



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

M. Adnan Jouned^{a,b}, Julian Kager^c, Christoph Herwig^{a,c}, Tilman Barz^{b,*}

^a ICEBE, TU Wien, Gumpendorfer Straße 1a 166/4, 1060 Vienna, Austria

^b Center for Energy, AIT Austrian Institute of Technology GmbH, Giefinggasse 2, 1210 Vienna, Austria

^c 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 (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 θ as unknown parameter vector, the unconstrained and unbounded minimization problem reads:

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

Parameter initial guesses and estimates are given in Table 4. CO₂ and O₂ content in the off-gas was obtained from online gas analyzer and

Table 1
List of symbols.

Term	Description	Unit
α_1	Consumed O ₂ for oxidative growth on glucose	mol/mol
α_{10}	Produced biomass for oxidative growth on ethanol	mol/mol
α_{11}	CO ₂ yield for oxidative growth on ethanol	mol/mol
α_{12}	H ₂ O yield for oxidative growth on ethanol	mol/mol
α_2	Produced biomass for oxidative growth on glucose	mol/mol
α_3	CO ₂ yield for oxidative growth on glucose	mol/mol
α_4	H ₂ O yield for oxidative growth on glucose	mol/mol
α_5	Produced biomass for fermentative growth on glucose	mol/mol
α_6	CO ₂ yield for fermentative growth on glucose	mol/mol
α_7	H ₂ O yield for fermentative growth on glucose	mol/mol
α_8	Ethanol yield for fermentative growth on glucose	mol/mol
α_9	Consumed O ₂ 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 ⁻¹
F_{Base}	Base feed rate	L.h ⁻¹
F_{gas}	Gas in/out flow	L.h ⁻¹
F_S	Glucose feed rate	L.h ⁻¹
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
μ_{total}	Total growth rate	h ⁻¹
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 ⁻¹ .g ⁻¹
q_s	Specific glucose uptake rate per unit of biomass	g.h ⁻¹ .g ⁻¹
q_e	Specific ethanol uptake rate per unit of biomass	g.h ⁻¹ .g ⁻¹
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
$Y_{x/(e)(ox)}$	Biomass yield for oxidative growth on ethanol	g/g
$Y_{x/(s)(ox)}$	Biomass yield for oxidative growth on glucose	g/g
$Y_{x/(s)(red)}$	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 $\tilde{S} = U\Sigma V$, Σ matrix is found which holds the singular values of \tilde{S} . The singular values in Σ are then used to calculate: 1- the condition number (κ) which is a measure of the sensitivity of model

results to the perturbation of the parameters. 2- the collinearity index (γ) which quantifies the collinearity of the parameters. Empirical values for thresholds of κ and γ 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

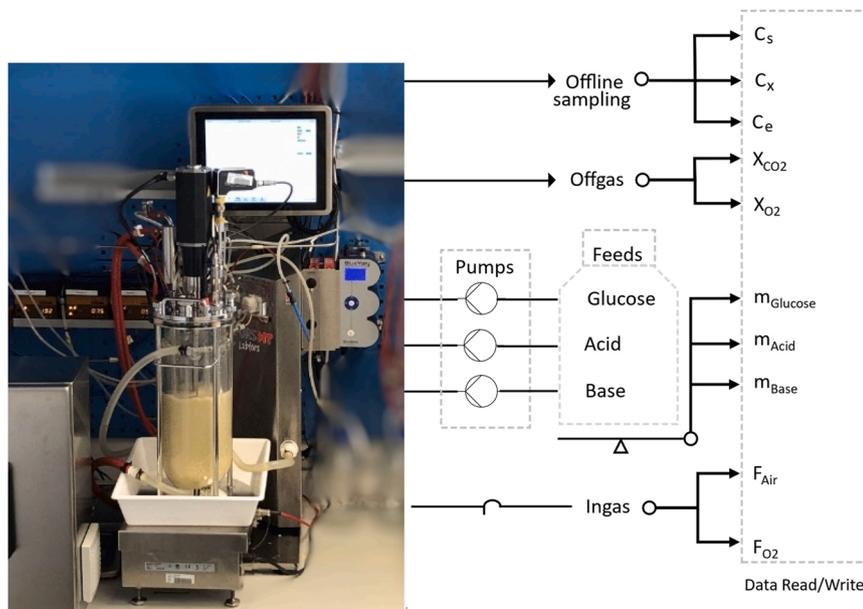


Fig. 1. Bioreactor image, and scheme indicating all collected data to simulate and to parametrize the model. Time-dependent inputs are determined as glucose, acid and base feed rates and are calculated from the respective balance signal ($m_{Glucose}$, m_{Acid} , m_{Base}). Online CER and OUR is calculated 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.

