Global parameter optimization for biokinetic modeling of simple batch experiments

Alma Mašić[a], Kai M. Udert[a], Kris Villez[a,*]

[a]*Eawag – Swiss Institute for Aquatic Science and Technology, Process Engineering, Dübendorf, Switzerland*

Abstract

Environmental process modeling is challenged by the lack of high quality data, stochastic variations, and nonlinear behavior. Conventionally, parameter optimization is based on stochastic sampling techniques to deal with the nonlinear behavior of the proposed models. Despite widespread use, such tools cannot guarantee globally optimal parameter estimates. It can be especially difficult in practice to differentiate between lack of algorithm convergence, convergence to a non-global local optimum, and model structure deficits. For this reason, we use a deterministic global optimization algorithm for kinetic model identification and demonstrate it with a model describing a typical batch experiment. A combination of interval arithmetic, reformulations, and relaxations allows globally optimal identification of all (six) model parameters. In addition, the results suggest that further improvements may be obtained by modification of the optimization problem or by proof of the hypothesized pseudo-convex nature of the problem suggested by our results.

Keywords: biotechnology, deterministic search, global optimality, kinetic modeling, nitrification, parameter estimation

*Corresponding author:

Postal address:	Eawag – Swiss Institute for Aquatic Science and Technology, Process Engineering, Überlandstrasse 133, P.O. Box 611, CH-8600 Dübendorf, Switzerland
E-Mail:	kris [dot] villez [at] eawag [dot] ch
Phone:	+41 58 765 5280
Fax:	+41 58 765 5802

Graphical abstract

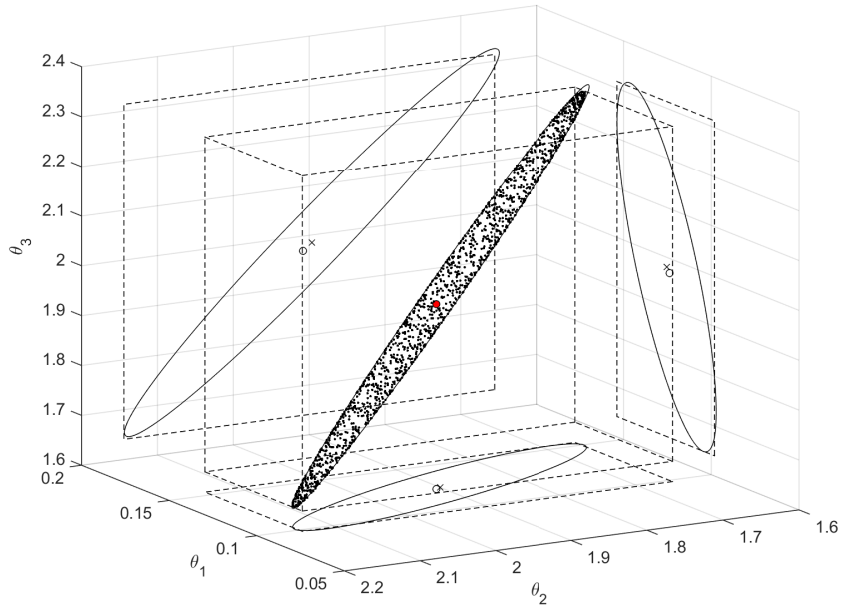

Figure 1: Graphical abstract.

1. Introduction

Despite abundant literature, model identification is a challenging task for environmental systems which keeps drawing considerable attention (Marsili-Libelli, 2010). In response, protocols have been developed to simplify the model identification task (Jakeman et al., 2006). One important aspect is that environmental process and system models are typically nonlinear in their parameters. Despite this problem, nonlinear parameter estimation is often solved with gradient-based optimization techniques that may not converge (Checchi et al., 2007) or which may converge to a local optimum (Jakeman et al., 2006; Rieger et al., 2012). Alternatively, stochastic optimization tools in combination with sensitivity-based parameter selection techniques (e.g., Benedetti et al., 2011; Sin et al., 2008) can ease this task. While fruitful in many cases, stochastic methods can still converge to a local optimum or may not converge at all. This is a significant drawback if the model structure itself is uncertain and subject to selection or modification. In other areas of engineering, deterministic optimization techniques are more popular. Whereas stochastic optimization methods *increase the chances* of finding global optima (in finite time), deterministic methods *find global optima without failure* (in finite time). Unfortunately, deterministic optimization still requires a deep understanding of the optimization problem and the most efficient algorithms tend to be tailored to a small set of optimization problems. However, with this work we show that deterministic optimization is at least applicable for modeling of simple batch respirometric experiments involving a single reaction. Since such experiments are typical for biological wastewater

---

Acronyms:

AOB	ammonia oxidizing bacteria
DO	dissoved oxygen
NOB	nitrite oxidizing bacteria
ODE	ordinary differential equations
OUR	oxygen uptake rate
QP	quadratic program
TNN	total nitrite nitrogen
WLS	weighted least squares
WRMSR	weighted root mean squared residuals

treatment process modeling, we argue that the provided parameter identification method is broadly applicable.

In addition to the nonlinear nature of the modeled processes, other factors complicating model identification include *(i)* the stochastic nature of their inputs, *(ii)* the lack of detailed understanding of metabolic pathways, and *(iii)* the large number of empirically determined parameters further leading to a lack of practical or even structural identifiability. While these issues are important, they are not addressed, diminished, or amplified by this work. Thus, we consider the experimental design and the produced experimental data as a given and focus on solving parameter estimation problems to global optimality.

To showcase the real-world applicability of the developed optimization method, a data set collected for the purpose of kinetic model identification of a biological urine treatment process is used. Separate collection and treatment of urine is a new approach to optimize sanitation. Two possible applications are the recovery and recycling of nutrients to agriculture (Udert and Wächter, 2012) and the prevention of corrosion in sewers by nitrate dosage (Jiang et al., 2011; Oosterhuis and van Loosdrecht, 2009). Nitrification of urine is applied in both approaches, either to stabilize volatile ammonia or to produce the electron acceptor nitrate. Stable nitrification requires balanced activities of both bacterial groups involved in the process, ammonium oxidizing bacteria (AOB) and nitrite oxidizing bacteria (NOB). However, stable nitrification is challenging in urine due to the high pH value and the high concentrations of ammonia, organic substances, and salts. Three major process failures can occur (Fumasoli, 2016). First, both AOB and NOB are inhibited at high pH values due to high concentrations of free ammonia. Second, at intermediate pH values AOB grow too fast and produce large amounts of nitrite, which inhibit NOB. Third, acid-tolerant AOB grow in when the operational pH is low. In turn, the pH value can decrease even further leading to the chemical production of large amounts of volatile nitrogen compounds, especially nitric oxide (Fumasoli et al., 2015). The main operational parameter is the pH value. It directly influences the energy generation of the bacteria, but it also determines *(i)* limitation effects by free ammonia and carbonate and *(ii)* inhibition effects by free ammonia and nitrous acid (Fumasoli, 2016). Keeping these effects apart and determining the respective kinetic constants is challenging. Consequently, mechanistic computer models can be a helpful tool to include all effects and the necessary chemical and microbial processes (Fumasoli, 2016). Jubany et al. (2005) showed that con-

secutive dosage of nitrite and fitting the oxygen uptake rate can be used to determine the kinetics of NOB in high-strength ammonia wastewaters. This approach to experimental data collection is also applied in our study in order to demonstrate our optimization algorithm.

The next section describes the experimental data and the applied optimization algorithm. Afterwards, results are shown and discussed in separate sections. The major conclusions are summarized at the end.

2. Materials and Methods

2.1. Notation and symbols

The notation conventions applied in this study are given in Table 1. All symbols used in this study are given in Table 2. In addition, inequalities of the form $\boldsymbol{x} \leq \boldsymbol{y}$ express that every element in $\boldsymbol{x}$ is smaller or equal to the corresponding element in $\boldsymbol{y}$, i.e. $\boldsymbol{x} \leq \boldsymbol{y} \Leftrightarrow \forall l : x_l = \boldsymbol{x}(l) \leq \boldsymbol{y}(l) = y_l$. Similarly, we write for matrices that $\boldsymbol{X} \leq \boldsymbol{Y} \Leftrightarrow \forall l, m : X_{l,m} = \boldsymbol{X}(l,m) \leq \boldsymbol{Y}(l,m) = Y_{l,m}$.

Table 1: Notation conventions

Notation	Description
x, θ	Scalar
$\boldsymbol{x}$, $\boldsymbol{x}_m$, $\boldsymbol{\theta}$	Column vector
x_l, $\boldsymbol{x}(l)$	l^{th} scalar element of vector $\boldsymbol{x}$
$X_{l,m}$, $\boldsymbol{X}(l,m)$	Scalar element of matrix $\boldsymbol{X}$ at row l and column m
$\boldsymbol{X}$	Matrix
$\boldsymbol{X}_{l,\cdot}$	l^{th} row from matrix $\boldsymbol{X}$
$\boldsymbol{X}_{\cdot,m}$	m^{th} column from matrix $\boldsymbol{X}$
$\tilde{x}$	Measurement
$\hat{x}$	Optimal estimate
$\breve{x}$, $\mathring{x}$	Relaxed estimate or value
x^+	Positive part of x ($\max(x,0)$)
x^-	Negative part of x ($\min(x,0)$)
$\underline{x}$	Lower bound (h,h_j,h_1,h_2), lower interval limit (q,s,θ)
$\overline{x}$	Upper bound (h,h_j,h_1,h_2), upper interval limit (q,s,θ)
$\mathcal{X}$	Set of feasible solutions

Table 2: List of symbols

Symbol	Description	Unit
a	Biomass activity	$[mgN\ L^{-1}\ h^{-1}]$
$a_{\max}$	Maximum biomass activity	$[mgN\ L^{-1}\ h^{-1}]$
b_{NOB}	Biomass decay rate of NOB	$[mg\ L^{-1}\ h^{-1}]$
$\boldsymbol{c}$	Regression constraint vector	$[-]$
d_k, d_{j,k_j}	Residuals	$[mgN\ L^{-1}]$, $[mg\mathrm{O}_2\ L^{-1}\ h^{-1}]$
e_{j,k_j}	Measurement errors	$[mgN\ L^{-1}]$, $[mg\mathrm{O}_2\ L^{-1}\ h^{-1}]$
h	Objective function	$[-]$
h_j, h_o, h_1, h_2	Objective function term	$[-]$
i_{decay}	Stoichiometric coefficient for oxygen in biomass decay reaction	$[mg\mathrm{O}_2\ mg^{-1}]$
i_{growth}	Stoichiometric coefficient for oxygen in biomass growth reaction	$[mg\mathrm{O}_2\ (mgN)^{-1}]$
j	Measured variable index	–
k_j	Measurement sample index	–
l, m	Integer index	–
p	Number of nonlinear parameters	–
pK_{a}, $pK_{\mathrm{a,HNO_2}}$, $pK_{\mathrm{a,HNO_3}}$	Acidity constants	$[-]$
q, q_{k_j}	Relative reaction rate	$[h^{-1}]$
r_{aer}	Oxygen mass transfer rate	$[mg\mathrm{O}_2\ L^{-1}\ h^{-1}]$
r_{endo}	Endogenous oxygen uptake rate	$[mg\mathrm{O}_2\ L^{-1}\ h^{-1}]$
r_{OUR}	Oxygen uptake rate	$[mg\mathrm{O}_2\ L^{-1}\ h^{-1}]$
s, s_{k_j}	Relative concentration	$[-]$
s_0	Initial relative concentration	$[-]$
$s_{\max}$	Relative concentration corresponding to the maximum growth rate	$[-]$
t	Time (continuous)	$[h]$
t_{k_j}	Measurement sampling times	$[h]$
t_{k_1}, $t_{k_{\mathrm{TNN}}}$	Sampling times for TNN measurements	$[h]$
t_{k_2}, $t_{k_{\mathrm{OUR}}}$	Sampling times for OUR measurements	$[h]$

