported yields of 1.08 and 1.2 [g.g<sup>-1</sup>] for  $Y_{O_2/S}$  and  $Y_{O_2/A}$  consecutively [38]. The first two contributions show modelling results for a lab scale reactor and the last one show reported results for minibioreactor scale. However, these contributions did not report on the corresponding  $K_{La}$  value.

A common challenge when estimating model parameters is to find  $K_{La}$  value that helps to set the delicate balance between the two components of DOT signal: the oxygen uptake rate and oxygen transfer rate. These components are mainly influenced by the values of the parameters  $K_{La}$  on one side and  $Y_{O_2/S}$  and  $Y_{O_2/A}$  on the other side (given  $q_{O_2}^{max}$  is estimated and has a fixed value). The positive correlation between the parameters means that high yields coefficients values imply high  $K_{La}$  value and vice versa. Therefore, high  $K_{La}$  is expected for the previous contributions.

The proposed DOT analysis can be of great importance to achieve simultaneous estimation of the  $Y_{O_2/S}$ ,  $Y_{O_2/A}$  and  $K_{La}$ . According to the author's knowledge, this is not presented before in literature for minibioreactor systems.

By estimating the parameters  $q_{O_2}^{max}$  and  $Y_{O_2/S}$ , the overflow switching condition  $q_{S(ox)}^{critical} = \frac{q_{O_2}}{Y_{O_2/S}}$  is identified. The oxygen affinity constant  $K_O$  appears to be an insensitive parameter in our analysis.

The slope of the 4<sup>th</sup> segment is directly linked to the volumetric mass transfer coefficient  $K_{La}$ . Almost constant values of the 4<sup>th</sup> slope can be seen for all runs in Figure 8, indicating an almost constant  $K_{La}$  value along the time course of each run. However, a comparison of the slope of the 4<sup>th</sup> segment between the runs, shows a positive drift with more feeding. The relevant process variables, like stirring speed and aeration rate are kept constants for all experimental runs in the analyzed time window. Therefore, the drift perhaps is only linked to the changes in the characteristics of the medium. With more feeding, viscosity of the medium changes mainly because of higher biomass concentration.

The hypothesis on the relationship between the 4<sup>th</sup> segment slope and  $K_{La}$  is supported by the quantitative analysis results in Table 3 which show relatively comparable  $K_{La}$  values that match literature findings in Table 4.

Further, the mechanistic model simulation results of DOT in section 4.2 which uses  $K_{La}$  values reported in Table 3 supports the hypothesis.

It is interesting to witness that highest estimated  $K_{La}$  values in Table 3 are reported for runs D and E. These runs are duplicates of the standard feeding plan reported in [8]. Higher feeding plans (i.e., runs F, G and H) seems to result in higher viscosity, resulting in lower delivery of oxygen from the gas to the liquid. Lower feeding plans (i.e., runs A, B and C) seems to result in lower working volume, which might affect the oxygen transfer rate delivery negatively.

The slopes of the 1<sup>st</sup> and 3<sup>rd</sup> segments in Figure 8 also have consistent values within each run, but slightly drifted values in between the runs. This can be explained by the changes in the uptake rates in the new cell generations caused by the intermittent feeding [15]. The slopes of these segments can be linked directly to the oxygen uptake rate  $q_{O_2}$ . However, given an almost instantaneous addition of the substrate, a maximum and constant value of the oxygen uptake rate is expected for most of the time windows of the 1<sup>st</sup> segment. For the 3<sup>rd</sup> segment, a similarly constant value of  $q_{O_2}$  is probable if enough

acetate accumulates. If that's the case, then the difference in the slopes of the 1<sup>st</sup> and 3<sup>rd</sup> segments can be explained by the difference in the oxygen to glucose yield  $Y_{O_2/S}$  and oxygen to acetate yield  $Y_{O_2/A}$ . However, the lack of enough acetate samples hinders checking this hypothesis, but the quantitative analysis provided in section 4.1.2, does suggest a difference in the values of the  $Y_{O_2/S}$  and  $Y_{O_2/A}$ .

The 2<sup>nd</sup> segment slope, in all experimental runs, starts with positive values for the first few hours, then becomes close to zero, and in some instances slightly negative. This indicates an increasing, almost flat, and slowly decreasing changes in the DOT signal consecutively. The range of slope changes in which the 2<sup>nd</sup> segment is captured is defined by the segmentation algorithm, in which tuning parameters are dynamically estimated when the algorithm is trained, see appendix 7.1 for details. Biomass concentration (not visible in Figure 12) and the working volume of all reactors have similar values at the beginning of the analyzed time window, therefore the high slope values in the first couple of hours can be explained exclusively by the large feeding pulses. A visual inspection Figure 5 reveals that the 2<sup>nd</sup> segment existence is usually linked to feeding pulses with high amplitudes.

#### 5.1.1. Newly observed DOT response behavior

Figure 6 shows the detected time length of the 2<sup>nd</sup> segment. The figure depicts values in the range of 8 to 68 seconds with a mean value of around 31 seconds.

[24], [25] reported on monitoring *E. coli* metabolic response to glucose pulses by using a strain with integrated bioluminescence protein that allows for an online monitoring of the changing metabolism. Their observations showed that *E. coli* can switch from overflow to acetate oxidation "rapidly" and this switch usually is aligned with an overshoot in the bioluminescence with a peak lasting for almost a minute.

This metabolic change could happen because of cell stress, or when part of the cell population switches while the rest do not, or as a mix of both factors.

[26], [27] applied nuclear magnetic resonance techniques to monitor the metabolic switches in *E. coli*. Their observation showed a rapid induction of "acs", the gene responsible for acetate synthase after the metabolic switch (from overflow to acetate oxidation). An overlapping time window between acetate consumption and acetate production might have an effect on cell metabolism in the time window around the switch in which a co-utilization of acetate and glucose happens. Also, the authors reported on a drop in the growth rate directly after the switch.

[28] used Isotope Dilution Mass Spectrometry (IDMS) to analyze the metabolic changes after a glucose pulse at seconds timescale. Interestingly: *E. coli*, at the scale of tens of seconds after overflow, can store relevant amounts of carbon, to be used in short periods of glucose starvation.

[16] also showed a pulse-based method for the determination of the maximum uptake capacities for glucose and oxygen in glucose limited cultivations. Their observations showed that acetate is formed after a glucose pulse. However, the redirection of the acetate flow from production to consumption takes some time. The authors did not report on a specific time duration. In their contribution, the sampling time interval of the DOT signal was 5 seconds, and a change in the DOT signal similar to the 2<sup>nd</sup> segment presented in this contribution was shown for a couple of sampling points. However, this is neither highlighted nor discussed. The authors also report on the increase in  $q_o$  rate after glucose pulses. This is described by the uncoupling effect (inhibition effect) of acetate, although the added acetate concentration was low.

The previous observations suggest a metabolic switching time similar to the time range reported in Figure 6, and provide possible explanations of the metabolic changes in this time window. Hence, incorporating an "adaptation state" in the model, which represents a reduction in the metabolic activities after switching from the overflow metabolism, seems feasible.

The results of Figure 11, shows that considering adaptation state (submodel II), in which the metabolism is paused for a short time, results in DOT signal changes similar to the 2<sup>nd</sup> segments seen in the raw data. The simulations also show that after a certain threshold the 2<sup>nd</sup> segment emerges clearly and becomes more pronounced with larger feeding volumes. larger feeding volume means cells need more time to fully consume the glucose added to the medium. As the cell is already working at its maximum uptake rate, the excess sugar is metabolized anaerobically, and acetate accumulates in the medium. The 2<sup>nd</sup> segment becomes more pronounced as a result of prolonged adaptation state caused by higher acetate accumulation.

### 5.2. Industrial relevance

The minibioreactor systems are increasingly seen as useful tools in pharmaceutical and bioprocessing industry for many purposes like strain screening and experimental design. They do not inherit some larger scale issues like inhomogeneity, mixing and aeration difficulties and they offer an economically viable option to cut costs. However, scaling experiments up/down from/to milliliter scale is still a challenging issue. [38] recently reported on this, where they showed deviations in parameters values compared to their reference cultivation and reported on an increased amount of some amino acids particularly norvaline when bolus feeding is used. However, our observations further show that the frequent metabolic switching could have a negative impact of that on key parameters of the cell.

The hypothesized adaptation phenomenon that repeatedly happens in minibioreactor systems with intermittent bolus feeding seems to cause frequent cell stress. The relatively low values of the estimated parameters (e.g.,  $q_{O_2}^{max}$ ,  $Y_{O_2/S}$  and  $Y_{O_2/A}$ ) and the general tendency towards lower values in Table 3, shows the negative impact of the frequent switching on cell metabolism.

Additionally, in larger reactor scales, the inhomogeneities in the medium can trigger a similar behavior of metabolic switching in some local regions in the reactor [15]. The proposed analysis, by quantifying the metabolic adaptation time, can be used as a strain selector to choose strains that can better endure these effects.

### 5.3. Future Outlook

Further in-vitro investigation on the physiology behind the adaptation state in minibioreactor systems is needed to reveal more on this phenomenon on genetic, proteomic and metabolic level.

Future experimental plans to overcome the practical limitation of the used minibioreactor systems can help to provide more detailed analysis and provide more information on the cell status. For example, by sampling immediately before and after the glucose pulse, additional information on the maximum cell substrates uptake rates can be obtained.

Further model-based analysis to calculate the sensitivities of all model parameters to the DOT signal, and the degree of metabolic reduction in the time window of the adaptation state could help to assess how much information on model parameters could be encoded in the DOT signal.

However, with the current level of understanding it is possible to incorporate the dynamics of the adaptation state in the models to better control cultures: to prevent oxygen depletion, optimize glucose feeding, and understand the influence bolus feeding on cell behavior. The authors plan to report on that in the future.

## 6. Conclusions

A segmentation and analysis method, and a mechanistic modelling approach to model the dissolved oxygen tension signal in minibioreactor systems are proposed in this

work. The segmentation and analysis method revealed the existence of four segments in some DOT pulses and investigated possible relationships between descriptive metrics of the segments and metabolic activities and process dynamics. The findings of the analysis hypothesize a repeated metabolic switching behavior in *E. coli* after each substrate pulse, where the metabolic states are linked to the identified segments of the DOT pulses. A newly observed DOT segment, not reported in literature, is likely to be linked to a metabolic adaptation behavior. In this segment, the cell is likely to pause or attenuate the metabolism.

A quantitative analysis and mechanistic model simulations support this hypothesis. The quantitative analysis shows model parameters values within acceptable literature ranges. The mechanistic model simulations show a possibility to reproduce DOT segments found in the raw data by using the parameters estimated from the quantitative analysis and extending the model of [29] by a metabolic adaptation submodel. The time length of this submodel is a function of the inhibitory acetate concentration.

For our quantitative analysis, the estimation of model parameters of the overflow switching condition was possible using only DOT signal and biomass samples, given the feeding and reactor working volumes were known.

The analysis suggests a negative impact on some model parameters such as the maximum oxidative uptake rate and oxygen yields on glucose and acetate, which might be caused by the frequent metabolic switching.

The proposed methods highlight the potential of considering the DOT signal to gain additional (unexploited) information on *E. coli* metabolism and its modelling parameters, especially when the number of samples is limited, and offer means to understand the influence of intermittent bolus feeding on cell behavior, and by that, help to address MBRs issues of oxygen supply and feeding plans optimization.

**Author Contributions:** Conceptualization, A.J., J.K. and T.B.; methodology, A.J.; software, A.J.; validation, A.J., T.B., J.K. and C.H.; investigation, A.J.; resources, V.R.; writing—original draft preparation, A.J.; writing—review and editing, A.J., T.B., J.K. and C.H.; visualization, A.J.; supervision, T.B. and C.H.; project administration, C.H.

All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was partially funded by the Austrian Research Funding Association (FFG) within both: the program Bridge 1 in the project “AdaMo” (No. 864705).