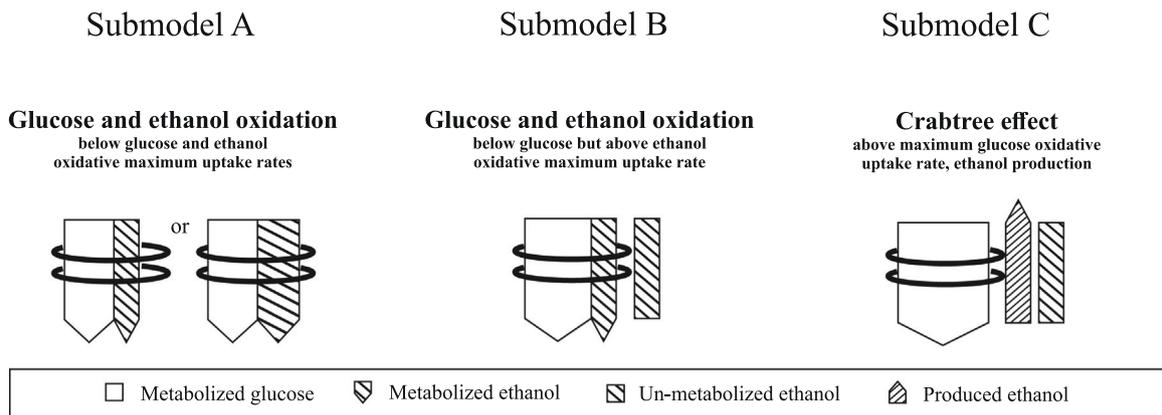


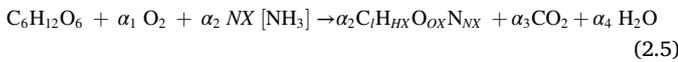
Fig. 2. The "bottleneck" concept in Sonnleitner and Käppli [51] yeast fermentation model. If the sum of substrate fluxes does not exceed, or equals, the cell's maximum oxidative uptake rate (shown as two rings in the figure), then the flux is subcritical, or critical, (left). If the glucose flux is higher 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.

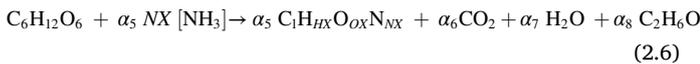
2.5. Growth model of *S. cerevisiae*

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 (α_1 - α_{12}):

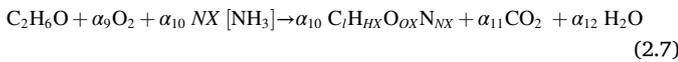
-Oxidative conversion of glucose (into biomass and CO₂)



-Reductive conversion of glucose (into biomass, CO₂ and ethanol)



-Oxidative conversion of ethanol (into biomass and CO₂)



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 α_2 , α_5 , and α_{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 α_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 O₂ 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:

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

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

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

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

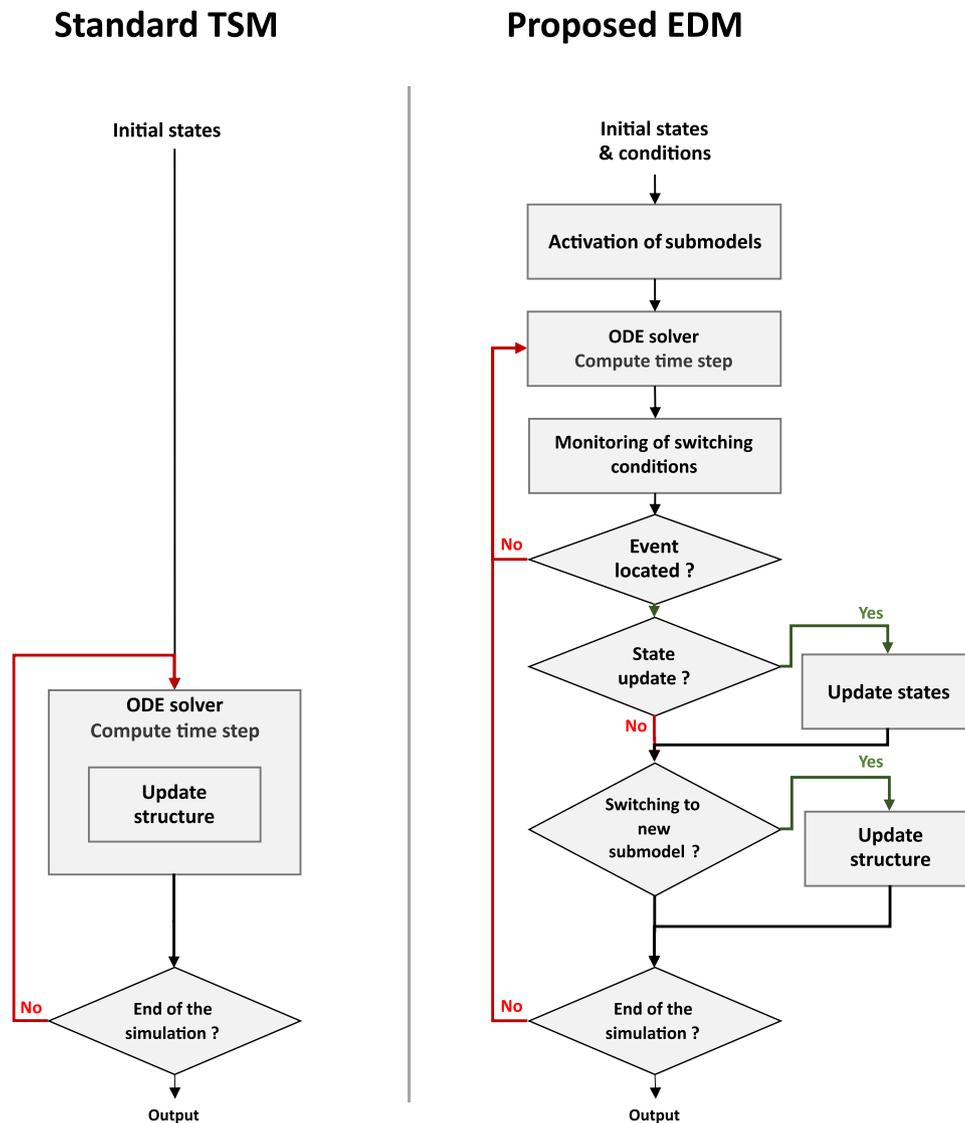


Fig. 3. Computational schemes of the standard time stepping method (TSM) and the proposed event driven method (EDM). The EDM scheme accounts explicitly for any discontinuities in the model by monitoring conditions and switching to the corresponding submodels (or updating state values).

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:

$$OUR = \left(q_{s(ox)} \cdot \alpha_1 \frac{MW_{O_2}}{MW_s} + q_{e(ox)} \cdot \alpha_9 \frac{MW_{O_2}}{MW_e} \right) C_x \cdot V \quad (2.19)$$

$$CER = \left(q_{s(ox)} \cdot \alpha_3 \frac{MW_{CO_2}}{MW_s} + q_{e(ox)} \cdot \alpha_{11} \frac{MW_{CO_2}}{MW_e} + q_{s(red)} \cdot \alpha_6 \frac{MW_{CO_2}}{MW_s} \right) C_x \cdot V \quad (2.20)$$

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 (θ)

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:

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

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 $F_{sampling}$ (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 $F_{sampling}$ 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 Käppeli [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:

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

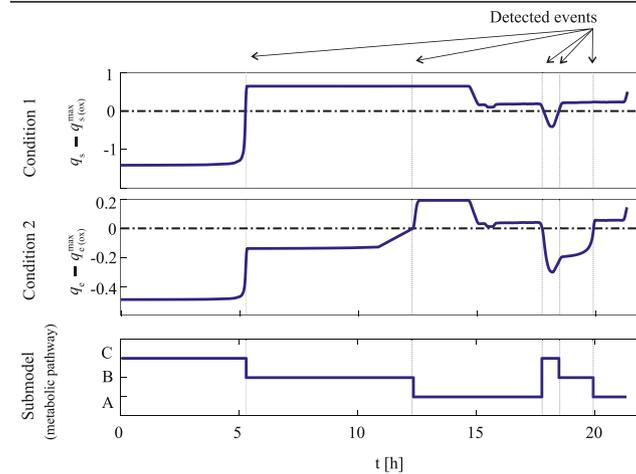
3.1.2.1. Accounting for additional discontinuities and non-physical solutions in the model. Non-physical solutions and highly nonlinear kinetic terms: Nonlinear kinetic models such as Monod-type growth models can exhibit stiff behaviors, especially when the affinity of the microorganisms to the used 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	$\mathcal{A} = \text{diag}[0 \ 1 \ 1 \ 0 \ 0]$

Conditions & submodels



States

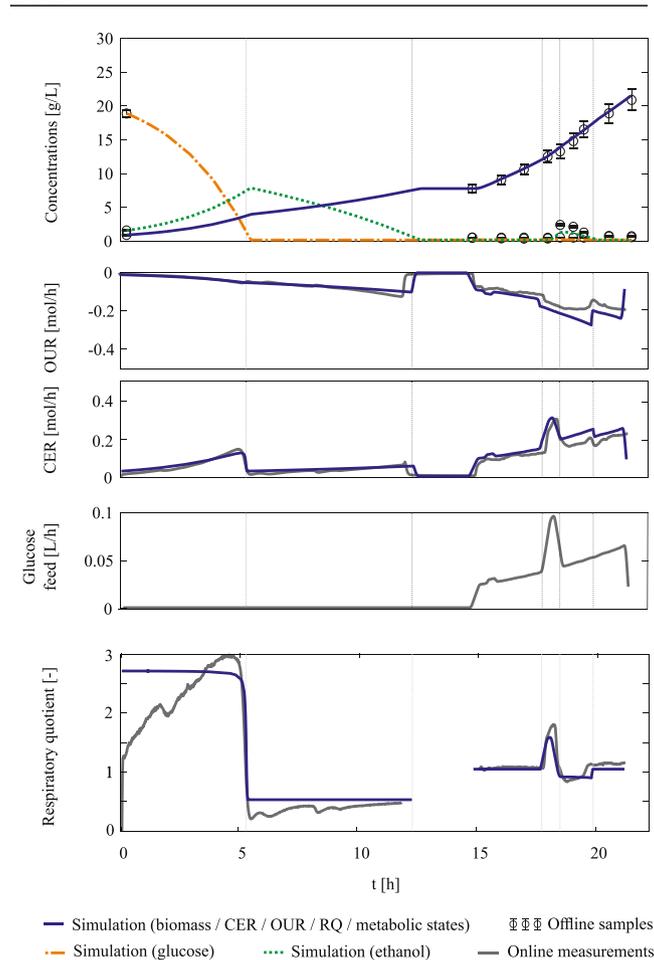


Fig. 4. Fitting the model following the EDM implementation to data from experiment 1. Above: Trigger functions (blue) and condition thresholds (dotted line). Detected events (blue line crosses dotted line) trigger a switch between submodels. Submodel C indicates a metabolism described by the Crabtree 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 $\sim 13 - 15$ [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_{Sampling}) usually show comparatively smooth curves. These curves are represented by discrete signals and can be transformed to smooth functions with relatively little