Continued on next page

List of symbols

Symbol	Description	Unit
y_k, y_{j,k_j}, y_{1,k_1}, $y_{\mathrm{TNN},k_{\mathrm{TNN}}}$, y_{2,k_2}, $y_{\mathrm{OUR},k_{\mathrm{OUR}}}$	Model output	$[mgN\ L^{-1}]$, $[mg\mathrm{O}_2\ L^{-1}\ h^{-1}]$
$\boldsymbol{B}$	Regression constraint matrix	–
C	Constant	–
J	Number of measured variables	–
K_j	Number of measurement samples	–
K_{S}	NOB affinity constant for HNO_2	$[mgN\ L^{-1}]$
K_{I}	NOB inhibition constant for HNO_2	$[mgN\ L^{-1}]$
M	Number of regression inputs	–
S_{O_2}	DO concentration	$[mg\mathrm{O}_2\ L^{-1}]$
S_{TNN}	TNN concentration	$[mgN\ L^{-1}]$
$S_{\mathrm{TNN},0}$	Initial TNN concentration	$[mgN\ L^{-1}]$
$\mathcal{T}$	Feasible set of parameter vectors	–
X_{NOB}	NOB concentration	$[mg\ L^{-1}]$
$\boldsymbol{X}$	Regression inputs	$[-]$
Y_{NOB}	Biomass growth yield coefficient	$[mg\ (mgN)^{-1}]$
$\boldsymbol{\beta}$, $\boldsymbol{\beta}_j$, $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$	Regression parameter vectors	$[mgN\ L^{-1}]$, $[mg\mathrm{O}_2\ L^{-1}\ h^{-1}]$
$\beta_{j,0}$, $\beta_{j,1}$, $\beta_{j,2}$, $\beta_{1,0}$, $\beta_{1,1}$, $\beta_{1,2}$, $\beta_{2,0}$, $\beta_{2,1}$, $\beta_{2,2}$	Regression parameters	$[mgN\ L^{-1}]$, $[mg\mathrm{O}_2\ L^{-1}\ h^{-1}]$
$\boldsymbol{\gamma}$	Parameter vector	Mixed units
$\mu(t)$	Specific growth rate	$[h^{-1}]$
$\mu_{\max}$	Maximum specific growth rate	$[h^{-1}]$
σ_k, σ_{j,k_j}, σ_{1,k_1}, $\sigma_{\mathrm{TNN},k_{\mathrm{TNN}}}$, σ_{2,k_2}, $\sigma_{\mathrm{OUR},k_{\mathrm{OUR}}}$	Measurement error standard deviations	$[mgN\ L^{-1}]$, $[mg\mathrm{O}_2\ L^{-1}\ h^{-1}]$
$\boldsymbol{\theta}$, $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$	Dimensionless kinetic parameter vectors	$[h^{-1}]$
θ_j, θ_1, θ_2, θ_3	Dimensionless kinetic parameter	$[h^{-1}]$
Γ	Convex set	–
Ω, Ω_j	Convex set	–

2.2. Assumed model structure and general problem statement

The parameter optimization method as developed in this work applies to process models whose dynamics can be formulated as follows:

$$\dot{s}(t) = -q\big(s(t), \boldsymbol{\theta}\big), \qquad\qquad s(0) = 1 \tag{1}$$

with $s(t)$ the single state variable and $q\big(s(t), \boldsymbol{\theta}\big)$ a single rate expression. The state variable can only take on nonnegative values ($s(t) \geq 0$) and the rate expression $q\big(s(t), \boldsymbol{\theta}\big)$ is nonnegative and non-increasing in its parameters ($\boldsymbol{\theta}$, dimensions: $p \times 1$) over its whole domain:

$$\forall s \in \mathbb{R}_{\geq 0}, \forall \boldsymbol{\theta}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathbb{R}^p : \begin{cases} q\big(s, \boldsymbol{\theta}\big) \geq 0 \\ \boldsymbol{\theta}_1 \leq \boldsymbol{\theta}_2 \Leftrightarrow q\big(s, \boldsymbol{\theta}_1\big) \geq q\big(s, \boldsymbol{\theta}_2\big) \end{cases} \tag{2}$$

The process state and/or the rate of change ($s(t)$ and $q(t)$) are measured through equations of the following form:

$$\tilde{y}_{j,k_j} = y_{j,k_j} + e_{j,k_j}, \qquad\qquad e_{j,k_j} \sim \mathcal{N}(0, \sigma_{j,k_j}) \tag{3}$$

$$\begin{aligned} y_{j,k_j} &= \beta_{j,0} + \beta_{j,1}\ s(t_{k_j}) + \beta_{j,2}\ q(s(t_{k_j}), \boldsymbol{\theta}) \\ &= \begin{bmatrix} 1 & s(t_{k_j}) & q(s(t_{k_j}), \boldsymbol{\theta}) \end{bmatrix} \boldsymbol{\beta}_j, \qquad j = 1, \ldots, J, \quad k_j = 1, \ldots, K_j \end{aligned} \tag{4}$$

$$\boldsymbol{\beta}_j \in \Omega_j \subset \mathbb{R}^3_{\geq 0} \qquad\qquad j = 1, \ldots, J \tag{5}$$

These measurement equations deliver K_j measurements $\tilde{y}_{j,k_j}$ of J measured variables y_{j,k_j} at sampling times t_{k_j}, where $k_j = 1, \ldots, K_j$ and $j = 1, \ldots, J$. The measurement errors e_{j,k_j} are assumed to be sampled independently from zero mean normal distributions with standard deviations σ_{j,k_j}. These standard deviations are assumed known. In addition, the vectors $\boldsymbol{\beta}_j = [\beta_{j,0}\ \beta_{j,1}\ \beta_{j,2}]^{\mathrm{T}}$ ($j = 1, \ldots, J$) are bound to belong to a subset of the non-negative real space, Ω_j. These subsets are assumed known and are required to be convex. The vectors $\boldsymbol{\theta}$ and $\boldsymbol{\beta}_j$ ($j = 1, \ldots, J$) constitute the parameters of the model and are to be estimated.

2.3. Parameter estimation methods

2.3.1. Definition of optimality

We define optimal parameter estimation as maximum likelihood estimation, that is, we aim to find the values for the parameters which maximize the likelihood. Let $\boldsymbol{\gamma}$ denote the vector containing all parameters :

$$\boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\theta}^{\mathrm{T}} & \boldsymbol{\beta}_1{}^{\mathrm{T}} & \boldsymbol{\beta}_2{}^{\mathrm{T}} & \ldots & \boldsymbol{\beta}_j{}^{\mathrm{T}} & \ldots & \boldsymbol{\beta}_J{}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \tag{6}$$

and let $h(\boldsymbol{\gamma})$ be the negative log-likelihood function. The optimization problem is then:

$$\hat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}} h(\boldsymbol{\gamma}) \tag{7}$$

Given assumptions and definitions discussed above, the negative log-likelihood corresponds to the following weighted least squares (WLS) objective function:

$$\begin{aligned} h(\boldsymbol{\gamma}) &= C + \sum_{j=1}^{J} h_j(\boldsymbol{\theta}, \boldsymbol{\beta}_j) \\ &= C + \sum_{j=1}^{J} \sum_{k_j=1}^{K_j} \left(\frac{d_{j,k_j}}{\sigma_{j,k_j}} \right)^2 \end{aligned} \tag{8}$$

$$d_{j,k_j} = y_{j,k_j}(\boldsymbol{\gamma}) - \tilde{y}_{j,k_j} \tag{9}$$

with C a constant which can be ignored during optimization, d_{j,k_j} residuals, and $y_{j,k_j}(\boldsymbol{\gamma})$ simulated values for y_{j,k_j} obtained with parameter vector $\boldsymbol{\gamma}$ according to (1)-(5).

2.3.2. Estimation of $\boldsymbol{\beta}_j$ conditional to $\boldsymbol{\theta}$

A major benefit of the model formulation given above is that the parameter vectors $\boldsymbol{\beta}_j$ appear linearly and separately in the measurement equations. Indeed, the measurements are linear in these parameters. This means that optimal WLS values for these parameters can be found easily provided that $s(t)$ and $q\big(s(t), \boldsymbol{\theta}\big)$ are known at the measurement sampling times. Indeed, given $\boldsymbol{\theta}$, the state $s(t)$ can be evaluated for any time t by

$$s(t, \boldsymbol{\theta}) = s_0 - \int_0^t q\big(s(t), \boldsymbol{\theta}\big)\, dt, \quad s(0) = s_0 = 1 \tag{10}$$

To find estimates for $\boldsymbol{\beta}_j$, equation (10) is evaluated for every instant t_{k_j} by solving the following problems:

$$j = 1, \ldots, J: \ \hat{\boldsymbol{\beta}}_j(\boldsymbol{\theta}) = \arg\min_{\boldsymbol{\beta}_j \in \Omega_j} \ h_j(\boldsymbol{\theta}, \boldsymbol{\beta}_j) = \sum_{k_j=1}^{K_j} \left(\frac{d_{j,k_j}}{\sigma_{j,k_j}} \right)^2 \tag{11}$$

subject to

$$\begin{aligned} d_{j,k_j} &= y_{j,k_j}(\boldsymbol{\gamma}) - \tilde{y}_{j,k_j} \\ &= \left[\begin{array}{cccc} 1 & s(t_{k_j}) & q(\,s(t_k)|_{\boldsymbol{\theta}}\,,\boldsymbol{\theta}) \end{array}\right] \boldsymbol{\beta}_j - \tilde{y}_{j,k_j} \end{aligned} \tag{12}$$

This problem is convex in β_j and is therefore solved efficiently by means of interior-point programming in the general case. If the set Ω_j is polyhedral (i.e., described completely by linear equalities and inequalities) the above problem is a quadratic program (QP) in $\boldsymbol{\beta}_j$.