**Data Availability Statement:** not applicable.

**Acknowledgments:** The authors thank Boehringer Ingelheim RCV GmbH & Co KG for their participation in the project and for the fruitful discussions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## 7. Appendices 728

### 7.1. Appendix A: Segmentation algorithm 729

#### Pulse detection algorithm 730

Segmentation algorithm has the following steps: 731

- 1) divide the whole signal into  $k \geq$  time windows. 732
- 2) for each time window calculate the mean of the signal. The result is a vector containing all the means. 733
- 3) The intersection of this vector with DOT signal, gives the intersection points. 734
- 4) a loop checks if within two consecutive intersection points, the DOT signal has two maximums and one minimum. If so, a pulse is identified, the time points of the two maxima define the beginning and the end times  $t_i^{start}$  and  $t_i^{end}$ . The time point of the minimum is  $t_i^{min}$ . 735

$k$  represents a tuning parameter, usually set to large values with very changing DOT signals. 736

The result is a vector of individual DOT pulses  $P$  : 737

$$P(i) = [t_i^{start} ; t_i^{min} ; t_i^{end}] \quad ; P = \bigcup_1^n P(i) \quad (12) \quad \text{742}$$

Where: 743

$i$ : the index number of the pulse of the glucose pulse (and the DOT signal). 744

$n$ : the total number of glucose pulses (and the DOT signal). 745

$t_i^{start}$ : the start time of each DOT pulse. 746

$t_i^{min}$ : the time corresponds to the minimum value of each DOT pulse. 747

$t_i^{end}$ : the end time of each DOT pulse. 748

The appendix is an optional section that can contain details and data supplemental to the main text—for example, explanations of experimental details that would disrupt the flow of the main text but nonetheless remain crucial to understanding and reproducing the research shown; figures of replicates for experiments of which representative data is shown in the main text can be added here if brief, or as Supplementary data. Mathematical proofs of results not central to the paper can be added as an appendix. 749

#### Segmentation algorithm 750

For an experimental run  $j$ , and a vector  $P(i)$ : 751

- 1) split the pulse  $P(i)$  into two pieces by the  $t_i^{min}$ . These results into: the “up-down” part (contains 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> segments) and “down-up” part (contains only 4<sup>th</sup> segment) 752
- 2) for the “up-down” part, the algorithm advances along time steps and monitors the change of the slope. 753
- 3) once a recognizable slope change  $\alpha_1^j$  is observed, the algorithm considers an end of a segment and a beginning of another segment. 754
- 4) repeat the previous step until the 3 segments are found, the corresponding slope ( $\alpha_1^j, \alpha_2^j$ ) are defined. 755
- 5) if not all the 3 segments are definable, assume only the 1<sup>st</sup> for the “up-down” part. 756

The algorithm is trained for each experimental run by the help of a visual inspection of an expert on a part of the DOT signal (training dataset). The training process estimates the coefficients  $\alpha_1, \alpha_2$ . The rest of the signal is used as a testing dataset.

776

777

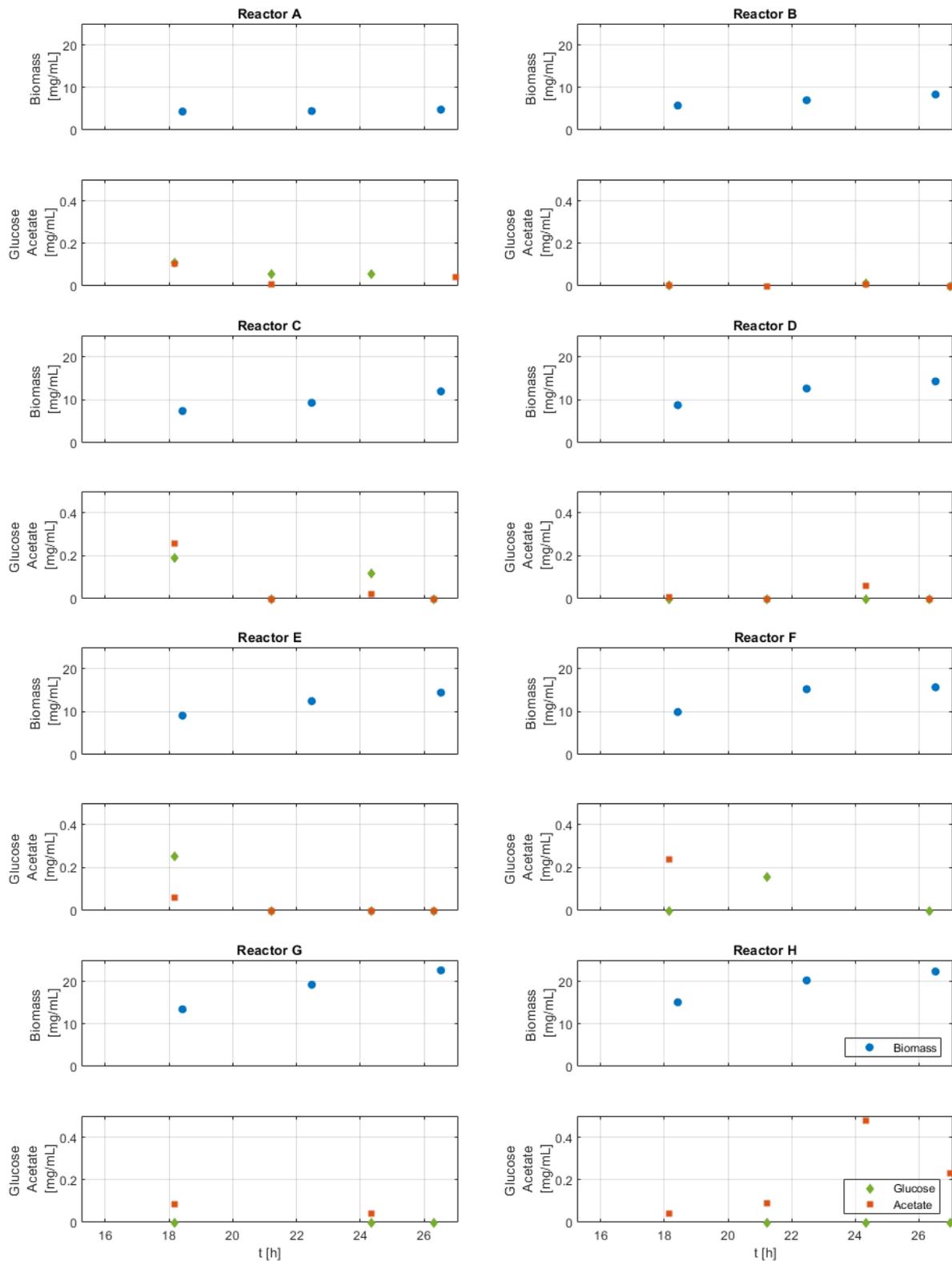
778

779

7.2. Appendix B: Additional Figures

780

781



782

783

784

**Figure 12.** Biomass, glucose, and acetate concentrations of each run in the time window of the analysis. Due to working volume limitation, only a limited number of off-line/at-line samples are possible.

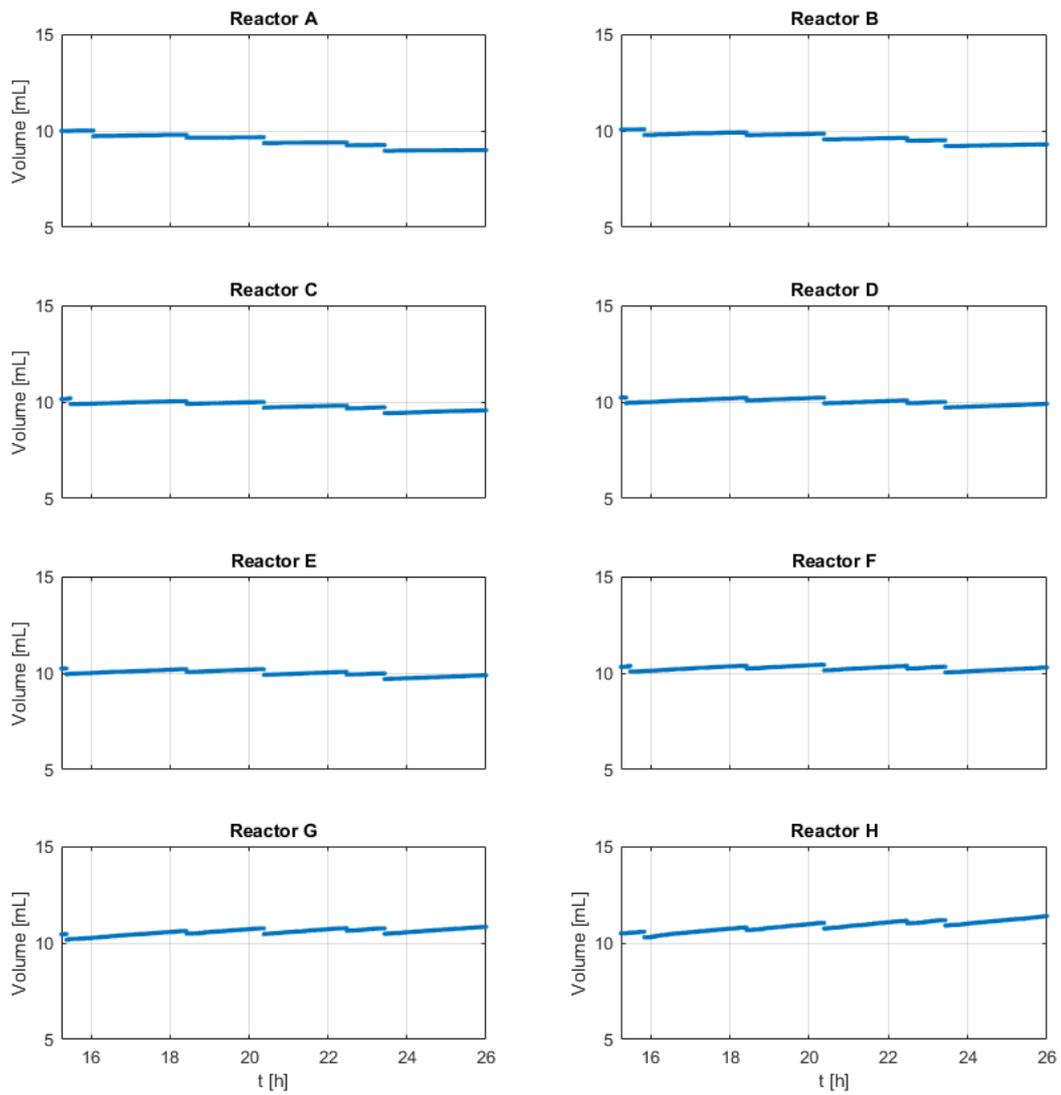


Figure 13. Reactor working volumes of all runs.

7.3. Appendix C: *E. coli* model and nomenclature

$$q_s = q_s^{max} \frac{C_s}{C_s + K_s}$$

$$q_A = q_A^{max} \frac{C_A}{C_A + K_A}$$

$$q_{O_2} = q_{O_2}^{max} \frac{DOT}{DOT + K_O}$$

$$q_{s(ox)}^{critical} = \frac{q_{O_2}}{Y_{O_2/s}}$$

$$q_{s(ox)} = \begin{cases} q_s & \text{if } q_s \leq q_{s(ox)}^{critical} \\ q_{s(ox)}^{critical} & \text{if } q_s > q_{s(ox)}^{critical} \end{cases}$$

$$q_{A(ox)}^{critical} = \frac{q_{O_2} - Y_{O_2/s} \cdot q_{s(ox)}}{Y_{O_2/A}}$$

$$q_{A(ox)} = \begin{cases} q_s & \text{if } q_A \leq q_{A(ox)}^{critical} \\ q_{A(ox)}^{critical} & \text{if } q_A > q_{A(ox)}^{critical} \end{cases}$$

$$q_{s(red)} = q_s - q_{s(ox)}^{max}$$

$$\mu_{total} = Y_{x/s(ox)} \cdot q_{s(ox)} + Y_{x/s(red)} \cdot q_{s(red)} + Y_{x/A(ox)} \cdot q_{A(ox)}$$

$$\frac{d}{dt} \begin{bmatrix} C_x \\ C_s \\ C_A \\ DOT \end{bmatrix} = \begin{bmatrix} Y_{x/s(ox)} & Y_{x/s(ox)} & Y_{x/s(red)} & Y_{x/A(ox)} & Y_{x/A(ox)} \\ -1 & -1 & -1 & 0 & 0 \\ 0 & 0 & Y_{e/s} & -1 & -1 \\ -Y_{O_2/s} & -Y_{O_2/s} & 0 & -Y_{O_2/A} & -Y_{O_2/A} \end{bmatrix} \cdot \mathcal{A} \cdot \begin{bmatrix} q_s \\ q_{s(ox)}^{critical} \\ q_{s(red)} \\ q_A \\ q_{A(ox)}^{critical} \end{bmatrix} C_x - \frac{F_s}{V} \begin{bmatrix} C_x \\ C_s \\ C_e \\ 0 \end{bmatrix} + K_{La} \begin{bmatrix} 0 \\ 0 \\ 0 \\ DOT^* \end{bmatrix} - K_{La} \begin{bmatrix} 0 \\ 0 \\ 0 \\ DOT \end{bmatrix}$$

Table 5. Activation matrix values for metabolic states defined in section 4.2.

Metabolic state / active submodel	Activation matrix $\mathcal{A}$
I) Overflow metabolism	$\mathcal{A} = \text{diag} [0 \ 1 \ 1 \ 0 \ 0]$
II) Adaptation state	$\mathcal{A} = \text{diag} [0 \ 0 \ 0 \ 0 \ 0]$
III) Acetate oxidation	$\mathcal{A} = \text{diag} [0 \ 1 \ 0 \ 0 \ 1]$
IV) Static state	$\mathcal{A} = \text{diag} [0 \ 0 \ 0 \ 0 \ 0]$

Sampling and feeding volumes are considered as sets of algebraic equations solved outside the ODE system. The timepoints of these pulses are considered as explicit time events, more on that in [39], [40] The values of the  $q_m$  the specific maintenance coefficient term is assumed to be zero for simplification. This value is not relevant for the scope of this contribution.