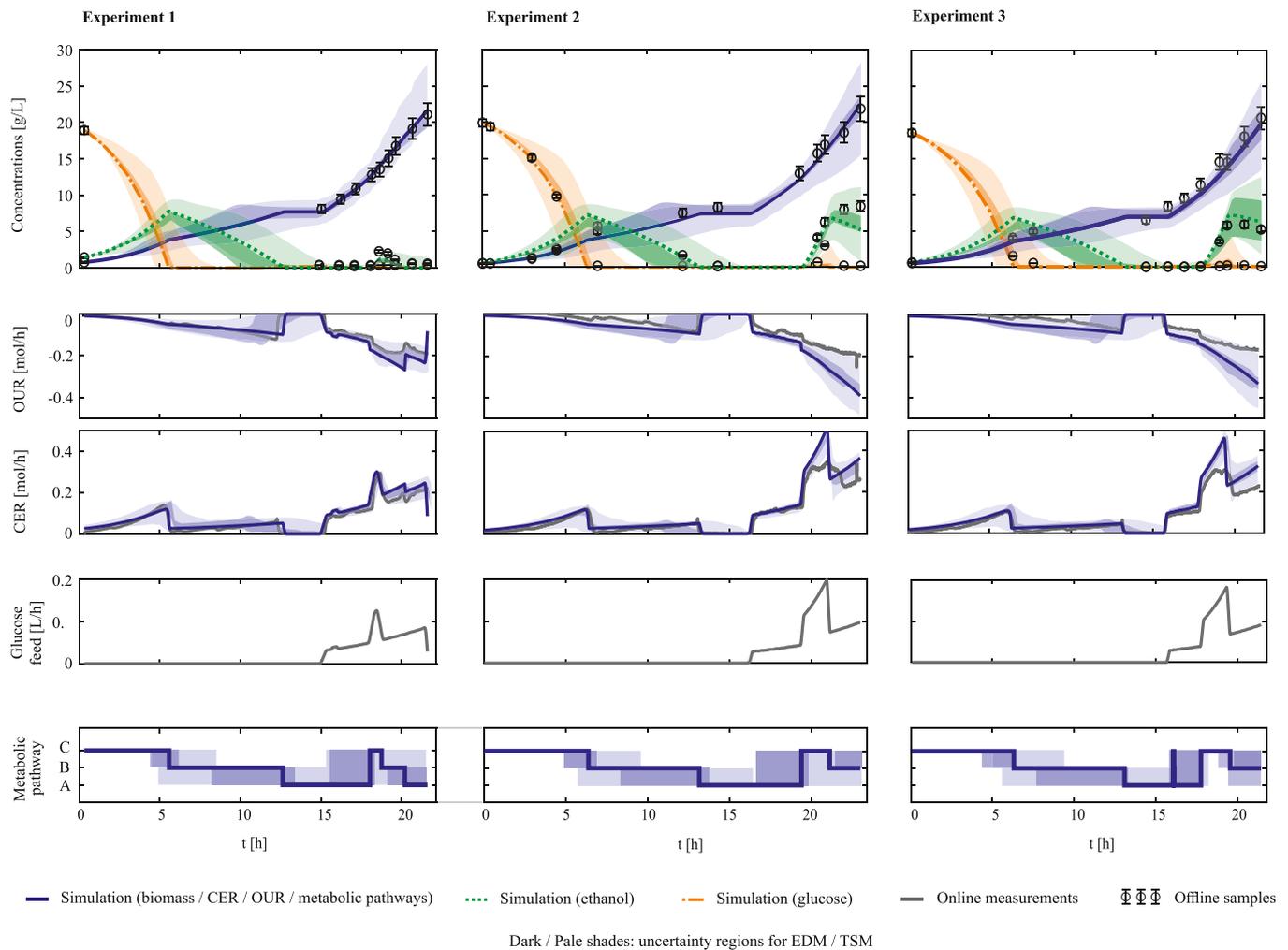


Fig. 5. Fitting the model to data from experiment 1, 2, and 3. The black solid lines and markers represent measurements, the colored lines represent the best fit (reference solution obtained by EDM). Uncertainties in the model predictions and in the location of switches between submodels are depicted 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_{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 ~ 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 (~ 15 h) subcritical glucose flux (metabolic pathway A) is aimed by a small exponential ramp. An RQ of ~ 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 % σ represents the relative standard deviation of the estimated parameters.

	Initial guess	Identifiability analysis Parameter subset Selection	Uncertainty quantification			
			Estimated value μ	% σ	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 ~ 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 ≈ 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).