Given the above solutions $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})_j$, one can now write the original optimization problem as follows:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}}\, h(\boldsymbol{\theta}) = \sum_{j=1}^{J} h_j\big(\boldsymbol{\theta}, \boldsymbol{\beta}_j\big) \tag{13}$$

subject to

$$h_j\big(\boldsymbol{\theta}, \boldsymbol{\beta}_j\big) = \sum_{k_j=1}^{K_j} \left(\frac{d_{j,k_j}}{\sigma_{j,k_j}}\right)^2 \tag{14}$$

$$d_{j,k_j} = \left[\begin{array}{cccc} 1 & s(t_{k_j}) & q(\,s(t_k)|_{\boldsymbol{\theta}}\,,\boldsymbol{\theta}) \end{array}\right] \boldsymbol{\beta}_j - \tilde{y}_{j,k_j} \tag{15}$$

This optimization problem is nonlinear in the remaining parameters $\boldsymbol{\theta}$. The particular instance of this problem studied in this work is solved in two ways. Firstly, it is solved by means of the branch-and-bound algorithm explained in the next paragraphs. Secondly, it is solved by means of the quasi-newton algorithm as is discussed below as well.

2.3.3. Deterministic optimization

The branch-and-bound algorithm is a popular and general method for solving nonlinear optimization problems in a deterministic fashion and is based on a divide-and-conquer strategy. Its applicability depends on the availability of provable upper and lower bounds to the objective function for subsets of the parameter search space. For the parameter optimization described above, such bounds are given below.

The basic procedure for branch-and-bound optimization is as follows. The branch-and-bound algorithm is initiated with a polyhedral set (i.e., a box in the multidimensional parameter space) containing all considered solutions

(parameter vectors) to the problem. This primary set is called the *root set* or *root*. The algorithm proceeds by *branching* from this root set. In this step, the set is halved into two non-overlapping sets by separating the solutions above and below the center value for one of the parameters over the set. The produced sets are called *leaf sets* or *leaf nodes*. This branching continues at every iteration by selecting a leaf node and branching from this node. Sets which have been branched into leaf nodes are called *branches* as they form a hierarchical solution tree. To prevent complete enumeration of all parameter subsets, bounding procedures are implemented. For every newly generated leaf set, one computes an upper bound and lower bound to the objective function for that set. These are classically defined as follows:

- *Upper bound.* An upper bound is a value which is guaranteed to be higher or equal to the objective function value for at least one feasible solution within the considered set.

- *Lower bound.* A lower bound is a value which is guaranteed to be lower or equal to the objective function value for every feasible solution within the considered set.

At every iteration of the algorithm, one now compares the lower and upper bounds for every pair of available leaf sets. If for a given set $\mathcal{A}$ the lower bound is higher than the upper bound for another set $\mathcal{B}$ then the set $\mathcal{A}$ cannot contain the global optimum. As a result, one can ignore the set $\mathcal{A}$ during the remainder of the algorithmic search for the global optimum. This is called *fathoming* and speeds up algorithm convergence as parts of the solution tree can be ignored without jeopardizing global optimality. For efficient convergence, the lower bound should be as close as possible to the actual minimum objective function within each of the considered leaf sets. The fathomed leaf sets are referred to as *dead nodes*. The remaining leaf sets, *live nodes*, remain available for continued branching and bounding. The algorithm is terminated when a predetermined stopping criterion is met. For more information of deterministic optimization schemes we refer to Floudas (1999); Nocedal and Wright (2006); Forst and Hoffmann (2010).

Additional implementation choices are as follows:

- *Stopping criterion.* The algorithm is terminated when all live leaf nodes are smaller than a set resolution in every dimension.

- *Node selection.* At every iteration of the branch-and-bound optimization algorithm, one must choose a node to be branched from. For the presented work, the node with the lowest lower bound value was chosen at every iteration.

- *Branching.* At every branching step, one must choose along which parameter one must split the given set into two new leaf sets. In our implementation of the branch-and-bound algorithm, branching is executed by splitting evenly along the longest dimension of the considered polyhedral set.

2.3.4. Bounds to the objective function

Assume that at a given iteration of the search algorithm a leaf set of values for $\boldsymbol{\theta}$ are given as a box set, $\mathcal{T}$, defined as follows:

$$\forall \boldsymbol{\theta} \in \mathcal{T} : \underline{\boldsymbol{\theta}} \leq \boldsymbol{\theta} \leq \overline{\boldsymbol{\theta}} \tag{16}$$

The following paragraphs describe how upper bounds and lower bounds to the objective function in (13) can be computed.

Upper bound. Computing a valid upper bound is a fairly simple task, as is typical for most optimization problems. In our implementation we compute the value for $h(\boldsymbol{\theta})$ twice, namely for $\underline{\boldsymbol{\theta}}$ and $\overline{\boldsymbol{\theta}}$. Given the expressions for $h_j(\boldsymbol{\theta}, \beta_j)$ in (13), it is necessary to obtain the values $\hat{\boldsymbol{\beta}}_j(\underline{\boldsymbol{\theta}})$ and $\hat{\boldsymbol{\beta}}_j(\overline{\boldsymbol{\theta}})$ ($j = 1, \ldots, J$) by solving the problems (11). Upon evaluation of both objective function values, $h\big(\underline{\boldsymbol{\theta}}\big)$ and $h\big(\overline{\boldsymbol{\theta}}\big)$, one obtains a valid upper bound by selecting the minimum of both. We write

$$\overline{h}(\mathcal{T}) = \min\left\{h\big(\underline{\boldsymbol{\theta}}\big), h\big(\overline{\boldsymbol{\theta}}\big)\right\}. \tag{17}$$

It is fairly trivial to see that the obtained value for the upper bound satisfies the definition given above.

Lower bound. Computing a lower bound is not trivial and deserves careful attention. In what follows, we describe the development of the obtained lower bound which is based on some of the simplest rules of interval arithmetic (Hansen and Walster, 2003; Moore et al., 2009) and a relaxation of WLS regression. In the results section, the lower bounding procedure is demonstrated in detail.

To construct a lower bound computing procedure, we first consider that the rate expression $q\big(s,\boldsymbol{\theta}\big)$ can be bounded as follows:

$$q\big(s,\overline{\boldsymbol{\theta}}\big) \leq q\big(s,\boldsymbol{\theta}\big) \leq q\big(s,\underline{\boldsymbol{\theta}}\big), \quad \forall s \in \mathbb{R}_{\geq 0}. \tag{18}$$

This implies that the slowest obtainable rate, given s and for any choice for $\boldsymbol{\theta} \in \mathcal{T}$, is obtained by setting $\boldsymbol{\theta} = \overline{\boldsymbol{\theta}}$. This is fairly trivial since the rate expression was defined to be monotonically decreasing in every element of $\boldsymbol{\theta}$. Similarly, the fastest attainable reaction rates are found by setting $\boldsymbol{\theta} = \underline{\boldsymbol{\theta}}$.

Using the previous definition in (10), it follows that the state $s(t)$ evaluated in t_{k_j} ($s_{k_j} = s(t_{k_j})\big|_{\boldsymbol{\theta}}$) can be bounded as follows:

$$s(t_{k_j})\big|_{\underline{\boldsymbol{\theta}}} = \underline{s_{k_j}} \leq s_{k_j} \leq \overline{s_{k_j}} = s(t_{k_j})\big|_{\overline{\boldsymbol{\theta}}}, \ \ \forall \boldsymbol{\theta} \in \mathcal{T}, k_j = 1, \ldots, K_j \tag{19}$$

Hence, the lowest relative concentrations are obtained for the lowest values of $\boldsymbol{\theta}$ within the set $\mathcal{T}$. This is fairly intuitive, as the highest process rates will deliver the fastest decreases of the state and thus the lowest values for the state. Similarly, the highest relative concentrations are obtained for the highest values of $\boldsymbol{\theta}$. These bounds are tight and are easy to obtain thanks to the required properties of the rate function (2).

Next, bounds on the process rate $q(s,\boldsymbol{\theta})$ are sought at every time instant. That is, we seek to find bounding values $\underline{q_{k_j}}$ and $\overline{q_{k_j}}$ which bound $q_{k_j} = q\big(\, s(t_{k_j})\big|_{\boldsymbol{\theta}}, \boldsymbol{\theta}\big)$ as follows:

$$\underline{q_{k_j}} \leq q_{k_j} \leq \overline{q_{k_j}}, \forall \boldsymbol{\theta} \in \mathcal{T}, \forall s_{k_j} \in \left\{ s_{k_j} | \underline{s_{k_j}} \leq s_{k_j} \leq \overline{s_{k_j}} \right\}, \\ j = 1, \ldots, J, k_j = 1, \ldots, K_j. \tag{20}$$

The bounds on the rate expression should define an interval within which any possible rate expression evaluation lies for any feasible parameter vector $\boldsymbol{\theta}$ and for any feasible value s.

Due to the monotonically decreasing property of $q\big(s,\boldsymbol{\theta}\big)$ (2), the lowest (highest) reaction rate can only be obtained for the highest (lowest) values of $\boldsymbol{\theta}$ within $\mathcal{T}$, namely $\overline{\boldsymbol{\theta}}$ ($\underline{\boldsymbol{\theta}}$). It is however more difficult to evaluate at which value of s within the interval $\left[\underline{s_{k_j}}, \overline{s_{k_j}}\right]$ one obtains a minimum (maximum) for $q\big(s,\overline{\boldsymbol{\theta}}\big)$. To handle this, several options are available. First, if the shape of the function is known simple rules of interval airthmetic can be applied. For instance, rate expressions that are monotonically increasing in s find their minimum (maximum) at $\underline{s}_{k_j}$ ($\overline{s}_{k_j}$) and rate expressions

that are pseudo-convex have a single minimum inside the interval $\left[\underline{s_{k_j}}, \overline{s_{k_j}}\right]$ which may be available analytically. The maximum of pseudo-convex rate expressions is found at either $\underline{s_{k_j}}$ and $\overline{s_{k_j}}$ following explicit evaluation at both locations. Secondly, for sufficiently simple expressions one can find all local minima (maxima) within the interval $\left[\underline{s_{k_j}}, \overline{s_{k_j}}\right]$ by solving one or more algebraic equations of the form $q\big(s, \boldsymbol{\theta}\big) = 0$. The overall minimum and maximum then follows from taking the minimum (maximum) among these minima (maxima) and values obtained at the interval boundaries. Finally, interval arithmetic rules can be applied in an automatic fashion to bound the rate expressions. To this end, specialized code libraries are available (Rump, 1999). As will be shown below, the first option can be applied to the case studied in this work. In what follows, we assume that provable bounds ($\underline{q_{k_j}}$, $\overline{q_{k_j}}$) are available.