**Table 6.** Model terms description and parameter values.

822

PAR.	UNIT	SIMULATION VALUES	DESCRIPTION
$\mu_{max}$	$h^{-1}$	-	Maximum growth rate
$q_S^{max}$	$g(s).g(x)^{-1}.h^{-1}$	0.15	Maximum specific glucose uptake rate of the Monod function
$q_{O_2}^{max}$	$g(O).g(x)^{-1}.h^{-1}$	2.7	Maximum specific oxygen uptake rate of the Monod function
$q_A^{max}$	$g(A).g(x)^{-1}.h^{-1}$	0.8	Maximum specific acetate uptake rate of the Monod function
$q_{s(ox)}^{critical}$	$g(O).g(x)^{-1}.h^{-1}$	-	Maximum specific glucose uptake rate defined by the maximum oxidative capacity
$q_{A(ox)}^{critical}$	$g(A).g(x)^{-1}.h^{-1}$	-	Maximum specific acetate uptake rate by the maximum oxidative capacity
$q_{s(ox)}$	$g(s).g(x)^{-1}.h^{-1}$	-	Actual specific glucose uptake rate
$q_{A(ox)}$	$g(A).g(x)^{-1}.h^{-1}$	-	Actual specific acetate uptake rate
$Y_{X/S(red)}$	$g(x).g(s)^{-1}$	0.4	Biomass yield for reductive growth on glucose
$Y_{X/S(ox)}$	$g(x).g(s)^{-1}$	0.5	Biomass yield for oxidative growth on glucose
$Y_{X/A(ox)}$	$g(x).g(A)^{-1}$	0.5	Biomass yield for oxidative growth on acetate
$Y_{A/S}$	$g(A).g(s)^{-1}$	0.4	Ethanol yield from glucose fermentation
$Y_{O_2/A}$	$g(O_2).g(A)^{-1}$	0.5	Oxygen (stoichiometric) yield on acetate
$Y_{O_2/S}$	$g(O_2).g(s)^{-1}$	0.1	Oxygen (stoichiometric) yield on glucose
$K_{la}$	$h^{-1}$	225	Oxygen mass transfer coefficient from the gas phase to the liquid phase
$K_A$	$g(A).L^{-1}$	0.001	Time affinity constant of the acetate
$K_S$	$g(S).L^{-1}$	0.001	Time affinity constant of the glucose
$K_O$	$g(O_2).L^{-1}$	0.001	Time affinity constant of the oxygen
H	$\frac{\%}{g(O_2).L^{-1}}$	14000	Henry Law derived constant
$C_X$	$g(X).L^{-1}$	-	Biomass concentration
$C_S$	$g(S).L^{-1}$	-	Glucose concentration
$C_A$	$g(A).L^{-1}$	-	Acetate concentration
DOT	[%]	-	Dissolved oxygen tension
DOT <sub>m</sub>	[%]	-	Measured Dissolved oxygen tension
DOT*	[%]	-	Dissolved oxygen tension at saturation
$\tau$	[h]	0.01	Dissolved oxygen probe response time
V	[L]	-	Working volume
OUR	$g(O_2).L^{-1}.h^{-1}$	-	Oxygen uptake rate

OTR	$g(O_2).L^{-1}.h^{-1}$	-	Oxygen transfer rate
-----	------------------------	---	----------------------

823

824

## References

- [1] G. Faust, N. H. Janzen, C. Bendig, L. Römer, K. Kaufmann, and D. Weuster-Botz, 'Feeding strategies enhance high cell density cultivation and protein expression in milliliter scale bioreactors', *Biotechnol J*, vol. 9, no. 10, pp. 1293–1303, Oct. 2014, doi: 10.1002/biot.201400346.
- [2] J. I. Betts and F. Baganz, 'Miniature bioreactors: Current practices and future opportunities', *Microbial Cell Factories*, vol. 5, May 25, 2006. doi: 10.1186/1475-2859-5-21.
- [3] Q. Long *et al.*, 'The development and application of high throughput cultivation technology in bioprocess development', *J Biotechnol*, vol. 192, no. PB, pp. 323–338, Dec. 2014, doi: 10.1016/j.jbiotec.2014.03.028.
- [4] R. Bhambure, K. Kumar, and A. S. Rathore, 'High-throughput process development for biopharmaceutical drug substances', *Trends Biotechnol*, vol. 29, no. 3, pp. 127–135, Mar. 2011, doi: 10.1016/J.TIBTECH.2010.12.001.
- [5] T. v Kirk and N. Szita, 'Oxygen Transfer Characteristics of Miniaturized Bioreactor Systems', *Biotechnol. Bioeng*, vol. 110, pp. 1005–1019, 2013, doi: 10.1002/bit.24824/abstract.
- [6] D. Schäpper, S. M. Stocks, N. Szita, A. E. Lantz, and K. v. Gernaey, 'Development of a single-use microbioreactor for cultivation of microorganisms', *Chemical Engineering Journal*, vol. 160, no. 3, pp. 891–898, Jun. 2010, doi: 10.1016/J.CEJ.2010.02.038.
- [7] D. Weuster-Botz, R. Puskeiler, A. Kusterer, K. Kaufmann, G. T. John, and M. Arnold, 'Methods and milliliter scale devices for high-throughput bioprocess design', *Bioprocess Biosyst Eng*, vol. 28, no. 2, pp. 109–119, Nov. 2005, doi: 10.1007/s00449-005-0011-6.
- [8] N. H. Janzen, G. Striedner, J. Jarmer, M. Voigtmann, S. Abad, and D. Reinisch, 'Implementation of a Fully Automated Microbial Cultivation Platform for Strain and Process Screening', *Biotechnol J*, vol. 14, no. 10, p. 1800625, Oct. 2019, doi: 10.1002/biot.201800625.
- [9] B. Haby *et al.*, 'Integrated Robotic Mini Bioreactor Platform for Automated, Parallel Microbial Cultivation With Online Data Handling and Process Control', *SLAS Technol*, vol. 24, no. 6, pp. 569–582, Dec. 2019, doi: 10.1177/2472630319860775/ASSET/IMAGES/LARGE/10.1177\_2472630319860775-FIG2.JPEG.
- [10] Niels Krausch *et al.*, 'High-throughput screening of optimal process conditions using model predictive control', 2022, doi: 10.22541/au.164600498.85996662/v2.
- [11] R. Hortsch and D. Weuster-Botz, 'Milliliter-scale stirred tank reactors for the cultivation of microorganisms', in *Advances in Applied Microbiology*, vol. 73, no. C, Academic Press Inc., 2010, pp. 61–82. doi: 10.1016/S0065-2164(10)73003-3.
- [12] C. Li *et al.*, 'Numerical and experimental assessment of a miniature bioreactor equipped with a mechanical agitator and non-invasive biosensors', *Journal of Chemical Technology & Biotechnology*, vol. 94, no. 8, pp. 2671–2683, Aug. 2019, doi: 10.1002/JCTB.6076.
- [13] P. M. Doran, *Bioprocess engineering principles: Second edition*, vol. 9780080917. Academic Press, 2018.
- [14] T. Ferenci, 'Hungry bacteria--definition and properties of a nutritional state', *Environ Microbiol*, vol. 3, no. 10, pp. 605–611, 2001, doi: 10.1046/J.1462-2920.2001.00238.X.
- [15] E. Vasilakou, M. C. M. van Loosdrecht, and S. A. Wahl, 'Escherichia coli metabolism under short-term repetitive substrate dynamics: adaptation and trade-offs', *Microbial Cell Factories 2020 19:1*, vol. 19, no. 1, pp. 1–19, May 2020, doi: 10.1186/S12934-020-01379-0.
- [16] H. Y. Lin, B. Mathiszik, B. Xu, S. O. Enfors, and P. Neubauer, 'Determination of the maximum specific uptake capacities for glucose and oxygen in glucose-limited fed-batch cultivations of Escherichia coli', *Biotechnol Bioeng*, vol. 73, no. 5, pp. 347–357, Jun. 2001, doi: 10.1002/bit.1068.

- [17] R. Puskeiler, K. Kaufmann, and D. Weuster-Botz, 'Development, parallelization, and automation of a gas-inducing milliliter-scale bioreactor for high-throughput bioprocess design (HTBD)', *Biotechnol Bioeng*, vol. 89, no. 5, pp. 512–523, Mar. 2005, doi: 10.1002/bit.20352. 866–868
- [18] K. S. Lee, P. Boccazzi, A. J. Sinskey, and R. J. Ram, 'Microfluidic chemostat and turbidostat with flow rate, oxygen, and temperature control for dynamic continuous culture', *Lab Chip*, vol. 11, no. 10, pp. 1730–1739, 2011, doi: 10.1039/c1lc20019d. 869–871
- [19] A. Schmideder, T. S. Severin, J. H. Cremer, and D. Weuster-Botz, 'A novel milliliter-scale chemostat system for parallel cultivation of microorganisms in stirred-tank bioreactors', *J Biotechnol*, vol. 210, pp. 19–24, Sep. 2015, doi: 10.1016/j.jbiotec.2015.06.402. 872–874
- [20] R. Puskeiler and D. Weuster-Botz, 'Combined sulfite method for the measurement of the oxygen transfer coefficient  $k_La$  in bioreactors', *J Biotechnol*, vol. 120, no. 4, pp. 430–438, Dec. 2005, doi: 10.1016/j.jbiotec.2005.06.016. 875–876
- [21] S. Suresh, V. C. Srivastava, and I. M. Mishra, 'Techniques for oxygen transfer measurement in bioreactors: A review', *Journal of Chemical Technology and Biotechnology*, vol. 84, no. 8, pp. 1091–1103, 2009, doi: 10.1002/jctb.2154. 877–878
- [22] L. de Maré, L. Andersson, and P. Hagander, 'Probing control of glucose feeding in *Vibrio cholerae* cultivations', *Bioprocess Biosyst Eng*, vol. 25, no. 4, pp. 221–228, 2003, doi: 10.1007/s00449-002-0304-y. 879–880
- [23] P. Hagander and J. P. Axelsson, 'Probing control of fed-batch cultivations: analysis and tuning', *Control Engineering Practice*, no. 9, pp. 709–723, 2001. 881–882
- [24] S. Sunya, N. Gorret, F. Delvigne, J. L. Uribelarrea, and C. Molina-Jouve, 'Real-time monitoring of metabolic shift and transcriptional induction of *yciG*:: LuxCDABE E. Coli reporter strain to a glucose pulse of different concentrations', *J Biotechnol*, vol. 157, no. 3, pp. 379–390, 2012, doi: 10.1016/j.jbiotec.2011.12.009. 883–885
- [25] S. Sunya, F. Delvigne, J. L. Uribelarrea, C. Molina-Jouve, and N. Gorret, 'Comparison of the transient responses of *Escherichia coli* to a glucose pulse of various intensities', *Appl Microbiol Biotechnol*, vol. 95, no. 4, pp. 1021–1034, 2012, doi: 10.1007/s00253-012-3938-y. 886–888
- [26] B. Enjalbert, M. Coccagn-Bousquet, J. C. Portais, and F. Letisse, 'Acetate exposure determines the diauxic behavior of *Escherichia coli* during the glucose-acetate transition', *J Bacteriol*, vol. 197, no. 19, pp. 3173–3181, 2015, doi: 10.1128/JB.00128-15. 889–891
- [27] B. Enjalbert, F. Letisse, and J.-C. Portais, 'Physiological and Molecular Timing of the Glucose to Acetate Transition in *Escherichia coli*', *Metabolites*, vol. 3, no. 3, pp. 820–837, 2013, doi: 10.3390/metabo3030820. 892–893
- [28] H. Taymaz-Nikerel, W. M. van Gulik, and J. J. Heijnen, '*Escherichia coli* responds with a rapid and large change in growth rate upon a shift from glucose-limited to glucose-excess conditions', *Metab Eng*, vol. 13, no. 3, pp. 307–318, 2011, doi: 10.1016/j.ymben.2011.03.003. 894–896
- [29] B. Xu, M. Jahic, and S. O. Enfors, 'Modeling of overflow metabolism in batch and fed-batch cultures of *Escherichia coli*', *Biotechnol Prog*, vol. 15, no. 1, pp. 81–90, 1999, doi: 10.1021/bp9801087. 897–898
- [30] M. A. Jouned, J. Kager, C. Herwig, and T. Barz, 'Event driven modeling for the accurate identification of metabolic switches in fed-batch culture of *S. cerevisiae*', *Biochem Eng J*, vol. 180, p. 108345, Mar. 2022, doi: 10.1016/j.bej.2022.108345. 899–901
- [31] E. Anane, D. C. López C, P. Neubauer, and M. N. Cruz Bournazou, 'Modelling overflow metabolism in *Escherichia coli* by acetate cycling', *Biochem Eng J*, vol. 125, pp. 23–30, Sep. 2017, doi: 10.1016/J.BEJ.2017.05.013. 902–903
- [32] P. Vrabel, R. G. J. M. van der Lans, F. N. van der Schot, K. C. A. M. Luyben, B. Xu, and S. O. Enfors, 'CMA: Integration of fluid dynamics and microbial kinetics in modelling of large-scale fermentations', *Chemical Engineering Journal*, vol. 84, no. 3, pp. 463–474, 2001, doi: 10.1016/S1385-8947(00)00271-0. 904–906