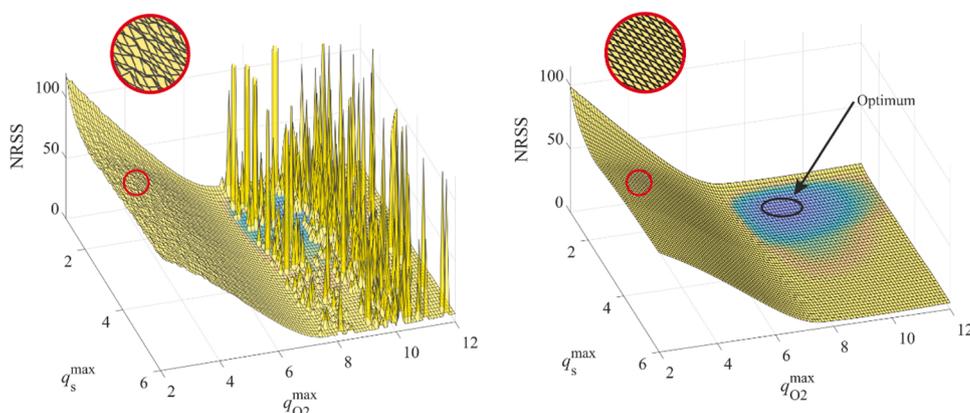


Fig. 6. NRSS (normalized residual sum of squared errors) optimization surfaces of a reduced two-dimensional parameter estimation problem for the model implementation following TSM (left) and EDM (right). The surfaces are constructed by evaluations (repeated simulations) for a grid of parameters. The 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 (% σ) 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 % σ	TSM % σ	EDM % σ	TSM % σ	EDM % σ	TSM % σ	EDM % σ
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