To enable the computation of a lower bound to h_j ($j = 1, \ldots, J$), the QPs as executed for the upper bound (11) are relaxed as follows:

$$j = 1, \ldots, J: \ \breve{\boldsymbol{\beta}}_j(\mathcal{T}) = \arg \min_{\boldsymbol{\beta}_j \in \Omega_j} \underline{h_j}(\mathcal{T}, \boldsymbol{\beta}_j) \tag{21}$$

subject to

$$\underline{h_j}(\mathcal{T}, \boldsymbol{\beta}_j) = \sum_{k_j=1}^{K_j} \left(\frac{\breve{d}^{+}_{j,k_j}}{\sigma_{j,k_j}} \right)^2 + \sum_{k_j=1}^{K_j} \left(\frac{\breve{d}^{-}_{j,k_j}}{\sigma_{j,k_j}} \right)^2 \tag{22}$$

$$\breve{d}^{+}_{1,k_j} = \left[\begin{array}{ccc} 1 & \underline{s_{k_j}} & \underline{q_{k_j}} \end{array} \right] \boldsymbol{\beta}_j - \tilde{y}_{j,k_j} \tag{23}$$

$$\breve{d}^{-}_{j,k_j} = \left[\begin{array}{ccc} 1 & \overline{s_{k_j}} & \overline{q_{k_j}} \end{array} \right] \boldsymbol{\beta}_j - \tilde{y}_{j,k_j} \tag{24}$$

This relaxed regression problem is (non-strictly) convex in the parameters $\boldsymbol{\beta}_j$ and is thus solved efficiently to a global optimum. The problem is also closely related to interval regression (e.g., Inuiguchi and Tanino, 2006). More importantly, its solution leads to a lower bound for h_j. This is proven in 2.3.5. Mathematically, one can write that:

$$\underline{h_j}\big(\mathcal{T}, \breve{\boldsymbol{\beta}}_j(\mathcal{T})\big) \leq h_j\big(\boldsymbol{\theta}, \hat{\boldsymbol{\beta}}_j\big), \quad \forall \boldsymbol{\theta} \in \mathcal{T}. \tag{25}$$

In words, the minimum objective function value for the relaxed regression problem, $\underline{h_j}\big(\mathcal{T}, \breve{\boldsymbol{\beta}}_j(\mathcal{T})\big)$, is lower than or equal to any value that can be obtained for $h_j(\boldsymbol{\theta})$ with any parameter vector $\boldsymbol{\theta} \in \mathcal{T}$.

Combining the relaxed regressions (25), $j = 1, \ldots, J$, one can now write:

$$\begin{aligned}\underline{h}(\mathcal{T}) &= \underline{h}\big(\mathcal{T}, \breve{\boldsymbol{\beta}}_1(\mathcal{T}), \ldots, \breve{\boldsymbol{\beta}}_J(\mathcal{T})\big) \\ &= \sum_{j=1}^{J} \underline{h_j}\big(\mathcal{T}, \breve{\boldsymbol{\beta}}_j(\mathcal{T})\big) \leq \sum_{j} h_j\big(\boldsymbol{\theta}, \hat{\boldsymbol{\beta}}_j(\boldsymbol{\theta})\big) = h(\boldsymbol{\theta})\end{aligned} \tag{26}$$

Thus, computing $\underline{h}(\mathcal{T})$ with the procedure described above gives a provable lower bound to the objective function given the set of feasible parameter vectors $\mathcal{T}$. This completes the description of the optimization method used to find globally optimal values for $\boldsymbol{\theta}$.

In Fig. 2 one can see that the computation of the bounds is possible by only considering the two extremal parameter vectors $\underline{\boldsymbol{\theta}}$ and $\overline{\boldsymbol{\theta}}$ in the considered set $\mathcal{T}$. This is thanks to the particular choice for the upper bound procedure. It has the practical advantage that only two integrations of the ODE in (10) are necessary per considered set of parameter values: one to compute values for $\underline{\boldsymbol{s}_{k_j}}$ and one to compute $\overline{\boldsymbol{s}_{k_j}}$.

2.3.5. Relaxation of weighted least-squares regression

Let $\boldsymbol{X}_{K\times M}$ be an input matrix to a constrained WLS regression problem. Furthermore, let the elements of $\boldsymbol{X}$ be functions of some parameters $\boldsymbol{\theta}$. Consider then that the following holds

$$\underline{\boldsymbol{X}} = \underline{\boldsymbol{X}}(\mathcal{T}) \leq \boldsymbol{X} = \boldsymbol{X}(\boldsymbol{\theta}) \leq \overline{\boldsymbol{X}} = \overline{\boldsymbol{X}}(\mathcal{T}), \quad \forall \boldsymbol{\theta} \in \mathcal{T}. \tag{27}$$

Let $\tilde{y}_k$ be measurements ($k = 1, \ldots, K$) and $\boldsymbol{\beta}$ a vector of parameters of size $N \times 1$. Define the constrained WLS regression problem as

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \arg\min_{\boldsymbol{\beta}\in\Omega} h_o(\boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_{k=1}^{K} \left(\frac{d_k}{\sigma_k}\right)^2 \tag{28}$$

subject to

$$d_k = \boldsymbol{X}_{k,\cdot}(\boldsymbol{\theta}) \cdot \boldsymbol{\beta} - \tilde{y}_k, \tag{29}$$

where:

$$\forall \boldsymbol{\beta} : \{\boldsymbol{\beta} | \boldsymbol{\beta} \in \Omega\} \Rightarrow \boldsymbol{\beta} \geq 0 \tag{30}$$

A globally optimal solution to the constrained WLS problem (28) is obtained when $\boldsymbol{X}$, $\tilde{\boldsymbol{y}}$, and $\boldsymbol{\theta}$ are given.

Since the regression problem is quadratic in the residuals, it can be rewritten equivalently by expressing the residuals as $d_k = d_k^+ + d_k^-$ such that

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \arg\min_{\boldsymbol{\beta}\in\Omega} h_o(\boldsymbol{\theta}, \boldsymbol{\beta}) \tag{31}$$

subject to

$$h(\boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_{k=1}^{K} \left(\frac{d_k^+ + d_k^-}{\sigma_k} \right)^2 \tag{32}$$

$$d_k^+ = \max\left\{ \boldsymbol{X}_{k,\cdot}(\boldsymbol{\theta}) \cdot \boldsymbol{\beta} - \tilde{y}_k, 0 \right\} \tag{33}$$

$$d_k^- = \min\left\{ \boldsymbol{X}_{k,\cdot}(\boldsymbol{\theta}) \cdot \boldsymbol{\beta} - \tilde{y}_k, 0 \right\}. \tag{34}$$

The equivalence follows from the fact that $d_k^+ \cdot d_k^- = 0, k = 1, \ldots, K$.

Lower bound. In order to find a lower bound to h_o for all values $\boldsymbol{\theta} \in \mathcal{T}$, the upper and lower bounds to $\boldsymbol{X}$ are used in the following relaxed regression problem

$$\breve{\boldsymbol{\beta}}(\mathcal{T}) = \arg\min_{\boldsymbol{\beta}\in\Omega} \underline{h_o}(\mathcal{T}, \boldsymbol{\beta}) = \sum_{k=1}^{K} \left(\frac{\breve{d}_k^+ + \breve{d}_k^-}{\sigma_k} \right)^2 \tag{35}$$

subject to

$$\breve{d}_k^+ = \max\left\{ \underline{\boldsymbol{X}_{k,\cdot}}(\mathcal{T}) \cdot \boldsymbol{\beta} - \tilde{y}_k, 0 \right\} \tag{36}$$

$$\breve{d}_k^- = \min\left\{ \overline{\boldsymbol{X}_{k,\cdot}}(\mathcal{T}) \cdot \boldsymbol{\beta} - \tilde{y}_k, 0 \right\}. \tag{37}$$

with $\breve{d}_k^+$ and $\breve{d}_k^+$ relaxed residuals. This problem is non-strictly convex and can thus be solved to global optimality with local optimization algorithms.

Theorem 2.1. *Let $\boldsymbol{\theta} \in \mathcal{T}$ and let the regression problems $\breve{\boldsymbol{\beta}}(\mathcal{T})$ and $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ be defined as above. Then, the objective function $\underline{h_o}(\mathcal{T}, \breve{\boldsymbol{\beta}}(\mathcal{T}))$, evaluated with the relaxed regression problem $\breve{\boldsymbol{\beta}}(\mathcal{T})$, is a lower bound to $h_o(\boldsymbol{\theta}, \boldsymbol{\beta}(\boldsymbol{\theta}))$ such that*

$$\underline{h_o}(\mathcal{T}, \breve{\boldsymbol{\beta}}(\mathcal{T})) \leq h_o(\boldsymbol{\theta}, \boldsymbol{\beta}(\boldsymbol{\theta})), \quad \forall \boldsymbol{\theta} \in \mathcal{T},\ \forall \boldsymbol{\beta} \in \Omega. \tag{38}$$

Proof. From (28) and (35), we know that $\hat{\boldsymbol{\beta}}(\theta)$ is a global minimizer of $h_o(\boldsymbol{\theta}, \boldsymbol{\beta})$ and that $\breve{\boldsymbol{\beta}}(\mathcal{T})$ is a global minimizer of $\underline{h_o}(\mathcal{T}, \boldsymbol{\beta})$, that is

$$h_o(\theta, \hat{\boldsymbol{\beta}}(\theta)) \leq h_o(\theta, \boldsymbol{\beta}),\ \forall \boldsymbol{\theta} \in \mathcal{T},\ \forall \boldsymbol{\beta} \in \Omega \tag{39}$$

$$\underline{h_o}(\mathcal{T}, \breve{\boldsymbol{\beta}}(\mathcal{T})) \leq \underline{h_o}(\mathcal{T}, \boldsymbol{\beta}),\ \forall \boldsymbol{\beta} \in \Omega. \tag{40}$$

It follows from (30), (33)-(34) and (36)-(37) that for every feasible $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ the magnitudes of the relaxed residuals are smaller than or equal to the magnitudes of the original residuals. We have

$$0 \leq \breve{d}_k^+ \leq d_k^+ \text{ and } d_k^- \leq \breve{d}_k^- \leq 0, \quad \forall \boldsymbol{\theta} \in \mathcal{T},\ \forall \boldsymbol{\beta} \in \Omega,\ k = 1, \ldots, K. \tag{41}$$

From (31)-(37) and (41), it follows that

$$\underline{h_o}(\mathcal{T}, \boldsymbol{\beta}) \leq h_o(\boldsymbol{\theta}, \boldsymbol{\beta}), \quad \forall \boldsymbol{\theta} \in \mathcal{T},\ \forall \boldsymbol{\beta} \in \Omega. \tag{42}$$

The same holds when replacing $\boldsymbol{\beta}$ with the constrained WLS solution $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ given by (31)-(34)

$$\underline{h_o}(\mathcal{T}, \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})) \leq h_o(\boldsymbol{\theta}, \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})), \quad \forall \boldsymbol{\theta} \in \mathcal{T}. \tag{43}$$

Finally, combining (39), (40), and (43), one can write:

$$\underline{h_o}(\mathcal{T}, \breve{\boldsymbol{\beta}}(\mathcal{T})) \leq \underline{h_o}(\mathcal{T}, \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})) \leq h_o(\boldsymbol{\theta}, \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})) \leq h_o(\boldsymbol{\theta}, \boldsymbol{\beta}), \quad \forall \boldsymbol{\theta} \in \mathcal{T},\ \forall \boldsymbol{\beta} \in \Omega \tag{44}$$

which proves the theorem. □

2.3.6. Improved relaxation of weighted least-squares regression

The following development is added for reasons of completeness despite not being used to generate the reported optimization results. The reasons for this are explained in the Discussion section.