- [33] Z. Zhang, N. Szita, P. Boccazzi, A. J. Sinskey, and K. F. Jensen, 'A well-mixed, polymer-based microbioreactor with integrated optical measurements', *Biotechnol Bioeng*, vol. 93, no. 2, pp. 286–296, Feb. 2006, doi: 10.1002/BIT.20678. 907  
908  
909
- [34] N. Szita, P. Boccazzi, Z. Zhang, P. Boyle, A. J. Sinskey, and K. F. Jensen, 'Development of a multiplexed microbioreactor system for high-throughput bioprocessing', *Lab Chip*, vol. 5, no. 8, pp. 819–826, 2005, doi: 10.1039/b504243g. 910  
911  
912
- [35] S. R. Lamping, H. Zhang, B. Allen, and P. Ayazi Shamlou, 'Design of a prototype miniature bioreactor for high throughput automated bioprocessing', *Chem Eng Sci*, vol. 58, no. 3–6, pp. 747–758, 2003, doi: 10.1016/S0009-2509(02)00604-8. 913  
914  
915
- [36] R. Bareither and D. Pollard, 'A review of advanced small-scale parallel bioreactor technology for accelerated process development: Current state and future need', *Biotechnol Prog*, vol. 27, no. 1, pp. 2–14, Jan. 2011, doi: 10.1002/btpr.522. 916  
917  
918
- [37] N. K. Gill, M. Appleton, F. Baganz, and G. J. Lye, 'Quantification of power consumption and oxygen transfer characteristics of a stirred miniature bioreactor for predictive fermentation scale-up', *Biotechnol Bioeng*, vol. 100, no. 6, pp. 1144–1155, Aug. 2008, doi: 10.1002/BIT.21852. 919  
920  
921
- [38] E. Anane, A. Sawatzki, P. Neubauer, and M. N. Cruz-Bournazou, 'Modelling concentration gradients in fed-batch cultivations of *E. coli* – towards the flexible design of scale-down experiments', *Journal of Chemical Technology and Biotechnology*, vol. 94, no. 2, pp. 516–526, 2019, doi: 10.1002/jctb.5798. 922  
923  
924
- [39] M. A. Jouned, J. Kager, C. Herwig, and T. Barz, 'Event driven modeling for the accurate identification of metabolic switches in fed-batch culture of *S. cerevisiae*', *Biochem Eng J*, vol. 180, p. 108345, Mar. 2022, doi: 10.1016/j.bej.2022.108345. 925  
926  
927
- [40] T. B. 463 M. Adnan JOUNED, Julian KAGER, Christoph HERWIG, 'Event driven analysis to enhance model calibration of experiments with high offline sampling rates', 2021, p. 463. 928  
929  
930

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content. 931  
932  
933

## 3 Conclusion, Impact, and Outlook

## 3.1 Conclusions, Impact and Outlook

### 3.1.1 Conclusions

To achieve the thesis aims of having adaptive and robust modelling approaches, comparative analyses are applied to different datasets using different methods. These comparisons alongside the proposed method helped to: reveal the obstacles and the possible remedies to achieve improved bioprocesses.

#### Summarized achievements

Part of the main achievements of this work are the comparative studies presented in sections 2.2.4, 2.2.3, 2.2.2, and 2.2.1. The comparative study in section 2.2.1 highlights the importance of proper selection of the fitting criteria inside the objective function which is used to parameterize the model. The physical nature of the measurements must be taken into account to get representative unbiased parameters. The proposed method that combines DTW and LSQ criteria can obtain extra information that is usually smoothed when exclusively LSQ-based criteria are used.

The comparative analysis in section 2.2.2 illustrates the danger of error propagation in mechanistic models even when a small deviation exists. This section shows the possible consequences of improper considering of explicit discontinues of volume changes because of sampling volumes in lab-scale reactors. The work also proposes a simple approach to get rid of these consequences.

The work in sections 2.2.3 and 2.2.4 extends the previous work in section 2.2.2 and shows a comparative study that investigates all possible discontinuous behaviors in bioprocessing models (shown for the model of [82] for yeast cultivation). The work highlights the main sorts of discontinuities in a generic bioprocessing model. They are A) Metabolic discontinuities: switches that happen between the metabolic pathways. B) Process-related discontinuities: switches or abrupt change in system states because of process-related change. C) Discontinuities caused by highly nonlinear kinetic terms (non-physical solutions): some kinetics exhibit a stiff behavior when they have parameter values in certain ranges. Bioprocesses usually operate in these ranges, which make this a common modelling problem.

The work shows a workflow to properly consider these discontinuities, and an approach to convert mechanistic models into matrix-like models using decision tree and Boolean condition functions. The work also shows the improvements of the proposed method on model prediction accuracy, model output robustness against noise and initial parameters and concentrations perturbations.

Section 2.2.5 presents a method to segment the dissolved oxygen signal for *E. coli* cultivations in minibioreactor systems. Having the segments allowed to extract discrete features for each segment. Data-driven analysis of the features revealed a pattern of responses in particular DOT segments linked to specific metabolic states. The analysis hypothesizes on the nature of the underlying metabolic activities and the DOT signal segments.

The quantitative analysis in this section, and the mechanistic model-based simulations support the hypothesis. The quantitative analysis provides a possibility to obtain key growth parameters of the overflow switching condition using only DOT signal and biomass samples.

## Novelty

The novelty of this thesis can be summarized by the:

- a) detailed analysis of the discontinuities in bioprocessing, and the EDM based workflow to properly handle bioprocessing models discontinuous behaviors in sections 2.2.2, 2.2.3 and 2.2.4.
- b) new method that combines DTW and LSQ terms in the objective function to improve the estimation of model parameters in yeast cultivation with off-gas measurements in section 2.2.1.
- c) novel segmentation algorithm, data-driven and model-based analysis of the DOT signal for *E. coli* cultivations in MBRs which suggest a pattern of responses in particular DOT segments which are linked to specific metabolic states, and highlight a plausible relationship between metabolic adaptation behavior and a newly observed DOT segment in section 2.2.5.
- d) novel model-based workflow to extract model parameters from DOT signals in section 2.2.5.

## Value

The value of this thesis lies in the analyses and methods designed to overcome the challenges described in section 1.1.3 which prevent an effective deployment of mechanistic models.

These methods help to improve the output of the models and give incentive for more work in the direction of developing mechanistic models in bioprocessing. Although, data-driven models and hybrid models are getting more interest, this thesis results suggest that more work to improve mechanistic models in bioprocessing is helpful to increase the information content and the applicability of these models. The impact of this thesis is described in more detail in section 3.1.2.

The following factors are proved to be key factors for the success of this thesis:

- Interdisciplinarity
- Collaboration with industry
- Combining methods in generic workflows

## Success factors

### I) Interdisciplinarity

Modelling bioprocesses is absolutely an interdisciplinary task. For the success of this thesis, an understanding of the differential calculus principles was needed to analyze the roots of model discontinuities on a mathematical level. Some knowledge in biology and cell physiology is needed to understand the meaning of model discontinuities on a metabolic level. The knowledge of data-driven methods also plays an essential part. These methods are the basics of many steps during the modelling procedure, from analytics (pre-)processing, to investigating possible latent relationships that don't have clear interpretations or mathematical representations.

Further, some knowledge in the mathematical optimization field is an advantage. This is essential for the formulation of the objective functions and the analysis and interpretation of the optimization algorithms results.

### II) Collaboration with industry

Academia can't thrive without a successful industry, and this thesis is no exception.

The quality and quantity of the data provided by the industrial partner are essential to achieve reliable and consistent research results. For the work of this thesis, the industrial partner provided large-volume data sets with enough variability to investigate the interesting phenomena. Historical industrial data, not shown in this thesis, is additionally provided by the industrial partner and helped to develop the methods.

Also, the university-industry collaboration meant more practical and academic knowledge were transferred in both directions.

### III) Combining methods in workflows

The main focus of having effective and successful mechanistic models in bioprocessing doesn't mean other methods are off the scene. On the contrary, this helped to understand what is missing for the mechanistic models, and by that helped to address the reasons behind that. This work has the implicit hypothesis that mechanistic models, although well established and used for decades now, still have room for improvement. Investigating and combining methods helped to have improved mechanistic modelling workflows.

### 3.1.2 Impact on Academic and Industrial Level

In this work, new methods to generate models with accurate predictions, lower uncertainty, robust parameters and representative structure are provided. This would raise the interest to develop such models in industry and academia to achieve better processes, and justify the extra efforts needed. The thesis impact can be measured on two levels.

On an academic level, this work highlights the importance of proper handling of discontinuities in bioprocesses. The comparative analyses shown in 2.2.4, 2.2.3 and 2.2.2 allows researchers to understand the influence of process-related and metabolic-related discontinuities on model parameters uncertainty and model predictions accuracy at a very detailed level. These comparisons allow to understand the two-sided relationship between the accuracy of locating metabolic switches and the accuracy and robustness of the model predictions. These analyses highlight an interdisciplinary understanding in cell physiology, numerical solutions, mathematical optimization, and data science to correctly cross-link resulting phenomena to their underlying reasons.

This thesis proposes a method and a workflow to properly handle discontinuities in a generic bioprocessing model, the method and the workflow are shown in 2.2.3.

This work also proposes a method to consider the off-gas singles for yeast cultivations in section 2.2.1, and a method to analyze the dissolved oxygen tension signal for *E. coli* cultivations in section 2.2.5. These two methods can help bioprocessing scientists to better understand the underlying metabolic activities in the cultivations.

The method in section 2.2.5 to analyze the dissolved oxygen signal revealed a new signal segment (characteristic) not reported before and likely to be linked to metabolic adaptation behavior.

On an industrial level, the results shown in section 2.2.5 for the modelling of *E. coli* growth and the rapidly changing dissolved oxygen levels in milliliter scale are of a great importance to address minibioreactor scale issues of oxygen supply and optimizing intermittent feeding plans.

The results shown in this section are for a state-of-the-art platform that is commercially established and recently deployed in industry.

The results shown in sections 2.2.4, 2.2.2 and 2.2.1 are expected to improve the calibration of a model commonly used in industrial context for yeast cultivations, and by that reducing the number of required runs for calibration, leading to cut in the costs and time.

Additionally, the thesis is expected to:

- Increase the use of piecewise mechanistic models.
- Help other researchers to transfer the two piecewise discontinuous models (Yeast and *E. coli*) described in 2.2.5 and 2.2.3 to other strains and/or organisms.
- Open potential door for direct metabolism-based-control strategies.
- Increase the awareness of proper selection of objective function criteria in bioprocessing models.

- The adaptation time described in section 2.2.5 can be used as strain selector to choose metabolic switching resilient strains to improve cultivation output and reduce inhibitory by-products.

### 3.1.3 Outlook

The results of the proposed methods, models, and workflows encourage further investigations on different organisms and platforms to evaluate their generic applicability on a wider set of variable conditions. For that the following work can be considered:

- I. Further in-vitro investigation on the physiology behind the adaptation state in *E. coli* in mini-bioreactor systems is needed to reveal more on this phenomenon on genetic, proteomic and metabolic level. It is also interesting to investigate the metabolic adaptation behavior of other organisms like yeast and mammalian cells.
- II. Better experimental implementation that overcomes the current practical limitation: for example, for mini-bioreactor cultivations, setting feeding plans in which either the pulses amplitude or the frequency change, not both. Also, considering sampling immediately after the substrate pulse. Such experimental implementation improvements would make the analysis more consistent and straightforward.
- III. Further model-based analysis to calculate the sensitivities, and to test different modelling concepts for the metabolic activities indicated by process analytics. This would help to investigate how much information on model parameters could the analytics, especially the used online signals, have.

## 4 Appendix

## 4.1 Discontinuities and ODE solvers

### 4.1.1 Details on the numerical solution

#### Boolean trigger functions

Events are triggered when the sign of the condition in (2.2), changes, i.e., a zero crossing is detected. The logical propositions may contain a number of relational expressions and sets of connectives (e.g., NOT, AND, OR). For  $\mathcal{C}(t)$ , output values true are mapped to positive values and output values false are mapped to negative values such that the corresponding condition has a root at every change of the corresponding Boolean values:

$$\mathcal{C}(t) := \begin{cases} 1, & \text{if } \text{cond}(c(t), u(t), \theta, t) \geq 0 \\ 0, & \text{if } \text{cond}(c(t), u(t), \theta, t) < 0 \end{cases} \quad (4.1)$$

Moreover, in 2.2.4, a decision tree is formulated out of the values of the Boolean functions which links different conditions, see Figure 10. Once one or more events are detected, the new active submodel is selected based on the evaluation of the decision tree.

#### Events location

##### I) Discontinuity locking

In the event driven method, the system of equations for each subinterval is locked throughout the solution. This means that the system of equations cannot change even if one or more state conditions are satisfied [101]. The conditions are monitored continuously, and if any of them are satisfied, the exact time of occurrence is then located, equations are switched, new initial states might be calculated, and the integration is restarted [118].

##### II) Chattering/Sticking control

In order to reduce chattering, in this contribution, a hysteresis band is defined for the threshold in each condition as follows:

$$\begin{aligned} \text{cond}(\text{cond}(t), u(t), \theta, t) &= -\epsilon & \text{if } \mathcal{C}(t) = 1 \\ \text{cond}(\text{cond}(t), u(t), \theta, t) &= +\epsilon & \text{if } \mathcal{C}(t) = 0 \end{aligned} \quad (4.2)$$

where the magnitude of  $\epsilon$  defines the magnitude of the hysteresis band. It can be seen that the sign of  $\epsilon$  depends on the current state of the Boolean trigger function  $\mathcal{C}(t)$ .  $\epsilon$  is a tuning parameter, its value needs to be chosen individually for the specific problem, keeping in mind that using small values avoids any delay effects in the event detection. The concept for event detection with hysteresis is illustrated in Figure 10.