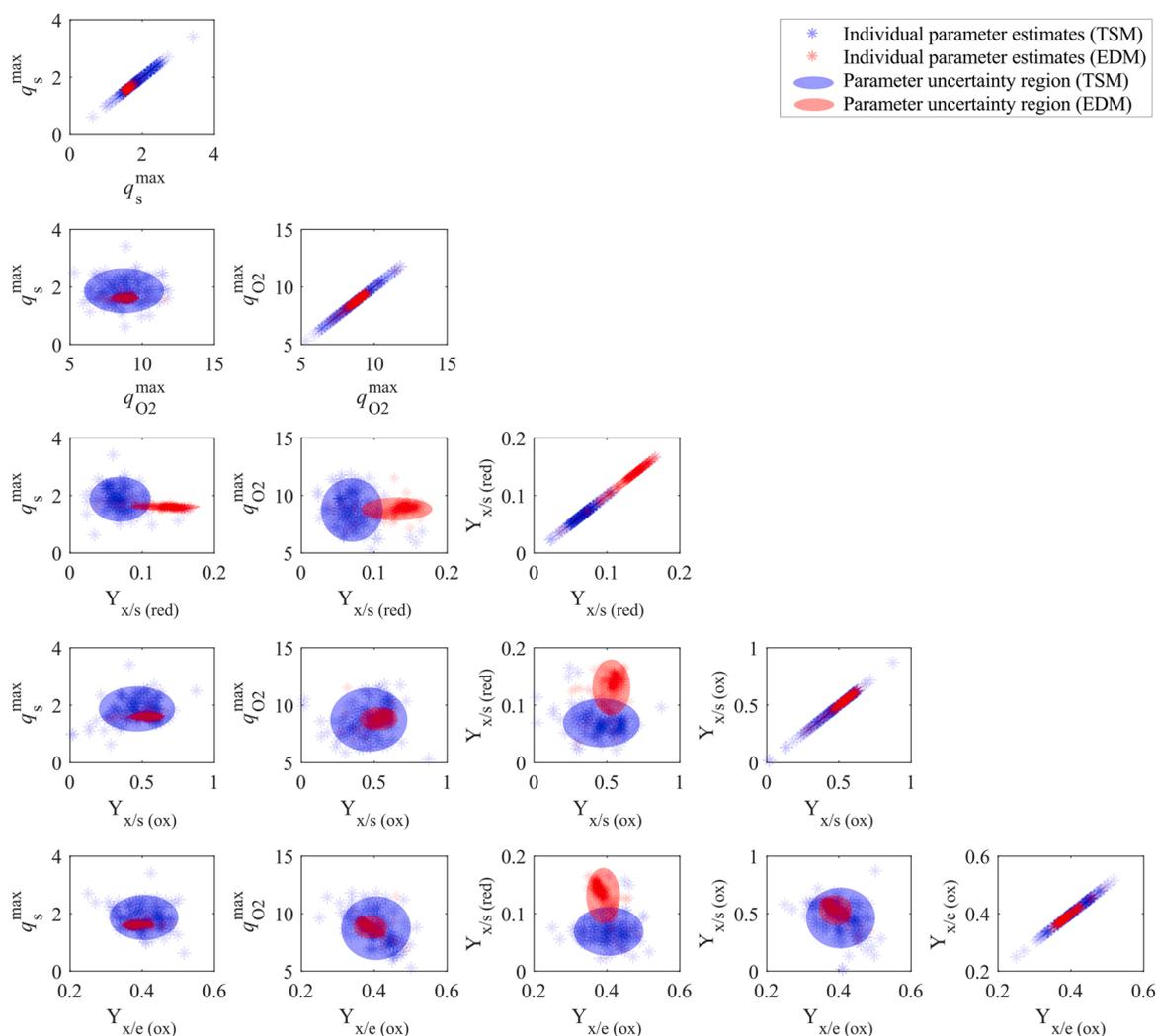


Fig. 7. Pairwise parameter estimates obtained by fitting 500 simulated data sets (generated by resampling the data from all experiments). For each fit the initial parameter guess was perturbed (perturbation levels 30%) from the best fitting parameter values in Table 4 (reference solution). The scatter 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 CO₂ and lower O₂ 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

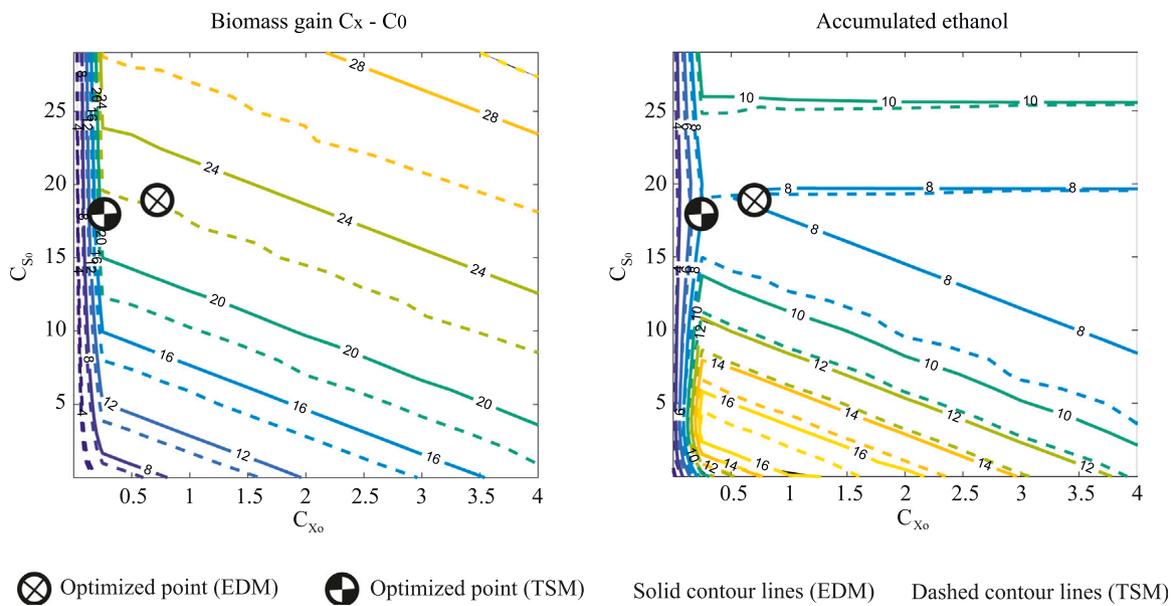


Fig. 8. Demonstration of the usage of the model for effective model-based DoE to optimize experiment 2. Right) The maximum accumulated ethanol, Left) The gain of biomass concentration during the time course of the experiment. Both are plotted against initial biomass and glucose concentrations. Lines: -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

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.

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].