An improved lower bound can be found by solving the following relaxed WLS problem:

$$\mathring{\boldsymbol{\beta}}(\mathcal{T}),\ \mathring{\boldsymbol{X}}(\mathcal{T}) = \arg \min_{\substack{\boldsymbol{\beta} \in \Omega \\ \boldsymbol{X} \in \Gamma}} \mathring{h}_o(\mathcal{T}, \boldsymbol{\beta}, \boldsymbol{X}) \tag{45}$$

subject to

$$\mathring{h}_o(\mathcal{T}, \boldsymbol{\beta}, \boldsymbol{X}) = \sum_{k=1}^{K} \left(\frac{\mathring{d}_k}{\sigma_k} \right)^2 \tag{46}$$

with the relaxed residuals

$$\mathring{d}_k = \boldsymbol{X}_{k,\cdot}(\boldsymbol{\theta})\ \boldsymbol{\beta} - \tilde{y}_k \tag{47}$$

In the above, $\Gamma(\mathcal{T})$ is the feasible set for $\boldsymbol{X}$ which is defined as follows:

$$\boldsymbol{X} \in \Gamma(\mathcal{T}) \Leftrightarrow \begin{cases} \underline{\boldsymbol{X}_{k,\cdot}}(\mathcal{T}) \leq \boldsymbol{X}_{k,\cdot}(\boldsymbol{\theta}) \leq \overline{\boldsymbol{X}_{k,\cdot}}(\mathcal{T}) \\ \overline{\boldsymbol{B}(\mathcal{T})} \cdot \mathbf{vec}\,(\boldsymbol{X}) \leq \boldsymbol{c}(\mathcal{T}) \end{cases} \quad , \quad \forall \boldsymbol{\theta} \in \mathcal{T} \tag{48}$$

with $\boldsymbol{B}(\mathcal{T})$ and $\boldsymbol{c}(\mathcal{T})$ describing constraints for estimates of the elements of $\boldsymbol{X}$. In the studied example, the profile of the state $s(t)$ is known to be monotonically decreasing with time. In addition, it is guaranteed to have a concave shape as long as $s(t) \geq \sqrt{\overline{\theta_1}/\underline{\theta_3}}$ (without proof). Similarly, the profile of $s(t)$ is guaranteed convex when $s(t) \leq \sqrt{\underline{\theta_1}/\overline{\theta_3}}$ (without proof). Applying such prior knowledge leads to shape constraints for s_{k_j} and q_{k_j} which are expressed with $\boldsymbol{B}(\mathcal{T})$ and $\boldsymbol{c}(\mathcal{T})$. The proposed relaxation is pseudo-convex for all-positive values for $\boldsymbol{\beta}$ and $\boldsymbol{X}$. (45)-(48) can be solved to convergence straightforwardly by means of (constrained) alternating least-squares, i.e. by iterating between optimization of $\boldsymbol{\beta}$ given the best known values for $\boldsymbol{X}$ and optimization of $\boldsymbol{X}$ given the best known $\boldsymbol{\beta}$. Each of these optimization problems is a QP. This relaxation delivers a tighter lower bound than the one discussed in 2.3.5:

$$\underline{h_o}(\mathcal{T}, \breve{\boldsymbol{\beta}}(\mathcal{T})) \leq \mathring{h}_o(\mathcal{T}, \mathring{\boldsymbol{\beta}}(\mathcal{T}), \mathring{\boldsymbol{X}}(\mathcal{T})) \leq h_o(\boldsymbol{\theta}, \boldsymbol{\beta}),\ \forall \boldsymbol{\theta} \in \mathcal{T},\ \forall \boldsymbol{\beta} \in \Omega \tag{49}$$

This is given here without proof.

2.3.7. Conventional optimization

The performance of the proposed deterministic optimization scheme is compared with a more conventional approach based on the quasi-newton algorithm. This algorithm is commonly applied for nonlinear optimization despite the risk of finding a local optimum (Jakeman et al., 2006; Rieger et al., 2012) or getting stuck in saddle-points (Dauphin et al., 2014). To

circumvent such problems it is common to start this algorithm from multiple initial estimates for the parameters. In this work, these initial estimates are obtained by uniform gridding in $\mathcal{T}$. The number of points on the grid is the same along each dimension and was set so that the computational effort, measured in terms of total batch process simulations, roughly matches the number of simulations executed during deterministic optimization. By means of this uniform sampling, the same level of prior ignorance is assumed for both the deterministic and conventional optimization method.

2.4. Application

The following paragraphs describe the developments specific to the application study used for demonstration of the proposed optimization method.

2.4.1. Experimental data

The experimental data used to demonstrate the parameter optimization method were collected for a batch respirometric experiment executed as follows. A laboratory-scale continuous-flow stirred tank reactor for biological urine nitrification was operated under aerobic conditions by means of a bang-bang oxygen controller (a.k.a. on-off controller, Levine (1996)) switching the aeration on (off) when measuring 6.0 mgO_2/L (6.2 mgO_2/L) with an optical oxygen sensor (WTW: TriOxmatic 700, without salinity correction). The pH level was left uncontrolled and the pH measurements (Mettler Toledo: 405-DXK-S8/225) remained between 5.66 and 5.68. Such a stable pH is currently explained by *(i)* direct inhibitory effects of the pH on the biological ammonia oxidation process (Fumasoli et al., 2015), *(ii)* complete consumption of the available inorganic carbon, and *(iii)* negligible net effects of other processes affecting the pH. Prior to the experiment, endogenous respiration conditions were achieved by stopping the inlet flow and attending the halting of the exogenous oxidation processes, in particular the oxidation of organic matter, ammonia, and nitrite. The mixed-liquor volume was 5.89 ± 0.10 L during the experiment. The complete experiment consists of four pulse additions of a nitrite stock solution, each time awaiting endogenous respiration conditions before adding the next pulse. Endogenous respiration conditions and a laboratory TNN concentration level measurement below the detection limit were awaited before each pulse addition. As the developed method is particularly suitable to single-pulse batch experiments, only the data regarding the last pulse is used in this study. At the start of this part of the experiment ($t = t_0 = 0$) a pulse of a nitrite stock solution was added

$(20 \pm 0.02\text{mL}, 5970 \pm 10\text{mgN/L})$ which increased the total nitrite nitrogen (TNN) concentration (S_{TNN}) in the reactor to 20.2 ± 0.5 mgN/L. The reactor was operated without feeding until endogenous respiration conditions could be recognized and a laboratory TNN concentration level measurement was below the detection limit. During the experiment, seven samples were taken at distinct times ($t_{k_{\text{TNN}}} = t_{k_1}$) to measure the TNN concentration ($\tilde{y}_{\text{TNN},k_{\text{TNN}}} = \tilde{y}_{1,k_1}$, $k_{\text{TNN}} = k_1 = 1, \ldots, K_1$, $K_{\text{TNN}} = K_1 = 7$). Each of these samples were filtered using micro-glass fibre paper (0.45μm, MGF, Munktell Filter AB, Falun, Sweden), diluted (sample 3: 1/25, all others: 1/20), and analyzed with colorimetric cuvette tests (LCK342 (samples 1-2) and LCK341 (samples 3-7), Hach-Lange, Berlin, Germany). Automated dissolved oxygen (DO) concentration (S_{O_2}) readings were collected throughout the experiment at a sampling interval of 5 seconds. These DO concentration measurements were corrected for salinity and processed to compute the oxygen uptake rate (OUR) as the slope of the line fitted in the least-squares sense to the linear segment of the DO concentration profiles obtained in each of the non-aerated periods. The $K_{\text{OUR}} = K_2 = 50$ OUR measurements and sampling times are further referred to as $\tilde{y}_{\text{OUR},k_{\text{OUR}}} = \tilde{y}_{2,k_2}$ and $t_{\text{OUR},k_{\text{OUR}}} = t_{2,k_2}$, where $k_{\text{OUR}} = k_2 = 1, \ldots, K_2$.

2.4.2. Proposed model structure

For the purpose of modeling, focus is given to the biological oxidation of nitrous acid (HNO_2) by means of the nitrite oxidizing bacteria (NOB):

$$\text{HNO}_2 + i_{\text{growth}}\ \text{O}_2 \rightarrow \text{HNO}_3 \tag{50}$$

with i_{growth} the stoichiometric coefficient for oxygen. When ignoring the need for oxygen for biomass internalization this coefficient equals 1/2 exactly. Here, we however assume the precise oxygen requirements are not known. This also allows accounting for potential deviations in the oxygen measurement, particular due to a deviation of the sensor's sensitivity. As such, i_{growth} becomes a parameter which lumps the stoichiometric requirement for the nitrite oxidation, the oxygen demand for biomass internalization during growth, and imperfections of the oxygen sensor together. Additional reactions include the acid-base reactions for the nitrite and nitrate species which influence the availability of the substrate HNO_2 and are assumed to

be in equilibrium at all times:

$$\mathrm{HNO_2} \rightleftarrows \mathrm{H^+ + NO_2^-} \tag{51}$$

$$\mathrm{HNO_3} \rightleftarrows \mathrm{H^+ + NO_3^-} \tag{52}$$

Because of the low pK_a value for the nitrate acid-base reaction (52) ($pK_{\mathrm{a,HNO_3}} \sim -1.4$) one can safely assume that for our study all nitrate is present in its ionic form ($\mathrm{NO_3^-}$) (at pH=5.66-5.68) and thus has little influence on the available $\mathrm{HNO_2}$ concentration. Furthermore, we consider that the pH is stable enough to assume it to be constant. The TNN concentration (S_{TNN}) is the sum of the nitrite and nitric acid concentrations, i.e. $S_{\mathrm{TNN}} = [\mathrm{HNO_2}] + [\mathrm{NO_2^-}]$. Given the stable pH, the experiment cannot be used to determine whether molecular nitrous acid or nitrite ions function as the substrate for the NOB. Here, we assume that nitrous acid is the substrate. Its concentration can be expressed as a linear function of S_{TNN}:

$$[\mathrm{HNO_2}] = \frac{S_{\mathrm{TNN}}}{1 + 10^{pH - pK_{\mathrm{a,HNO_2}}}}. \tag{53}$$

This allows us to formulate our model in terms of S_{TNN}. Given the above description and assumptions, the following model is cast to describe the dynamic behavior of a batch experiment (Henze et al., 2000):

$$\dot{S}_{\mathrm{TNN}}(t) = -\frac{1}{Y_{\mathrm{NOB}}}\ \mu(t)\ X_{\mathrm{NOB}}(t) \tag{54}$$

$$\dot{S}_{\mathrm{O_2}}(t) = \frac{r_{\mathrm{aer}}(t)}{V} - i_{\mathrm{decay}}\ b_{\mathrm{NOB}}\ X_{\mathrm{NOB}}(t) - \frac{i_{\mathrm{growth}}}{Y_{\mathrm{NOB}}}\ \mu(t)\ X_{\mathrm{NOB}}(t) \tag{55}$$

$$\dot{X}_{\mathrm{NOB}}(t) = \mu(t)\ X_{\mathrm{NOB}}(t) - b_{\mathrm{NOB}}\ X_{\mathrm{NOB}}(t) \tag{56}$$

with $S_{\mathrm{TNN}}(t)$ the TNN concentration, $S_{\mathrm{O_2}}(t)$ the DO concentration, $X_{\mathrm{NOB}}(t)$ the bacterial mass, $\mu(\cdot)$ the specific growth rate, and $r_{\mathrm{aer}}(t)$ the oxygen mass transfer rate. All remaining elements of the above equations are parameters and are listed in Table 2.