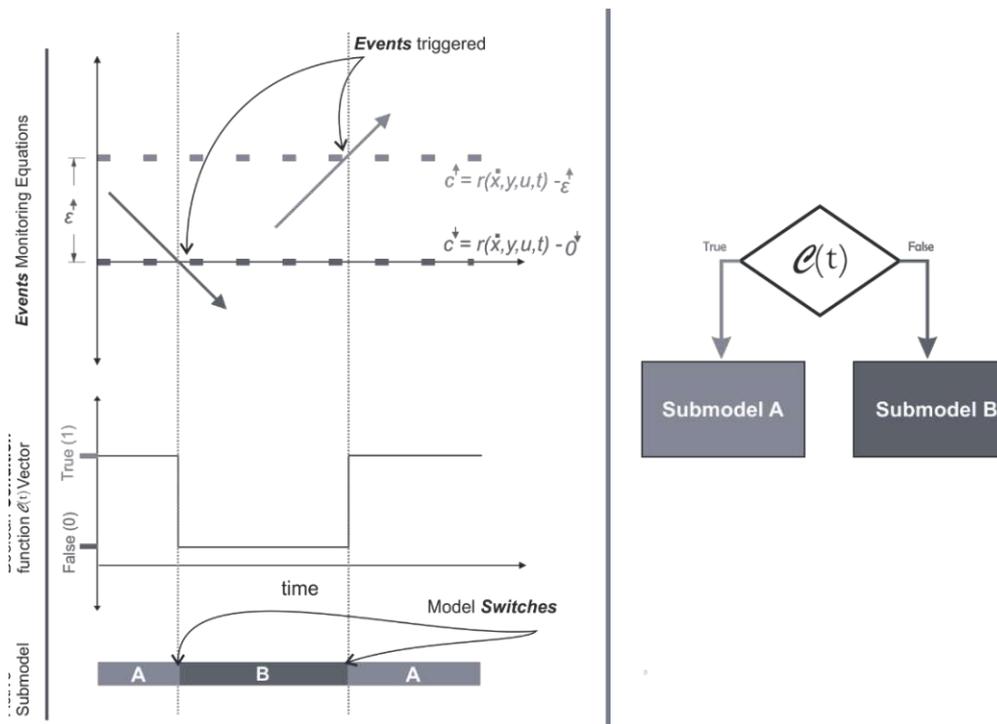


Figure 10: Event detection with hysteresis in order to avoid chattering/sticking behavior in the numerical solution of the ODE mechanistic models

## 4.1.2 State-of-the-art ODE solvers

solver	variable-step	order	method	stiffness
Ode23		2,3	pair of Bogacki and Shampine	moderate stiffness
Ode45		4,5	Dormand-Prince pair	non-stiff
Ode113	yes	1- 12	Adams-Bashforth-Moulton	moderate stiffness to non-stiff
Ode15s	yes	1- 5	NDFs-BDF/ Gear's method	stiff
Ode23s		2,3	modified Rosenbrock formula of order 2	stiff
Ode23t		2,3	trapezoidal rule	moderately stiff
Ode23tb		2,3	implicit Runge-Kutta formula with a trapezoidal rule	stiff

Table 2 main solvers used in MATLAB [83], [119], [120]

solver	variable-step	order	method	stiffness
RK45		4,5	Explicit Runge-Kutta method of order 5(4)	non-stiff
RK23		2,3	Explicit Runge-Kutta method of order 3(2)	non-stiff
DOP853		8	Explicit Runge-Kutta method of order 8	Non-stiff
Radau		5	Radau IIA family of order 5	stiff
BDF	yes	1-5	Implicit BDF multi-step variable-order (1 to 5) method	stiff
LSODA		varies	Adams/BDF method with automatic stiffness detection. Wrapper for ODEPACK in Fortran	

Table 3 main solvers used in Python [121]–[126]

solver	variable-step	order	method	stiffness
CV_BDF	yes	1-5	Implicit BDF linear multistep method	stiff
CV_ADAMS	yes	1-12	Adams-Moulton linear multistep method	non-stiff

Table 4 main solvers used in Modelica [127]

### 4.1.3 Mathematical discontinuities in yeast growth model

#### 4.1.3.1 Visual Representation

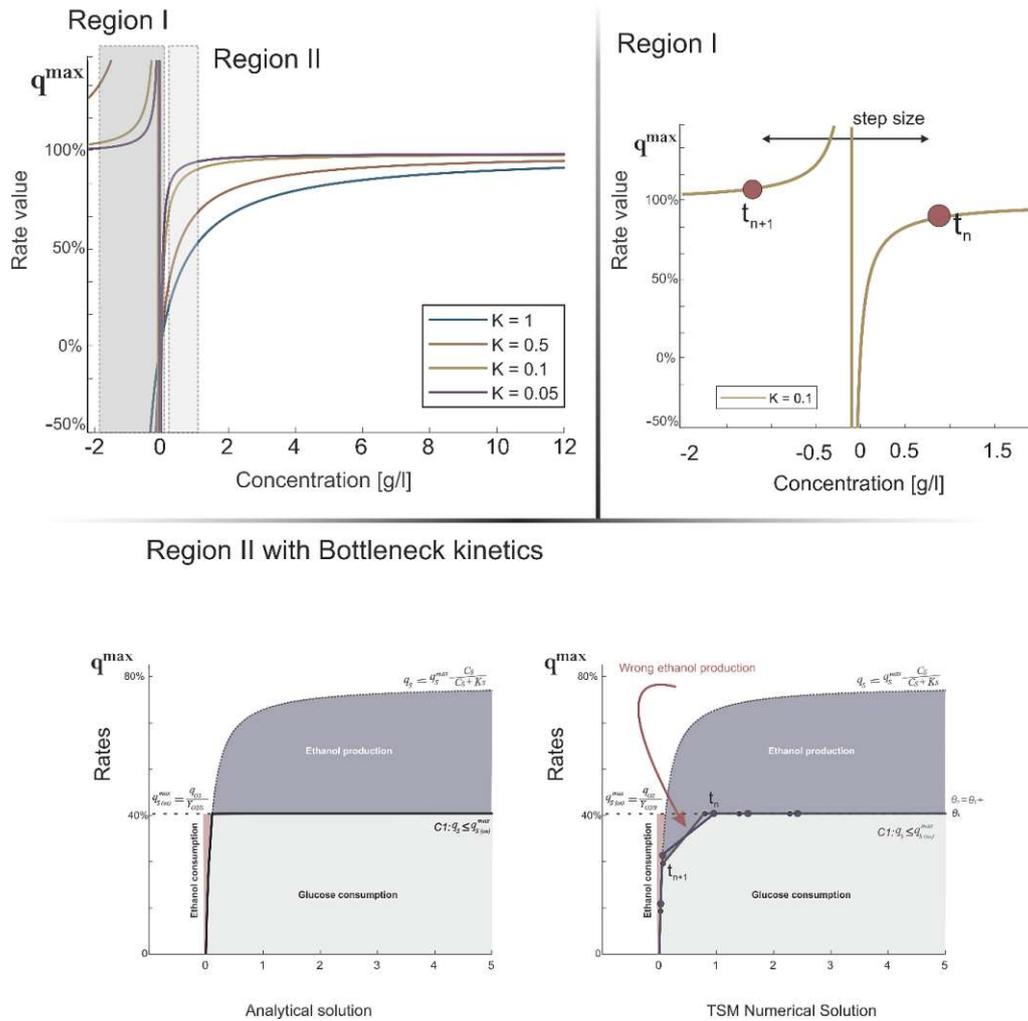


Figure 11: Common types of discontinuities happen in yeast cultivation model of [82]

### 4.1.3.2 Numerical quantification

Table 5: A batch is modelled with parameter set taken from section 0. The solver tolerances were systematically decreased starting from crude tolerances (default values with relative error tolerance  $RelTol = 10^{-3}$ , absolute error tolerance  $AbsTol = 10^{-3}$ ) [75], up to values of  $RelTol = AbsTol = 10^{-10}$ . Table 9 shows the influence of these values on the step size, the overstepping (measured as the smallest negative value) and on the computational cost (measured as the number of steps taken by the solver, i.e., all steps, failed and succeeded)

Method	Tolerance	Stiff solver								Non-stiff solver					
		Ode15s		Ode23s		Ode23t		Ode23tb		Ode45		Ode23		Ode113	
		smallest negative value	total steps	smallest negative value	total steps										
Time stepping	Crude tolerance	-16.223	180	-16.404	142	-16.2484	168	-16.228	129	$-1.105 \times 10^{-6}$	1728	$-1.695 \times 10^{-6}$	2411	$-8.862 \times 10^{-6}$	4202
	$10^{-5}$	-16.1863	363	$-1.814 \times 10^{-9}$	79	$-1.171 \times 10^{-7}$	691	$-5.507 \times 10^{-6}$	564	$-1.07 \times 10^{-6}$	1759	$-1.34 \times 10^{-6}$	2519	$-6.002 \times 10^{-6}$	4256
	$10^{-7}$	$-1.126 \times 10^{-6}$	878	$-5.752 \times 10^{-9}$	1429	$-1.255 \times 10^{-6}$	1730	$-1.82 \times 10^{-6}$	1390	$-1.067 \times 10^{-6}$	1770	$-1.445 \times 10^{-6}$	2779	$-8.256 \times 10^{-6}$	4414
	$10^{-10}$	$-5.009 \times 10^{-7}$	902	$-1.038 \times 10^{-9}$	1564	$-5.394 \times 10^{-7}$	1908	$-6.46 \times 10^{-8}$	1554	$-1.060 \times 10^{-6}$	1782	$-1.632 \times 10^{-6}$	2877	$-6.0887 \times 10^{-6}$	4341

### 4.1.4 Video explanation

A simplified video example of an optimization problem for EDM and TSM modelling results using MATLAB lsqnonlin function with a parameter estimation problem (for two parameters  $q_s$  &  $q_{02}$ ) can be found here:

[https://drive.google.com/drive/folders/1z\\_5TKBRF7BrmseI5tKAKTSxWZLCfSFYH](https://drive.google.com/drive/folders/1z_5TKBRF7BrmseI5tKAKTSxWZLCfSFYH)

## 4.2 Video contributions

A link to the video presentation of the contribution “Proper handling of metabolic and process-related discontinuities in the modelling of *S. cerevisiae* fed-batch cultures” in the VH Yeast Conference 2022 in Berlin, Germany

[https://www.vh-berlin.org/vhyc2022/videos/2.3.joune\\_d\\_lsizocbh.mp4](https://www.vh-berlin.org/vhyc2022/videos/2.3.joune_d_lsizocbh.mp4)

# 4.3 Posters

Adnan JOUNED <sup>a,b</sup> // Julian KAGER <sup>a,c</sup> // Judit AIZPURI <sup>b</sup> // Christoph HERWIG <sup>a,c</sup> // Tilman BARZ <sup>b</sup>  
<sup>a</sup> ECCE: TU Wien, Guspenhofer Straße 1a 10604, 1060 Wien, Austria  
<sup>b</sup> AIT Austrian Institute of Technology GmbH, Giefinggasse 2, 1210 Wien, Austria  
<sup>c</sup> CD Laboratory on Mechanistic and Physiological Methods for Improved Bioprocesses, TU Wien, Guspenhofer Straße 1a 10604, 1060 Wien, Austria  
 Adnan.jouned@ait.ac.at



## IMPROVING THE CALIBRATION OF KINETIC GROWTH MODELS USING DYNAMIC TIME WARPING

### Fitting off-gas signals in kinetic models, does LSQ criterion fit this task ?

Off-gas measurements give valuable information on the respiratory activity of organisms during fermentation processes. Standard algorithms for parameter estimation use the least-squares, LSQ does not necessarily yield off-gas representative model predictions and parameters, especially with rapid variability and corresponding sharp bends and kinks. A new Adapted algorithm using the Dynamic Time Warping (DTW) criterion is proposed and used. Both algorithms were used to fit experimental data of a kinetic model for *Saccharomyces cerevisiae* (Sonnleitner and Käppli 1986).



### Objective function 1: LSQ criterion (Standard)

$$\phi_{L+G}^{LSQ}(\theta) = \sum_{i=1}^M \sum_{j=1}^{N_L+N_G} (Y_{i,j}^m - Y_{i,j}(\theta))^2$$

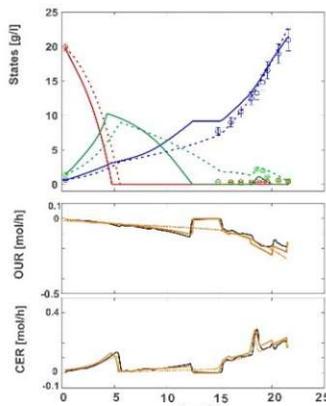
$$\min_{\theta} \phi_{L+G}^{LSQ}(\theta)$$

### Objective function 2: Combined LSQ and DTW criterion

$$\phi_G^{DTW} = \min_k \left( \sum_{i,j} Y_{i,j} \right)_k$$

$$\min_{\theta} \left( \phi_L^{LSQ}(\theta) + \alpha \cdot \phi_G^{DTW}(\theta) \right)$$

L:liquid components, G:gas components



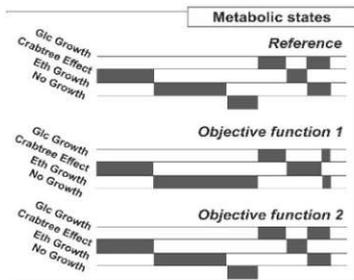
### Results and discussions

Comparison results showed difference in the metabolic states sequence. It can be seen that with standard LSQ criterion the optimizer overlooks some intermediate details in order to get a good fitting along the whole time horizon.