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 ϵ defines the magnitude of the hysteresis band. It can be seen that the sign of ϵ 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. ϵ 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).

$$\begin{array}{c} V \\ x \\ s \\ e \\ CO_2 \\ O_2 \\ NH_3 \\ H_2O \end{array} \begin{array}{ccc} r1 & r2 & r3 \\ \left[\begin{array}{ccc} 0 & 0 & 0 \\ \alpha_2 & \alpha_5 & \alpha_{10} \\ 1 & 1 & 0 \\ 0 & \alpha_8 & -1 \\ \alpha_3 & \alpha_6 & \alpha_{11} \\ -\alpha_1 & 0 & -\alpha_9 \\ -\alpha_2 & -\alpha_5 & -\alpha_{10} \\ \alpha_4 & \alpha_8 & \alpha_{12} \end{array} \right] \end{array} \cdot \begin{array}{c} V \\ x \\ s \\ e \\ CO_2 \\ O_2 \\ NH_3 \\ H_2O \end{array} \begin{array}{ccccccc} \left[\begin{array}{ccccccc} 0 & 1 & 6 & 2 & 1 & 0 & 0 & 0 \\ 0 & HX & 12 & 6 & 0 & 0 & 3 & 2 \\ 0 & OX & 6 & 1 & 2 & 2 & 0 & 1 \\ 0 & NX & 0 & 0 & 0 & 0 & 1 & 0 \end{array} \right] \end{array} \begin{array}{c} C \\ H = 0 \\ O \\ N \end{array} \quad (\text{B.1})$$

Table B.1
Yeast fermentation media.

	Batch (1,5 L)	Fed-Batch (1 L)
Glucose monohydrate	33 g	220/200/198 g
(NH4)2SO4	7,5 g	5 g
KH2PO4	4,5 g	3 g
MgSO4 * 7H2O	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.

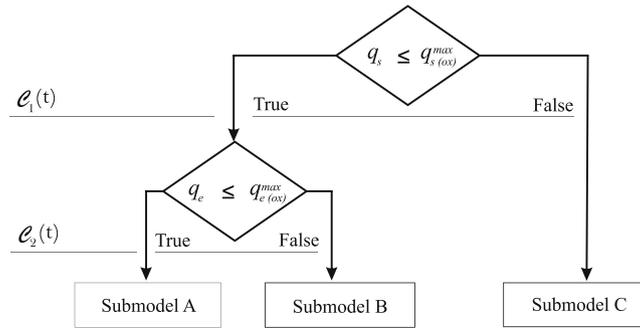


Fig. C1. Decision tree with two Boolean trigger functions $e_1(t)$, $e_2(t)$ and three submodels for the *Saccharomyces cerevisiae* fermentation model. Switches are made based on the metabolic flux capacity “bottleneck concept”. The decision tree is part of EDM computational scheme. EDM monitors the 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.

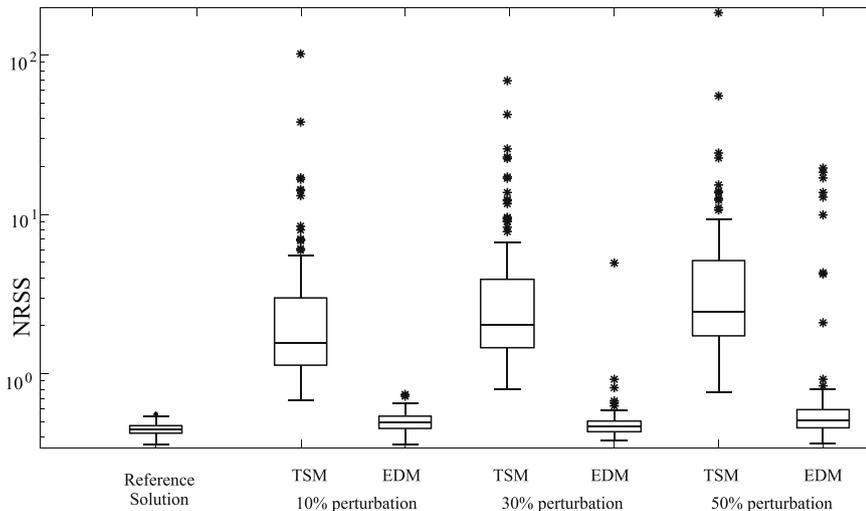


Fig. D1. Normalized residual sum of squares (NRSS) box plots obtained from fitting 500 simulated data sets (generated by resampling the data from all experiments) and starting with initial parameter values at different perturbation levels (10%,30% and 50%) from the best fitting parameter values in Table 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.

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/bpb.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 Ed. 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.
- [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.
- [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. Leclaire, 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.