Earlier work (Hellinga et al., 1999; Jubany, 2007) has suggested that the specific growth rate μ of NOB can be adequately described by means of Haldane kinetics, which include substrate affinity and inhibition effects in the rate expression (Andrews, 1968):

$$\mu(t) = \mu_{\max}\ \frac{S_{\mathrm{TNN}}}{K_{\mathrm{S}} + S_{\mathrm{TNN}} + {S_{\mathrm{TNN}}}^2/K_{\mathrm{I}}} \tag{57}$$

313 with parameters defined in Table 2.

In addition, the effect of aeration can be omitted from the equations by replacing equation (55) with the description of the oxygen uptake rate r_{OUR}:

$$r_{\mathrm{OUR}}(t) = i_{\mathrm{decay}}\ b_{\mathrm{NOB}}\ X_{\mathrm{NOB}}(t) + \frac{i_{\mathrm{growth}}}{Y_{\mathrm{NOB}}}\ \mu(t)\ X_{\mathrm{NOB}}(t). \tag{58}$$

Given that the active nitrifying biomass concentration cannot be measured directly, the parameters i_{decay}, i_{growth}, and Y_{NOB} cannot be identified simultaneously (lack of structural identifiability). To address this, one can assume to know one of these parameters, e.g. based on values in the literature. In this work, we instead assume that the net growth of the biomass is negligible during the batch experiment, which translates to X_{NOB} being a constant. Define the biomass activity $a(\cdot)$ as the substrate degradation rate:

$$a(t) = \frac{X_{\mathrm{NOB}}}{Y_{\mathrm{NOB}}}\ \mu(t) = a_{\max}\ \frac{S_{\mathrm{TNN}}}{K_{\mathrm{S}} + S_{\mathrm{TNN}} + {S_{\mathrm{TNN}}}^2/K_{\mathrm{I}}}, \tag{59}$$

where $a_{\max} = \mu_{\max}\ X_{\mathrm{NOB}}/Y_{\mathrm{NOB}}$. Then the dynamic behavior of the equations (54)-(58) can be reduced to the following set of two equations which are a single ordinary differential equation (ODE) and a linear equation in the reaction rate:

$$\dot{S}_{\mathrm{TNN}}(t) = -a(t) \tag{60}$$

$$r_{\mathrm{OUR}}(t) = r_{\mathrm{endo}} + i_{\mathrm{growth}}\ a(t) \tag{61}$$

with the initial condition

$$S_{\mathrm{TNN}}(0) = S_{\mathrm{TNN},0} \tag{62}$$

314 and $r_{\mathrm{endo}} = i_{\mathrm{decay}}\ b_{\mathrm{NOB}}\ X_{\mathrm{NOB}}$.

The data obtained during the experiment described above are given as vectors of measurements of the TNN and the OUR. As above, the measurement errors e_{1,k_1} and e_{2,k_2} are assumed to be sampled independently from zero mean normal distributions with known standard deviations σ_{1,k_1} and σ_{2,k_2}. The measurements are thus described as

$$\tilde{y}_{\mathrm{TNN},k_{\mathrm{TNN}}} = \tilde{y}_{1,k_1} = y_{1,k_1} + e_{1,k_1}, \quad y_{1,k_1} := S_{\mathrm{TNN}}(t_{k_1}), \quad k_1 = 1, \ldots, K_1 \tag{63}$$

$$\tilde{y}_{\mathrm{OUR},k_{\mathrm{OUR}}} = \tilde{y}_{2,k_2} = y_{2,k_2} + e_{2,k_2}, \quad y_{2,k_2} := r_{\mathrm{OUR}}(t_{k_2}), \quad k_2 = 1, \ldots, K_2. \tag{64}$$

315 *2.4.3. Model reformulation*

316 The full model is now described by equations (59)-(64). As the remain-
317 ing parameters are unknown, the parameter estimation problem consists of
318 finding optimal values for $a_{\max}$, K_{S}, K_{I}, $S_{\mathrm{TNN},0}$, i_{growth}, and r_{endo}. To ap-
319 ply the parameter optimization method described above this model is now
320 refomulated to match the form in (1)-(5).

The ODE defined above is nondimensionalized by writing it in terms of the relative substrate concentration $s(t)$. This is defined as

$$s(t) = \frac{S_{\mathrm{TNN}}(t)}{S_{\mathrm{TNN},0}}. \tag{65}$$

A new complete description of the process is then:

$$\dot{s}(t) = -a_{\max}\ \frac{s(t)}{K_{\mathrm{S}} + S_{\mathrm{TNN},0}\ s(t) + \frac{{S_{\mathrm{TNN},0}}^2}{K_{\mathrm{I}}}\ s(t)^2} \tag{66}$$

$$s(0) := s_0 = 1. \tag{67}$$

$$S_{\mathrm{TNN}} = S_{\mathrm{TNN},0}\ s(t) \tag{68}$$

$$r_{\mathrm{OUR}}(t) = r_{\mathrm{endo}} - i_{\mathrm{growth}}\ S_{\mathrm{TNN},0}\ \dot{s}(t) \tag{69}$$

To simplify the equations, we define the expression $q(\cdot)$ as the relative reaction rate:

$$q(s) = a_{\max}\ \frac{s}{K_{\mathrm{S}} + S_{\mathrm{TNN},0}\ s + \frac{{S_{\mathrm{TNN},0}}^2}{K_{\mathrm{I}}}\ s^2} \tag{70}$$

and the nonnegative parameters $\beta_{1,1}, \beta_{2,0}, \beta_{2,2}, \theta_1, \theta_2, \theta_3$ such that

$$\boldsymbol{\beta}_1 = \begin{bmatrix} 0 & \beta_{1,1} & 0 \end{bmatrix}^T \tag{71}$$

$$\boldsymbol{\beta}_2 = \begin{bmatrix} \beta_{2,0} & 0 & \beta_{2,2} \end{bmatrix}^T \tag{72}$$

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 \end{bmatrix}^T. \tag{73}$$

with the following equivalence relations:

$$\begin{aligned} \boldsymbol{\gamma} &= \begin{bmatrix} a_{\max} & K_{\mathrm{S}} & K_{\mathrm{I}} & S_{\mathrm{TNN},0} & r_{\mathrm{endo}} & i_{\mathrm{growth}} \end{bmatrix}^T \\ &= \begin{bmatrix} \frac{\beta_{1,1}}{\theta_2} & \frac{\theta_1\ \beta_{1,1}}{\theta_2} & \frac{\beta_{1,1}\ \theta_2}{\theta_3} & \beta_{1,1} & \beta_{2,0} & \frac{\beta_{2,2}}{\beta_{1,1}} \end{bmatrix}^T \end{aligned} \tag{74}$$

As a result, the ODE is now:

$$\dot{s}(t) = -q\big(s(t), \boldsymbol{\theta}\big) = -\frac{s(t)}{\theta_1 + \theta_2\ s(t) + \theta_3\ s(t)^2}, \qquad s(0) := s_0 = 1. \tag{75}$$

and the measurement equations can be rewritten as:

$$\tilde{y}_{1,k_1} = y_{1,k_1} + e_{1,k_1}, \quad y_{1,k_1} = \beta_{1,1}\ s(t_{k_1}), \qquad k_1 = 1, \ldots, K_1 \tag{76}$$

$$\tilde{y}_{2,k_2} = y_{2,k_2} + e_{2,k_2}, \quad y_{2,k_2} = \beta_{2,0} + \beta_{2,2}\ q\big(s(t), \boldsymbol{\theta}\big), \quad k_2 = 1, \ldots, K_2. \tag{77}$$

This matches the form of (1)-(4) except that in the proposed model several values in the vectors $\boldsymbol{\beta}_j$ are known to be zero, i.e. $\beta_{1,0} := 0$, $\beta_{1,2} := 0$, $\beta_{2,1} := 0$. The implicit assumptions are *(i)* that the TNN measurements are measured without an offset and *(ii)* that the measurements are affected by only one measured variable (i.e., no mixing effects in the measurements). Therefore, the subsets Ω_j are defined as follows:

$$\boldsymbol{\beta}_1 \in \Omega_1 \Leftrightarrow \beta_{1,0} = 0 \wedge \beta_{1,1} \geq 0 \wedge \beta_{1,2} = 0 \tag{78}$$

$$\boldsymbol{\beta}_2 \in \Omega_2 \Leftrightarrow \beta_{2,0} \geq 0 \wedge \beta_{2,1} = 0 \wedge \beta_{2,2} \geq 0 \tag{79}$$

These constraints define a polyhedral set. It is also verified easily that the expression for $q\big(s(t), \boldsymbol{\theta}\big)$ satisfies the requirements in (2). This is thanks to the specific choices made in (70) and (74). As a consequence, all requirements for the optimization method are satisfied. Interestingly, the reformulated model contains six parameters of which three are available through WLS regression. This means that the branch-and-bound algorithm needs to operate in only three dimensions to find the values of $\boldsymbol{\theta} = [\theta_1\ \theta_2\ \theta_3]^{\mathrm{T}}$.

A scheme of the procedures to compute these bounds is given in Fig. 2. This scheme reflects the special structure of the measurement equations (i.e., no mixing effects), as the lower bound for h_1 (h_2) only depends on the bounds for s (q).