This is problematic, as different metabolic states are assigned along the process, where metabolic states are indicated by sharp changes in the off-gas signal.

The length of a metabolic state is strongly determined by the parameters related to maximum reaction rates and the corresponding conversion coefficients (yields), which needed to be correctly assigned during model parametrization.

By performing a statistical analysis of the model calibration, DTW/LSQ criterion had a twice higher success rate in finding the exact parameters out of 1000 model calibrations, each with differently perturbed initial parameter sets.

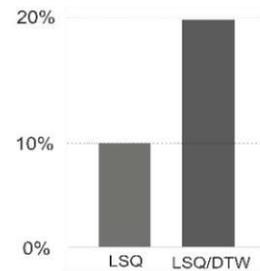


### Conclusion

Using DTW (Dynamic Time Warping) as a fitting criterion for the identification of *S. cerevisiae* fermentation models leads to more reliable parameters estimates compared with the standard LSQ fitting approach. This has been quantitatively proved for three experimental datasets with different metabolic states.

For the presented case study, it is shown that model predictions generated by LSQ fitting tend to smooth out measured off-gas signals losing specific details of the signals shape that might represent important metabolic changes.

### Successful Rate %



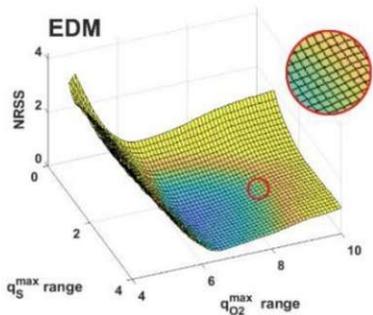
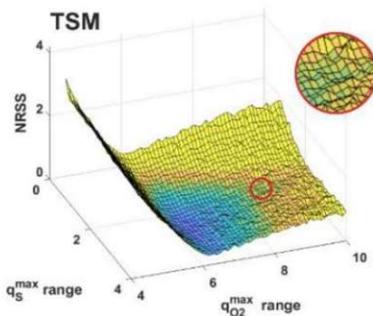
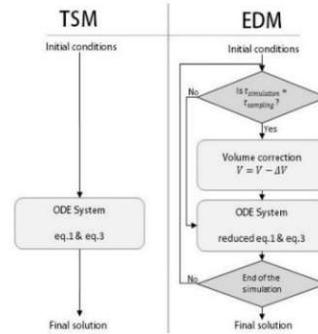
This work was partially funded by the Austrian Research Funding Association (FFG) within the program Bridge 1 in the project "AdaMo" (No. 864705).

Figure 12: Poster contribution in the 30th European Symposium on Computer-Aided Process Engineering (ESCAPE-30) 2020 in Milan, Italy

## EVENT DRIVEN ANALYSIS TO ENHANCE MODEL CALIBRATION OF EXPERIMENTS WITH HIGH OFFLINE SAMPLING RATES

### Does High Offline Sampling Rates in Pharmaceutical Bioprocessing Affect Mechanistic Models Calibration ?

The use of autosamplers connected to high throughput analytical devices allows for a high sampling frequency and analytics with reduced manual labor, leading to better process characterization (Maurer et al., 2015; Hofer et al., 2020). Increased sampling often leads to a significantly increased information content in the generated data. However, in combination with miniaturized or lab-scale reactors, the effect of volume change by the frequent sampling becomes challenging. How this volume can be correctly modelled ?



### Conclusion

Using EDM to consider sampling volume in *S. cerevisiae* fermentation models delivers more reproducible and accurate model predictions. Moreover, for model calibration, the quantitative analysis reveals a significantly improved convergence of the parameter estimation algorithm. Accordingly, EDM results are obtained with a lower prediction error and parameters uncertainty. The reason is a smoother optimization surface that leads to more robust convergence to the best fitting model parameters.

<sup>1</sup> This work was partially funded by the Austrian Research Funding Association (FFG) within the program Bridge 1 in the project "AdaMo" (No. 864705).

### Results and discussions

EDM (Event Driven Method) accurately locates all sampling events and uses an external algebraic equation to correct the volume and to restart the integration at each sampling timepoint. By doing this, it suppresses the noise and provides accurate and reproducible model simulations. This behavior is illustrated in the left graph.

The noisy nature of the surface in TSM (Time Stepping Method) reduces the effectiveness of the gradient-based optimizer to seek an optimum. The simulation results are shown below.

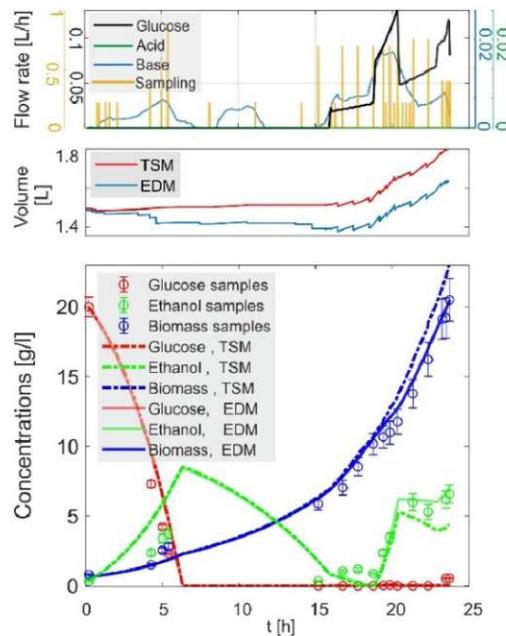


Figure 13: Poster contribution in the 31st European Symposium on Computer-Aided Process Engineering (ESCAPE-31) 2021 in Istanbul, Turkey

## 5 References

- [1] H. Narayanan *et al.*, “Bioprocessing in the Digital Age: The Role of Process Models,” *Biotechnol J*, vol. 15, no. 1, Jan. 2020, doi: 10.1002/biot.201900172.
- [2] Tanguy Catlin, Johannes-Tobias Lorenz, Bob Sternfels, and Paul Willmott, “A roadmap for a digital transformation | McKinsey,” 2017. Accessed: Dec. 30, 2022. [Online]. Available: <https://www.mckinsey.com/industries/financial-services/our-insights/a-roadmap-for-a-digital-transformation>
- [3] European Medicines Agency EMA and Food and Drug Administration FDA, “EMA-FDA pilot program for parallel assessment of Quality-by-Design applications: lessons learnt and Q&A resulting from the first parallel assessment,” Amsterdam , Aug. 2013. Accessed: Dec. 31, 2022. [Online]. Available: [www.ema.europa.eu](http://www.ema.europa.eu)
- [4] Food and Drug Administration, “Guidance for Industry PAT - A Framework for Innovative Pharmaceutical Development, manufacturing, and Quality Assurance,” Silver Spring, Maryland, 2004. Accessed: Dec. 31, 2022. [Online]. Available: <http://www.fda.gov/cvm/guidance/published.html>
- [5] P. Kleinebudde, J. Khinast, and J. Rantanen, *Continuous manufacturing of pharmaceuticals*. 2017. Accessed: Dec. 30, 2022. [Online]. Available: <https://www.wiley.com/en-us/Continuous+Manufacturing+of+Pharmaceuticals-p-9781119001324>
- [6] European Medicine Agency, “Mandate for Process Analytical Technology ,” Amsterdam, 2006. Accessed: Dec. 31, 2022. [Online]. Available: <http://www.emea.europa.eu>
- [7] A. S. Rathore, “Roadmap for implementation of quality by design (QbD) for biotechnology products,” *Trends Biotechnol*, vol. 27, no. 9, pp. 546–553, Sep. 2009, doi: 10.1016/J.TIBTECH.2009.06.006.
- [8] G. Faust, N. H. Janzen, C. Bendig, L. Römer, K. Kaufmann, and D. Weuster-Botz, “Feeding strategies enhance high cell density cultivation and protein expression in milliliter scale bioreactors,” *Biotechnol J*, vol. 9, no. 10, pp. 1293–1303, Oct. 2014, doi: 10.1002/biot.201400346.
- [9] J. I. Betts and F. Baganz, “Miniature bioreactors: Current practices and future opportunities,” *Microbial Cell Factories*, vol. 5, May 25, 2006. doi: 10.1186/1475-2859-5-21.
- [10] A. Silge, K. Weber, D. Cialla-May, L. Müller-Böttcher, D. Fischer, and J. Popp, “Trends in pharmaceutical analysis and quality control by modern Raman spectroscopic techniques,” *TrAC - Trends in Analytical Chemistry*, vol. 153, Aug. 2022, doi: 10.1016/j.trac.2022.116623.

- [11] A. G. Cardillo *et al.*, “Towards in silico Process Modeling for Vaccines,” *Trends Biotechnol*, vol. 39, no. 11, pp. 1120–1130, Nov. 2021, doi: 10.1016/j.tibtech.2021.02.004.
- [12] I. A. Udugama, P. C. Lopez, C. L. Gargalo, X. Li, C. Bayer, and K. v. Gernaey, “Digital Twin in biomanufacturing: challenges and opportunities towards its implementation,” *Systems Microbiology and Biomanufacturing*, vol. 1, no. 3, pp. 257–274, 2021, doi: 10.1007/s43393-021-00024-0.
- [13] Y. Chen, O. Yang, C. Sampat, P. Bhalode, R. Ramachandran, and M. Ierapetritou, “Digital twins in pharmaceutical and biopharmaceutical manufacturing: A literature review,” *Processes*, vol. 8, no. 9, Sep. 2020, doi: 10.3390/pr8091088.
- [14] P. Neubauer, N. Cruz, F. Glauche, S. Junne, A. Knepper, and M. Raven, “Consistent development of bioprocesses from microliter cultures to the industrial scale,” *Eng Life Sci*, vol. 13, no. 3, pp. 224–238, May 2013, doi: 10.1002/ELSC.201200021.
- [15] P. Satzer, D. Komuczki, M. Pappenreiter, A. L. Cataldo, B. Sissolak, and A. Jungbauer, “Impact of failure rates, lot definitions and scheduling of upstream processes on the productivity of continuous integrated bioprocesses,” *Journal of Chemical Technology & Biotechnology*, vol. 97, no. 9, pp. 2393–2403, Sep. 2022, doi: 10.1002/JCTB.6648.
- [16] I. Khanna, “Drug discovery in pharmaceutical industry: productivity challenges and trends,” *Drug Discov Today*, vol. 17, no. 19–20, pp. 1088–1102, Oct. 2012, doi: 10.1016/J.DRUDIS.2012.05.007.
- [17] C.-F. Mandenius, N. J. Titchener-Hooker, and A. (Anshuman) Bansal, “Measurement, monitoring, modelling and control of bioprocesses,” 2013.
- [18] G. Jifa, “Information Technology and Quantitative Management (ITQM2013) Data, Information, Knowledge, wisdom and meta-synthesis of wisdom-comment on wisdom global and wisdom cities,” *Procedia Comput Sci*, vol. 17, pp. 713–719, 2013, doi: 10.1016/j.procs.2013.05.092.
- [19] P. Kroll, P. Sagmeister, Wiel, Reichelt, L. Neutsch, and T. K. & C. Herwig, “Ex situ online monitoring: application, challenges and opportunities for biopharmaceuticals processes,” *Pharm Bioprocess*, vol. 2, no. 3, pp. 285–300, 2014, doi: 10.4155/PBP.14.22.
- [20] Shanley Agnes, “From the Editor: The DRIP (Data Rich, Information Poor) Syndrome | Pharma Manufacturing,” *Endeavor Business Media*, Nashville, 2005. Accessed: Dec. 31, 2022. [Online]. Available: <https://www.pharmamanufacturing.com/production/automation-control/article/11365560/from-the-editor-the-drip-data-rich-information-poor-syndrome>

- [21] R. M. C. Portela *et al.*, “When Is an In Silico Representation a Digital Twin? A Biopharmaceutical Industry Approach to the Digital Twin Concept,” *Adv Biochem Eng Biotechnol*, vol. 176, pp. 35–55, 2021, doi: 10.1007/10\_2020\_138.
- [22] C. Herwig, O. F. Garcia-Aponte, A. Golabgir, and A. S. Rathore, “Knowledge management in the QbD paradigm: manufacturing of biotech therapeutics,” *Trends Biotechnol*, vol. 33, no. 7, pp. 381–387, Jul. 2015, doi: 10.1016/J.TIBTECH.2015.04.004.
- [23] J. Smiatek, A. Jung, and E. Bluhmki, “Towards a Digital Bioprocess Replica: Computational Approaches in Biopharmaceutical Development and Manufacturing,” *Trends Biotechnol*, vol. 38, no. 10, pp. 1141–1153, Oct. 2020, doi: 10.1016/j.tibtech.2020.05.008.
- [24] D. A. Suarez-Zuluaga, D. Borchert, N. N. Driessen, W. A. M. Bakker, and Y. E. Thomassen, “Accelerating bioprocess development by analysis of all available data: A USP case study,” *Vaccine*, vol. 37, no. 47, pp. 7081–7089, Nov. 2019, doi: 10.1016/J.VACCINE.2019.07.026.
- [25] M. Sokolov, M. Morbidelli, A. Butté, J. Souquet, and H. Broly, “Sequential Multivariate Cell Culture Modeling at Multiple Scales Supports Systematic Shaping of a Monoclonal Antibody Toward a Quality Target,” *Biotechnol J*, vol. 13, no. 4, Apr. 2018, doi: 10.1002/BIOT.201700461.
- [26] M. Sokolov *et al.*, “Enhanced process understanding and multivariate prediction of the relationship between cell culture process and monoclonal antibody quality,” *Biotechnol Prog*, vol. 33, no. 5, pp. 1368–1380, Sep. 2017, doi: 10.1002/BTPR.2502.
- [27] J. C. Gunther, J. Baclaski, D. E. Seborg, and J. S. Conner, “Pattern matching in batch bioprocesses-Comparisons across multiple products and operating conditions,” *Comput Chem Eng*, vol. 33, no. 1, pp. 88–96, Jan. 2009, doi: 10.1016/J.COMPCHEMENG.2008.07.001.
- [28] Cosima Koch, “Advances in Bioprocess Monitoring by Mid-Infrared Spectroscopy,” *TUW*, 2015.
- [29] M. Sokolov, F. Feidl, M. Morbidelli, and A. Butté, “Big data in biopharmaceutical process development vice or virtue?,” *Chimica Oggi/Chemistry Today*, vol. 36, no. 6, pp. 26–29, 2018.
- [30] P. M. Doran, *Bioprocess engineering principles: Second edition*, vol. 9780080917. Academic Press, 2018.
- [31] A. S. Rathore and H. Winkle, “Quality by design for biopharmaceuticals,” *Nature Biotechnology*, vol. 27, no. 1. Nature Publishing Group, pp. 26–34, Jan. 25, 2009. doi: 10.1038/nbt0109-26.
- [32] J. Nielsen, *Physiological Engineering Aspects of Penicillium Chrysogenum*. WORLD SCIENTIFIC, 1997. doi: 10.1142/3195.