The only aspect of the optimization method left unattended is how bounds for the rate expressions (20) can be obtained. We assume that the intervals $\left[\underline{s_{k_j}}, \overline{s_{k_j}}\right]$ are computed already. As mentioned before, lower (upper) bounds for the rate expression are obtained with $\overline{\boldsymbol{\theta}}$ ($\underline{\boldsymbol{\theta}}$). Thanks to the pseudo-concave property of the rate expression (75), the lower bound for the relative reaction rate is easily obtained as the minimum of two evaluations of q obtained at the extrema of the given feasible interval for s. Mathematically one

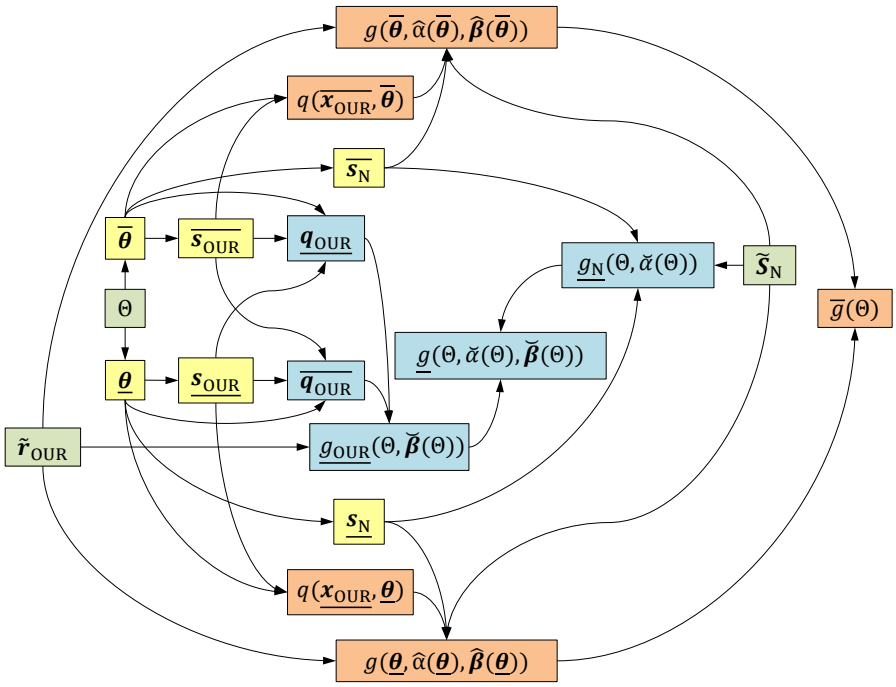

Figure 2: A schematic view of the bounding procedures. The inputs to the procedures are shown in green. The steps exclusively required for the upper (lower) bound are indicated in red (blue) and the steps used for both the upper and lower bound are indicated in yellow. The choice made for the upper bounding procedures means that only two ODE integrations are necessary, namely one time to compute the values for $\overline{s_{k_1}}$ and $\overline{s_{k_2}}$ and one time for $\underline{s_{k_1}}$ and $\underline{s_{k_2}}$.

339 obtains:

$$\underline{q_{k_j}} \quad = \quad \min\left\{q\left(\underline{s_{k_j}}, \overline{\boldsymbol{\theta}}\right), q\left(\overline{s_{k_j}}, \overline{\boldsymbol{\theta}}\right)\right\}. \tag{80}$$

To compute the upper bound, three distinct situations can occur as (70) has a unique (local and global) maximum within its domain, its location denoted as $s_{\max}$. First, $s_{\max}$ can lie left of $\underline{s_{k_j}}$, in which case the upper bound is found at $s = \underline{s_{k_j}}$. Secondly, $s_{\max}$ can lie within the interval $\left[\underline{s_{k_j}}, \overline{s_{k_j}}\right]$, in which case the upper bound is found at $s = s_{\max}$. Lastly, $s_{\max}$ can lie right of $\overline{s_{k_j}}$, in

which case the upper bound is found at $\overline{s_{k_j}}$. Mathematically, one writes:

$$\overline{q_{k_j}} = \begin{cases} q\big(\underline{s_{k_j}}, \underline{\boldsymbol{\theta}}\big) & s_{\max} \leq \underline{s_{k_j}}, \\ q\big(s_{\max}, \underline{\boldsymbol{\theta}}\big) & \underline{s_{k_j}} \leq s_{\max} \leq \overline{s_{k_j}}, \\ q\big(\overline{s_{k_j}}, \underline{\boldsymbol{\theta}}\big) & \overline{s_{k_j}} \leq s_{\max} \end{cases} \tag{81}$$

Fortunately, the location of the relevant maximum, $s_{\max}$, is available analytically given $\underline{\boldsymbol{\theta}}$ so that the upper bound computations remain fairly simple:

$$s_{\max} = \arg\max_s q\big(s, \underline{\boldsymbol{\theta}}\big) = \sqrt{\underline{\theta_1} \Big/ \underline{\theta_3}}. \tag{82}$$

2.5. Data and Software Availability

All computations were executed with Matlab R2012b (8.0.0.783, win32) on a desktop machine (CPU: Intel$^{\mathrm{R}}$ Core$^{\mathrm{TM}}$ i7-3770K 3.50 GHz, RAM: 8.00 GB, OS: Windows 7 Enterprise, Service Pack 1). To solve the nonlinear optimization problem we used the Spike_O toolbox (v1.1), an open-source package for deterministic optimization (Villez et al., 2013; Villez and Habermacher, 2016; Villez et al., 2016) in Matlab/Octave. All QPs were solved with the MOSEK optimization software (Version 7.1.0.30). All additional data and software used to generate the obtained results are released publicly under the GNU GPL license (Version 3) and provided in the Supplementary Information.

3. Results

In what follows, we first demonstrate the computation of the provided bounds. Thereafter, the basic result of the applied optimization strategy is shown. Lastly, additional results regarding the optimization algorithm are discussed.

3.1. Experimental data

In Fig. 3, the experimental data are shown. One can see that during the considered pulse experiment, the TNN concentration measurements ($\tilde{y}_{\mathrm{TNN},k_{\mathrm{TNN}}} = \tilde{y}_{1,k_1}$) decrease monotonically during the experiment from 19.22 mg N/L at 12' to 0.66 mg N/L at 3h12'. Simultaneously, the OUR is observed to increase first from about 155 mg O_2/L.h at the beginning of the experiment up to around 200 mg O_2/L.h at 2h30' in the experiment only

to decrease quickly afterwards to 55 mg N/L.h at the end of the experiment. The TNN measurements are assumed to be subject to measurement errors with a relative standard deviation of 2% with respect to the measured value ($\sigma_{\mathrm{TNN},k_{\mathrm{TNN}}} = \sigma_{1,k_1} = 2/100\ \tilde{y}_{1,k_1}$). The measurement error standard deviation for the OUR measurements ($\sigma_{\mathrm{OUR},k_{\mathrm{OUR}}} = \sigma_{2,k_2}$) is assumed to be constant and equal to 2 mg O_2/L.h. This is an educated guess based on the best fit obtained with a cubic spline function which is constrained to consist of increasing-convex, concave, and decreasing-convex segments (not shown, Villez et al., 2013). Fig. 3 shows the 3-σ bounds around the measurements.

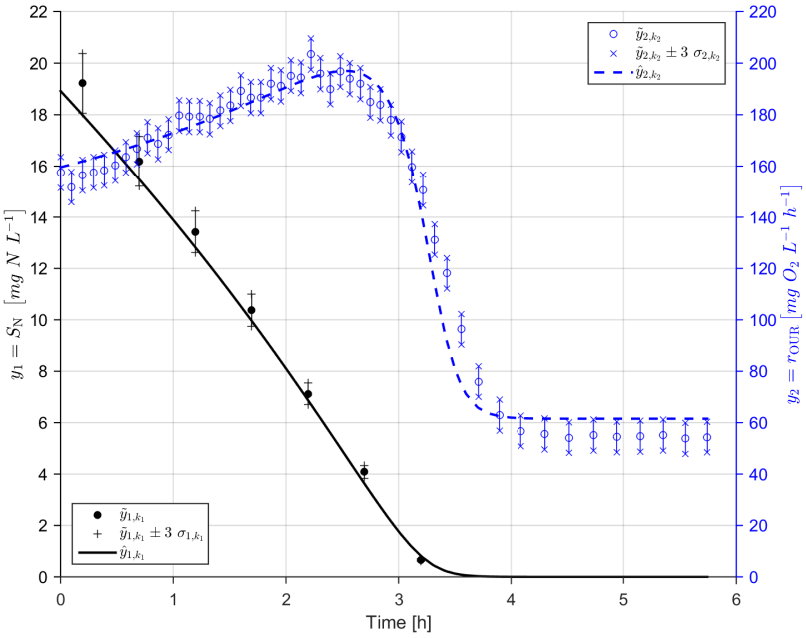

Figure 3: Data and best-fitting model simulation. Data are shown with 3-σ confidence intervals. The model simulation for the best-fitting values for $\boldsymbol{\theta}$ are shown with full and dashed lines.

3.2. Demonstrating the Bounds

The ranges for the parameter values $\boldsymbol{\theta}$ applied to demonstrate the bounds are given in Table 3. Using (18)-(19), one obtains the intervals $[\underline{s}(t), \overline{s}(t)]$ for the process state, s, at every time point during the experiment. The resulting

intervals are shown in Fig. 4. The intervals obtained at the measurement sampling times t_{1,k_1} and t_{2,k_2} are indicated as well ($\left[\underline{s_{1,k_1}}, \overline{s_{1,k_1}}\right]$ and $\left[\underline{s_{2,k_2}}, \overline{s_{2,k_2}}\right]$). A number of simulations corresponding to a regular grid of parameter vectors within the feasible set of parameter vectors are shown as well. One can see that the computed interval bounds effectively bound the obtained concentration values, thereby demonstrating this part of the bounding procedure. As far as understood, these bounds cannot be improved (i.e., these are the tightest bounds achievable).

Table 3: Considered set of parameter vectors ($\mathcal{T}$) to demonstrate the bounding procedures.)

Parameter (θ)	Lower bound ($\underline{\theta}$)	Upper bound ($\overline{\theta}$)
θ_1	0.050	0.45
θ_2	1.6	2.0
θ_3	1.6	2.0

With (20) and (80)-(82), values for $\underline{q}(t)$ and $\overline{q}(t)$ are obtained. The procedures for this are demonstrated in Fig. 5, which offers an intuitive insight into (80)-(82). One can clearly see how the position of the interval for the relative concentration affects the selected bounds for the reaction rate and corresponding relative concentrations. The figure also suggests that the bounds to the reaction rate are tight for a given concentration value. In Fig. 6, the profiles of relative concentrations corresponding to the lower and upper bounds to the relative reaction rate are shown in the top and bottom panel. One sees that the lower bound computations results in a discrete jump for the selected relative concentration as the concentration interval moves from high to low values. In contrast, the computation of the upper bound results in a constant section in the profile. The bounding intervals for the relative reaction rates are shown in the bottom panel together with simulated values for the reaction rates corresponding to the previously applied grid of feasible parameter vectors. The computed bounds appear valid as they bound the values obtained with random simulations. Unfortunately, the figure also suggests that these bounds are not tight (for the selected set of parameter vectors) as the white-space in the bottom panel between the most extremal simulation of the relative reaction rate and the proven bounds is fairly large, especially for the second half of the experiment.

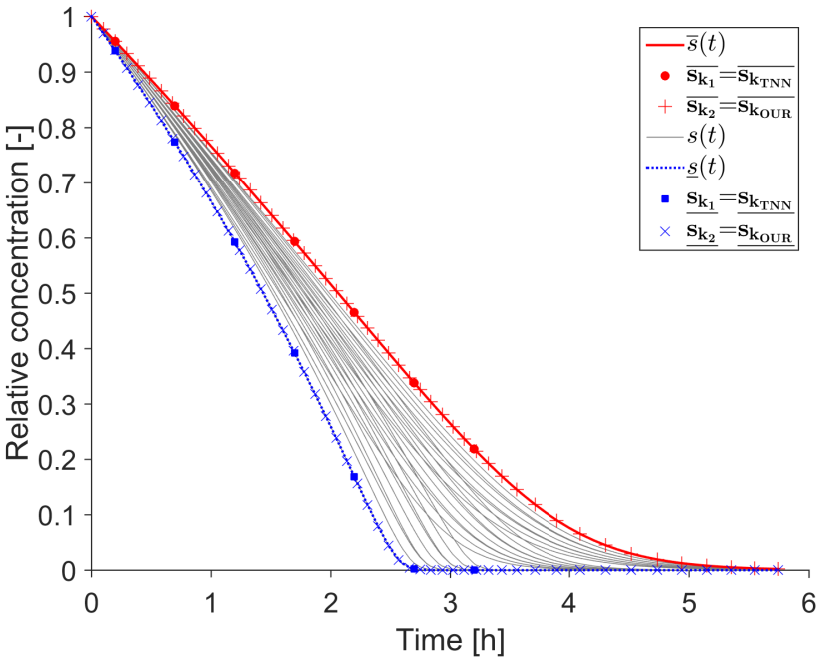

Figure 4: Bounding and selected time profiles of the relative concentration. Selected time profiles and upper and lower bounds for the relative concentrations at each time, corresponding to the lowest and highest values for $\boldsymbol{\theta}$.