- [33] K. V. Gernaey, A. E. Lantz, P. Tufvesson, J. M. Woodley, and G. Sin, “Application of mechanistic models to fermentation and biocatalysis for next-generation processes,” *Trends Biotechnol*, vol. 28, no. 7, pp. 346–354, 2010, doi: 10.1016/j.tibtech.2010.03.006.
- [34] I. Gosling, “Process simulation and modeling for industrial bioprocessing: Tools and techniques,” *Industrial Biotechnology*, vol. 1, no. 2, pp. 106–109, Jun. 2005, doi: 10.1089/ind.2005.1.106.
- [35] A. Toumi, C. Jürgens, C. Jungo, B. A. Maier, V. Papavasileiou, and D. Petrides, “Design and optimization of a large scale biopharmaceutical facility using process simulation and scheduling tools,” *Pharmaceutical Engineering*, vol. 30, pp. 22+24+26+28+30+32+34-36, Mar. 2010.
- [36] J. E. Jiménez-Hornero, I. M. Santos-Dueñas, and I. García-García, “Optimization of biotechnological processes. The acetic acid fermentation. Part I: The proposed model,” *Biochem Eng J*, vol. 45, no. 1, pp. 1–6, Jun. 2009, doi: 10.1016/j.bej.2009.01.009.
- [37] V. Abt *et al.*, “Model-based tools for optimal experiments in bioprocess engineering,” *Curr Opin Chem Eng*, vol. 22, pp. 244–252, 2018, doi: 10.1016/j.coche.2018.11.007.
- [38] M. N. Cruz Bournazou *et al.*, “Online optimal experimental re-design in robotic parallel fed-batch cultivation facilities,” *Biotechnol Bioeng*, vol. 114, no. 3, pp. 610–619, Mar. 2017, doi: 10.1002/bit.26192.
- [39] H. Narayanan *et al.*, “Bioprocessing in the Digital Age: The Role of Process Models,” *Biotechnol J*, vol. 15, no. 1, p. 1900172, Jan. 2020, doi: 10.1002/biot.201900172.
- [40] D. Dochain, “State and parameter estimation in chemical and biochemical processes: A tutorial,” *J Process Control*, vol. 13, no. 8, pp. 801–818, Dec. 2003, doi: 10.1016/S0959-1524(03)00026-X.
- [41] C. Dietzsch, O. Spadiut, and C. Herwig, “On-line multiple component analysis for efficient quantitative bioprocess development,” *J Biotechnol*, vol. 163, no. 4, pp. 362–370, Feb. 2013, doi: 10.1016/j.jbiotec.2012.03.010.
- [42] S. Natarajan and J. H. Lee, “Repetitive model predictive control applied to a simulated moving bed chromatography system,” in *Computers and Chemical Engineering*, Jul. 2000, vol. 24, no. 2–7, pp. 1127–1133. doi: 10.1016/S0098-1354(00)00493-2.
- [43] R. Singh, K. V. Gernaey, and R. Gani, “Model-based computer-aided framework for design of process monitoring and analysis systems,” *Comput Chem Eng*, vol. 33, no. 1, pp. 22–42, Jan. 2009, doi: 10.1016/j.compchemeng.2008.06.002.
- [44] M. Degerman, K. Westerberg, and B. Nilsson, “Determining critical process parameters and process robustness in preparative chromatography - A model-based approach,” *Chem Eng Technol*, vol. 32, no. 6, pp. 903–911, Jun. 2009, doi: 10.1002/ceat.200900019.

- [45] A. Meitz, P. Sagmeister, W. Lubitz, C. Herwig, and T. Langemann, “Fed-Batch Production of Bacterial Ghosts Using Dielectric Spectroscopy for Dynamic Process Control,” *Microorganisms*, vol. 4, no. 2, p. 18, Mar. 2016, doi: 10.3390/microorganisms4020018.
- [46] P. Czop, G. Kost, D. Slawik, and G. Wszolek, “Formulation and identification of First- Principle Data-Driven models,” *Journal of achievements in materials and manufacturing engineering*, 2011.
- [47] J. J. DiStefano, *Dynamic Systems Biology Modeling and Simulation*. Academic Press, 2015.
- [48] A. K. Gombert and J. Nielsen, “Mathematical modelling of metabolism,” *Curr Opin Biotechnol*, vol. 11, no. 2, pp. 180–186, Apr. 2000, doi: 10.1016/S0958-1669(00)00079-3.
- [49] O. D. Kim, M. Rocha, and P. Maia, “A Review of Dynamic Modeling Approaches and Their Application in Computational Strain Optimization for Metabolic Engineering,” *Front Microbiol*, vol. 9, Jul. 2018, doi: 10.3389/fmicb.2018.01690.
- [50] S.-O. Enfors, *Fermentation Process Technology*. Stockholm: Royal Institute of Technology (KTH), 2019.
- [51] G. Sin, P. Ödman, N. Petersen, A. E. Lantz, and K. v. Gernaey, “Matrix notation for efficient development of first-principles models within PAT applications: Integrated modeling of antibiotic production with *Streptomyces coelicolor*,” *Biotechnol Bioeng*, vol. 101, no. 1, pp. 153–171, Sep. 2008, doi: 10.1002/BIT.21869.
- [52] Isabel Cristina de Almeida Pereira da Rocha, *MODEL-BASED STRATEGIES FOR COMPUTER-AIDED OPERATION OF A RECOMBINANT E. COLI FERMENTATION Braga, 2003 MODEL-BASED STRATEGIES FOR COMPUTER-AIDED OPERATION OF A RECOMBINANT E. COLI FERMENTATION*. Braga: Escola de Engenharia Universidade do Minho, 2003.
- [53] M. Svensson, “The temperature-limited fed-batch technique for control of *Escherichia coli* cultures,” 2006.
- [54] K. Adamberg, S. Kask, T. M. Laht, and T. Paalme, “The effect of temperature and pH on the growth of lactic acid bacteria: A pH-auxostat study,” *Int J Food Microbiol*, vol. 85, no. 1–2, pp. 171–183, Aug. 2003, doi: 10.1016/S0168-1605(02)00537-8.
- [55] M. Svensson and S.-O. Enfors, “Modeling of temperatur-dependent growth of *Escherichia coli* in a PH-auxostat and temperatur-limited fed-batch culture,” 2006, Accessed: Jan. 03, 2023. [Online]. Available: <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-6270>
- [56] M. Muloiwa, S. Nyende-Byakika, and M. Dinka, “Comparison of unstructured kinetic bacterial growth models.,” *S Afr J Chem Eng*, vol. 33, pp. 141–150, Jul. 2020, doi: 10.1016/J.SAJCE.2020.07.006.

- [57] B. Srinivasan, “A guide to enzyme kinetics in early drug discovery,” *FEBS J*, 2022, doi: 10.1111/FEBS.16404.
- [58] J. Monod, “The Growth of Bacterial Cultures,” *Annu Rev Microbiol*, vol. 3, no. 1, pp. 371–394, Oct. 1949, doi: 10.1146/annurev.mi.03.100149.002103.
- [59] H. Moser, “The dynamics of bacterial populations maintained in the chemostat.,” *The dynamics of bacterial populations maintained in the chemostat.*, 1958.
- [60] F. F. Blackman, “Optima and limiting factors,” *Ann Bot*, vol. os-19, no. 2, pp. 281–296, Apr. 1905, doi: 10.1093/oxfordjournals.aob.a089000.
- [61] M. L. Shuler and J. D. Varner, “Cell growth dynamics,” in *Comprehensive Biotechnology*, Pergamon, 2011, pp. 32–38. doi: 10.1016/B978-0-444-64046-8.00065-3.
- [62] N. J. Stanford, T. Lubitz, K. Smallbone, E. Klipp, P. Mendes, and W. Liebermeister, “Systematic construction of kinetic models from genome-scale metabolic networks,” *PLoS One*, vol. 8, no. 11, p. e79195, Nov. 2013, doi: 10.1371/journal.pone.0079195.
- [63] S. Ulonska, D. Waldschitz, J. Kager, and C. Herwig, “Model predictive control in comparison to elemental balance control in an E. coli fed-batch,” *Chem Eng Sci*, vol. 191, pp. 459–467, Dec. 2018, doi: 10.1016/J.CES.2018.06.074.
- [64] M. N. Cruz Bournazou *et al.*, “Online optimal experimental re-design in robotic parallel fed-batch cultivation facilities,” *Biotechnol Bioeng*, vol. 114, no. 3, pp. 610–619, Mar. 2017, doi: 10.1002/BIT.26192.
- [65] E. A. del Rio-Chanona, D. Zhang, and V. S. Vassiliadis, “Model-based real-time optimisation of a fed-batch cyanobacterial hydrogen production process using economic model predictive control strategy,” *Chem Eng Sci*, vol. 142, pp. 289–298, Mar. 2016, doi: 10.1016/J.CES.2015.11.043.
- [66] IEEE, *IEEE Standard 610.12-1990 Glossary of Software Engineering Terminology (Reaffirmed 2002)*, vol. 121990, no. 1. 2002. doi: 10.1109/IEEESTD.1990.101064.
- [67] S. Craven, N. Shirsat, J. Whelan, and B. Glennon, “Process model comparison and transferability across bioreactor scales and modes of operation for a mammalian cell bioprocess,” *Biotechnol Prog*, vol. 29, no. 1, pp. 186–196, Jan. 2013, doi: 10.1002/BTPR.1664.
- [68] E. Anane, A. Sawatzki, P. Neubauer, and M. N. Cruz-Bournazou, “Modelling concentration gradients in fed-batch cultivations of E. coli – towards the flexible design of scale-down experiments,” *Journal of Chemical Technology and Biotechnology*, vol. 94, no. 2, pp. 516–526, 2019, doi: 10.1002/jctb.5798.
- [69] B. J. Sánchez, D. C. Soto, H. Jorquera, C. A. Gelmi, and J. R. Pérez-Correa, “HIPPO: An iterative reparametrization method for identification and calibration of dynamic bioreactor models of complex processes,” *Ind Eng Chem Res*, vol. 53, no. 48, pp. 18514–18525, Dec. 2014, doi: 10.1021/IE501298B/SUPPL\_FILE/IE501298B\_SI\_001.PDF.

- [70] S. Hoops *et al.*, “COPASI—a COmplex PAthway SIMulator,” *Bioinformatics*, vol. 22, no. 24, pp. 3067–3074, Dec. 2006, doi: 10.1093/BIOINFORMATICS/BTL485.
- [71] O. Chiş, J. R. Banga, and E. Balsa-Canto, “GenSSI: a software toolbox for structural identifiability analysis of biological models,” *Bioinformatics*, vol. 27, no. 18, p. 2610, Sep. 2011, doi: 10.1093/BIOINFORMATICS/BTR431.
- [72] J. A. Egea *et al.*, “MEIGO: An open-source software suite based on metaheuristics for global optimization in systems biology and bioinformatics,” *BMC Bioinformatics*, vol. 15, no. 1, pp. 1–9, May 2014, doi: 10.1186/1471-2105-15-136/FIGURES/8.
- [73] T. Hirmajer, E. Balsa-Canto, and J. R. Banga, “DOTcvpSB, a software toolbox for dynamic optimization in systems biology,” *BMC Bioinformatics*, vol. 10, no. 1, pp. 1–14, Jun. 2009, doi: 10.1186/1471-2105-10-199/FIGURES/8.
- [74] T. M. Alsoudani, *Discontinuities in Mathematical Modelling: Origin, Detection and Resolution*, no. March. 2016.
- [75] I. Gladwell, L. F. ; Shampine, and S. Thompson, *Solving ODEs with MATLAB*. Cambridge University Press, 2003.
- [76] D. R. Reynolds, A. C. Hindmarsh, and R. Serban, “User Documentation for ccode v4.1.0 ( sundials v4.1.0),” vol. 0, 2019.
- [77] K. Radhakrishnan and A. C. Hindmarsh, “Description and Use of LSODE , the Livermore Solver for Ordinary Differential Equations,” 1993.
- [78] A. Cuoci, A. Frassoldati, T. Faravelli, and E. Ranzi, “OpenSMOKE++: An object-oriented framework for the numerical modeling of reactive systems with detailed kinetic mechanisms,” *Comput Phys Commun*, 2015, doi: 10.1016/j.cpc.2015.02.014.
- [79] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*.
- [80] R. L. Burden and J. D. Faires, *Numerical Analysis*, Ninth Edt. Boston,: Richard Stratton, 2011.
- [81] L. Shampine *et al.*, “A User ’ s View of Solving Stiff Ordinary Differential Equations To cite this version : HAL Id : hal-01711390,” 2018.
- [82] B. Sonnleitner and O. Käppeli, “Growth of *Saccharomyces cerevisiae* is controlled by its limited respiratory capacity: Formulation and verification of a hypothesis,” *Biotechnol Bioeng*, vol. 28, no. 6, pp. 927–937, 1986, doi: 10.1002/bit.260280620.
- [83] L. F. Shampine and M. W. Reichelt, “Ode\_Suite,” *J Sci Comput*, vol. 18, pp. 1–22, 1997, doi: 10.1137/S1064827594276424.
- [84] P. Noll and M. Henkel, “History and Evolution of Modeling in Biotechnology: Modeling & Simulation, Application and Hardware Performance,” *Comput Struct Biotechnol J*, vol. 18, pp. 3309–3323, Jan. 2020, doi: 10.1016/J.CSBJ.2020.10.018.
- [85] Niels Krausch *et al.*, “High-throughput screening of optimal process conditions using model predictive control,” 2022, doi: 10.22541/au.164600498.85996662/v2.

- [86] P. Marshall, T. Szikszai, V. Lemay, and A. Kozak, “Testing the distributional assumptions of least squares linear regression,” *The forestry chronicle*, vol. 71, no. 2, pp. 213–2018, 1995.
- [87] C. Ratanamahatana and E. Keogh, “Everything you know about dynamic time warping is wrong,” *Third Workshop on Mining Temporal and Sequential Data*, no. January 2004, pp. 22–25, 2004, doi: 10.1097/01.CCM.0000279204.24648.44.
- [88] D. C. López, T. Barz, M. Peñuela, A. Villegas, S. Ochoa, and G. Wozny, “Model-based identifiable parameter determination applied to a simultaneous saccharification and fermentation process model for bio-ethanol production,” *Biotechnol Prog*, vol. 29, no. 4, pp. 1064–1082, 2013, doi: 10.1002/btpr.1753.
- [89] Bradley. Efron and Robert. Tibshirani, *An introduction to the bootstrap*. CRC Press, 1994.
- [90] E. Anane *et al.*, “A model-based framework for parallel scale-down fed-batch cultivations in mini-bioreactors for accelerated phenotyping,” *Biotechnol Bioeng*, vol. 116, no. 11, pp. 2906–2918, 2019, doi: 10.1002/bit.27116.
- [91] R. Hortsch and D. Weuster-Botz, “Milliliter-scale stirred tank reactors for the cultivation of microorganisms,” in *Advances in Applied Microbiology*, vol. 73, no. C, Academic Press Inc., 2010, pp. 61–82. doi: 10.1016/S0065-2164(10)73003-3.
- [92] T. v Kirk and N. Szita, “Oxygen Transfer Characteristics of Miniaturized Bioreactor Systems,” *Biotechnol. Bioeng*, vol. 110, pp. 1005–1019, 2013, doi: 10.1002/bit.24824/abstract.
- [93] N. H. Janzen, G. Striedner, J. Jarmer, M. Voigtmann, S. Abad, and D. Reinisch, “Implementation of a Fully Automated Microbial Cultivation Platform for Strain and Process Screening,” *Biotechnol J*, vol. 14, no. 10, p. 1800625, Oct. 2019, doi: 10.1002/biot.201800625.
- [94] F. Fröhlich, F. J. Theis, J. O. Rädler, and J. Hasenauer, “Parameter estimation for dynamical systems with discrete events and logical operations,” *Bioinformatics*, vol. 33, no. 7, pp. 1049–1056, Dec. 2017, doi: 10.1093/bioinformatics/btw764.
- [95] H. G. Crabtree, “Observations on the carbohydrate metabolism of tumours,” *Biochemical Journal*, vol. 23, no. 3, pp. 536–545, Jan. 1929, doi: 10.1042/bj0230536.
- [96] B. Xu, M. Jahic, and S. O. Enfors, “Modeling of overflow metabolism in batch and fed-batch cultures of *Escherichia coli*,” *Biotechnol Prog*, vol. 15, no. 1, pp. 81–90, 1999, doi: 10.1021/bp9801087.
- [97] L. Dieci and L. Lopez, “A survey of numerical methods for IVPs of ODEs with discontinuous right-hand side,” *J Comput Appl Math*, vol. 236, no. 16, pp. 3967–3991, 2012, doi: 10.1016/j.cam.2012.02.011.

- [98] C. C. Pantelides and P. I. Barton, “Equation-oriented dynamic simulation current status and future perspectives,” *Comput Chem Eng*, vol. 17, pp. S263–S285, 1993, doi: 10.1016/0098-1354(93)80240-N.
- [99] “scipy.integrate.solve\_ivp — SciPy v1.10.0 Manual,” 2023. [https://docs.scipy.org/doc/scipy/reference/generated/scipy.integrate.solve\\_ivp.html](https://docs.scipy.org/doc/scipy/reference/generated/scipy.integrate.solve_ivp.html) (accessed Jan. 09, 2023).
- [100] A. C. Hindmarsh *et al.*, “SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers,” *ACM Transactions on Mathematical Software*, vol. 31, no. 3, pp. 363–396, 2005, doi: 10.1145/1089014.1089020.
- [101] T. Park and P. I. Barton, “State event location in differential-algebraic models,” *ACM Transactions on Modeling and Computer Simulation*, vol. 6, no. 2, pp. 137–165, 1996, doi: 10.1145/232807.232809.
- [102] P. I. Barton and C. K. Lee, “Modeling, simulation, sensitivity analysis, and optimization of hybrid systems,” *ACM Transactions on Modeling and Computer Simulation*, vol. 12, no. 4, pp. 256–289, Oct. 2002. doi: 10.1145/643120.643122.
- [103] E. Fredriksson, C. Andersson, and J. Åkesson, “Discontinuities handled with events in Assimulo,” *Proceedings of the 10th International Modelica Conference, March 10-12, 2014, Lund, Sweden*, vol. 96, pp. 827–836, 2014, doi: 10.3384/ecp14096827.
- [104] K. Soetaert, T. Petzoldt, D. Germany, and R. W. Setzer, “Package deSolve: Solving Initial Value Differential Equations in R.”
- [105] M. Sofroniou and R. Knapp, “Advanced Numerical Differential Equation Solving in Mathematica,” ... *Inc. URL <http://reference.wolfram.com/mathematica/> ...*, pp. 1–372, 2008.
- [106] S. Kowalewski *et al.*, “A Case Study in Tool-Aided Analysis of Discretely Controlled Continuous Systems: The Two Tanks Problem,” Springer, Berlin, Heidelberg, 1999, pp. 163–185. doi: 10.1007/3-540-49163-5\_9.
- [107] P. J. Mosterman, “An overview of hybrid simulation phenomena and their support by simulation packages,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1999, vol. 1569, pp. 165–177. doi: 10.1007/3-540-48983-5\_17.
- [108] H. Lundvall, P. Fritzson, and B. Bachmann, “Event Handling in the OpenModelica Compiler and Runtime System.”
- [109] S. B. Petkov and R. A. Davis, “On-line biomass estimation using a modified oxygen utilization rate,” *Bioprocess Engineering*, vol. 15, no. 1, pp. 43–45, Jun. 1996, doi: 10.1007/BF00435527.
- [110] T. Anderlei, W. Zang, M. Papaspyrou, and J. Büchs, “Online respiration activity measurement (OTR, CTR, RQ) in shake flasks,” in *Biochemical Engineering Journal*, 2004, vol. 17, no. 3, pp. 187–194. doi: 10.1016/S1369-703X(03)00181-5.

- [111] K. Gollmer and C. Posten, “Supervision of bioprocesses using a dynamic time warping algorithm,” *Control Eng Pract*, vol. 4, no. 9, pp. 1287–1295, 1996, doi: 10.1016/0967-0661(96)00136-0.
- [112] S. Sunya, N. Gorret, F. Delvigne, J. L. UribeArrea, and C. Molina-Jouve, “Real-time monitoring of metabolic shift and transcriptional induction of *yciG*:: LuxCDABE E. Coli reporter strain to a glucose pulse of different concentrations,” *J Biotechnol*, vol. 157, no. 3, pp. 379–390, 2012, doi: 10.1016/j.jbiotec.2011.12.009.
- [113] S. Sunya, F. Delvigne, J. L. UribeArrea, C. Molina-Jouve, and N. Gorret, “Comparison of the transient responses of *Escherichia coli* to a glucose pulse of various intensities,” *Appl Microbiol Biotechnol*, vol. 95, no. 4, pp. 1021–1034, 2012, doi: 10.1007/s00253-012-3938-y.
- [114] B. Enjalbert, M. Coccagn-Bousquet, J. C. Portais, and F. Letisse, “Acetate exposure determines the diauxic behavior of *Escherichia coli* during the glucose-acetate transition,” *J Bacteriol*, vol. 197, no. 19, pp. 3173–3181, 2015, doi: 10.1128/JB.00128-15.
- [115] B. Enjalbert, F. Letisse, and J.-C. Portais, “Physiological and Molecular Timing of the Glucose to Acetate Transition in *Escherichia coli*,” *Metabolites*, vol. 3, no. 3, pp. 820–837, 2013, doi: 10.3390/metabo3030820.
- [116] H. Taymaz-Nikerel, W. M. van Gulik, and J. J. Heijnen, “*Escherichia coli* responds with a rapid and large change in growth rate upon a shift from glucose-limited to glucose-excess conditions,” *Metab Eng*, vol. 13, no. 3, pp. 307–318, 2011, doi: 10.1016/j.ymben.2011.03.003.
- [117] H. Y. Lin, B. Mathiszik, B. Xu, S. O. Enfors, and P. Neubauer, “Determination of the maximum specific uptake capacities for glucose and oxygen in glucose-limited fed-batch cultivations of *Escherichia coli*,” *Biotechnol Bioeng*, vol. 73, no. 5, pp. 347–357, Jun. 2001, doi: 10.1002/bit.1068.
- [118] F. E. Cellier, “Combined continuous/discrete system simulation by use of digital computers techniques and tools,” 1979. doi: 10.3929/ethz-a-010782581.
- [119] R. May, “The Numerical Solution of Ordinary Differential Equations: Initial Value Problems,” *North-Holland Mathematics Studies*, vol. 83, pp. 1–94, Jan. 1984, doi: 10.1016/S0304-0208(08)71200-3.
- [120] L. F. Shampine and S. N. Laboratories, “Implementation of Rosenbrock Methods,” vol. 8, no. 2, pp. 93–113, 1982.
- [121] “ODEPACK | Computing,” Livermore, 2023. Accessed: Jan. 04, 2023. [Online]. Available: <https://computing.llnl.gov/projects/odepack>
- [122] E. Hairer and G. Wanner, “Solving Ordinary Differential Equations II,” vol. 14, 1991, doi: 10.1007/978-3-662-09947-6.

- [123] P. Bogacki and L. F. Shampine, “A 3(2) pair of Runge - Kutta formulas,” *Appl Math Lett*, vol. 2, no. 4, pp. 321–325, 1989, doi: 10.1016/0893-9659(89)90079-7.
- [124] J. R. Dormand and P. J. Prince, “A family of embedded Runge-Kutta formulae,” *J Comput Appl Math*, vol. 6, no. 1, pp. 19–26, Mar. 1980, doi: 10.1016/0771-050X(80)90013-3.
- [125] A. C. Hindmarsh, “ODEPACK, A Systematized Collection of ODE Solvers,” *IMACS Transactions on Scientific Computation*, no. 1, pp. 55–64, 1983, Accessed: Jan. 04, 2023. [Online]. Available: [https://computing.llnl.gov/sites/default/files/ODEPACK\\_pub1\\_u88007.pdf](https://computing.llnl.gov/sites/default/files/ODEPACK_pub1_u88007.pdf)
- [126] L. Petzold, “Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations,” <https://doi.org/10.1137/0904010>, vol. 4, no. 1, pp. 136–148, Jul. 2006, doi: 10.1137/0904010.
- [127] “Solving Modelica Models — OpenModelica User’s Guide v1.20.0 documentation.” <https://www.openmodelica.org/doc/OpenModelicaUsersGuide/1.20/solving.html#hairer-1993> (accessed Jan. 04, 2023).