The objective function term h_1 is bounded by means of solving three regression problems. To this end, a linear model is fitted to the concentration data by manipulating the value of $\beta_{1,1}$. The first two problems are used to compute values for h_1 corresponding to the parameter vectors $\underline{\boldsymbol{\theta}}$ and $\overline{\boldsymbol{\theta}}$ and by means of solving (11) accordingly. The objective function term h_1 and the corresponding minima for $\beta_{1,1}$ are shown in Fig. 7. For the purpose of demonstration, other quadratic objective function profiles are shown corresponding to previously selected values of $\boldsymbol{\theta}$ with the considered feasible set. The convex nature of these regression problems (for $\beta_{1,1} \in \mathbb{R}_0^+$) is visually confirmed by inspection of the figure. It is also visible that the selected upper bound value is higher than a number of minima obtained for other parameter vectors. The upper bound is thus not a tight one. This problem however disappears as the considered sets ($\mathcal{T}$) become smaller during branch-and-bound optimization (not shown). The last regression problem is the relaxed regres-

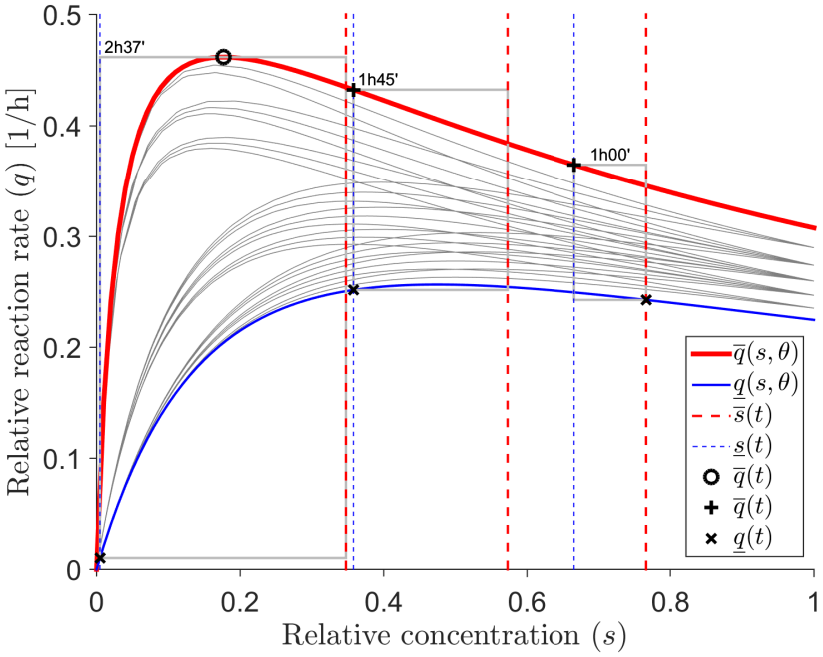

Figure 5: Bounding of the relative reaction rate. Dashed vertical lines indicate the intervals for the relative concentration at selected time instants (1h00', 1h45', and 2h37'). The grey rectangles, cross-hairs, and circle indicate the lower and upper bound values for the relative reaction rate for each of the selected time instants computed according to (80)-(82).

sion problem described in (21). The resulting objective value is also shown as function of $\beta_{1,1}$ and for the obtained value for $\beta_{1,1}$. One can easily see that the best objective function value for $\beta_{1,1}$ obtained through the relaxed regression problem delivers a lower bound to the objective function term h_1. The lower bound function exhibits an insensitive zone within which its value is constant and equal to zero. This effect disappears relatively quickly during branch-and-bound optimization (not shown). More important is that the gap between the lower bound function and the global minimum is rather large. Although this gap converges to zero as the volume of considered sets $\mathcal{T}$ is reduced, convergence of the lower bound is rather slow (see below). Similar observations are made for the bounds of the second objective function term h_2 in (13) (not shown).

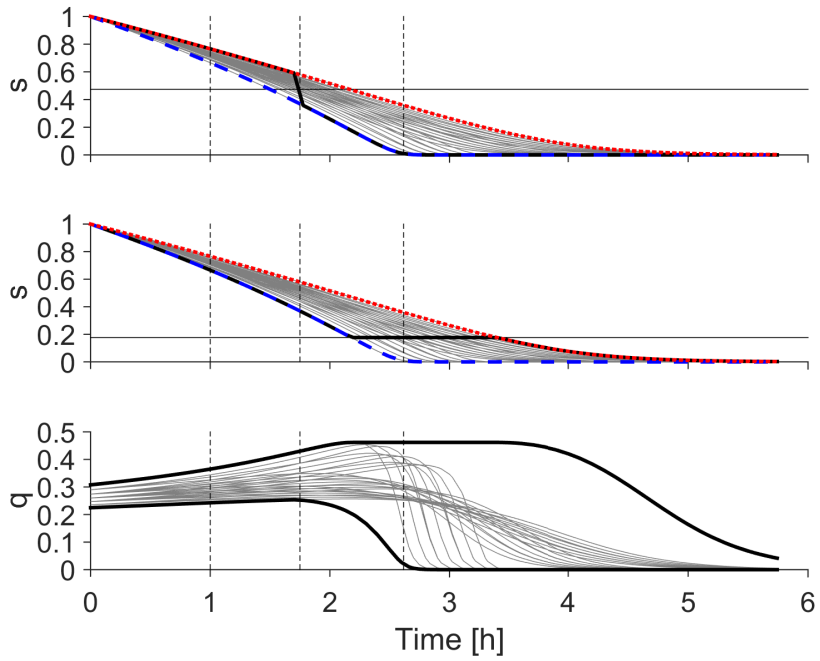

Figure 6: Computing bounds for the relative reaction rate. Top/Middle: Relative concentrations – Dashed lines: Bounds to the relative concentration, Full thick line: Relative concentration delivering the minimum (top) and maximum (middle) relative reaction rate, Full thin lines: Relative concentration profiles for selected parameter vectors within the considered set; Bottom: Relative reaction rates – Full thick lines: Bounds to the relative reaction rate, Full thin lines: Relative reaction rate profiles for selected parameter vectors within the considered set. In all panels, dashed vertical lines indicate previously selected time instants for use in Fig. 5 (1h00', 1h45', and 2h37'). The bottom panel suggests that the bounds on the relative reaction rate are not tight.

3.3. Parameter estimation

The feasible root set ($\mathcal{T}$) for the parameters ($\boldsymbol{\theta}$) is decribed in Table 4. Starting from this feasible set, the branch-and-bound algorithm is executed until a relative resolution of $1/8 = 0.125$ is reached for every live set. This requires 512 iterations in the worst case. A minimal bounding box is constructed around these live sets. The algorithm is then repeated starting with this bounding box until the same relative resolution is reached. The parameter intervals describing the resulting bounding boxes are shown in Fig. 8. One can see that the upper bound interval for every parameter can

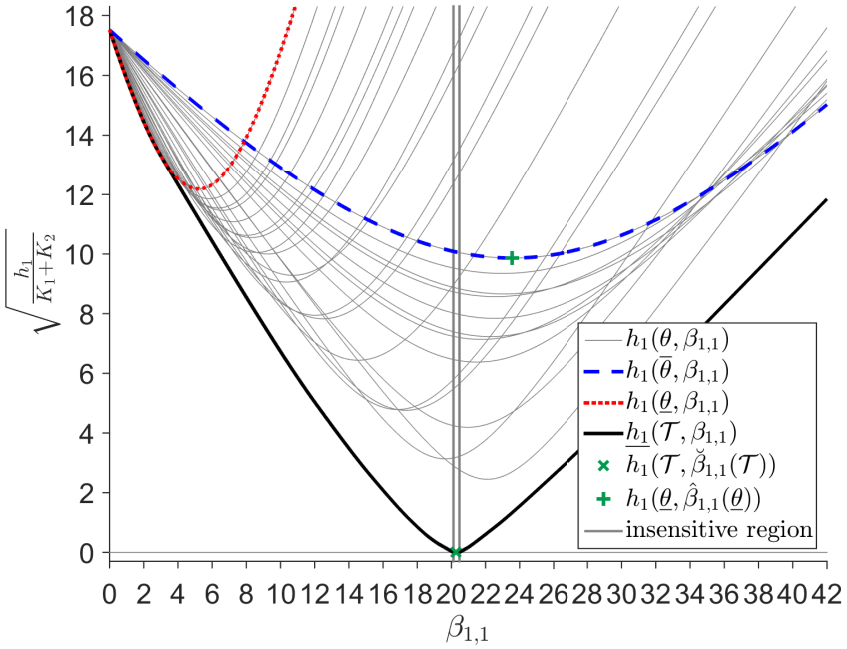

Figure 7: Bounding the objective function term h_1. Objective function profiles for the *(i)* bounding values of $\boldsymbol{\theta}$ (dashed lines), *(ii)* selected values within these bounds (full thin lines), and *(iii)* the relaxed regression problem (full thick line). The obtained value for $\underline{h_1}$ at $\hat{\beta}_{1,1}(\mathcal{T})$ is clearly lower than or equal to any other objective function value. Major observations are as follows: *(i)* the gap between the lower bound objective function and the (global) minimum is rather large; *(ii)* the selected upper bound to the objective function term is clearly larger than the global minimum value as some of the selected profiles can deliver a better overall fit; and *(iii)* an insensitive region exists for the lower bound objective function within which the lower bound is exactly equal to zero.

be reduced whereas the lower bound remains the same. The obtained upper bounds increase with the parameter index ($\overline{\theta_1} \leq \overline{\theta_2} \leq \overline{\theta_3}$), meaning that the absolute resolution decreases more slowly with increasing parameter index. At the fifth repetition of the algorithm, the bounding box cannot be reduced further. In Table 5 the number of iterations for each run of the algorithm as well as the relative volume of the bounding box is given. As can be seen, the volume is reduced to 2.36% of the original feasible parameter space with a total of 702 algorithm iterations. With each iteration, four simulations are executed (two for each leaf set), leading to a total of roughly 2800 simula-

tions. In what follows, the lastly obtained bounding box is referred to as the contracted set. The set described in Table 4 is called the original set.

Table 4: Feasible root set for parameter optimization ($\mathcal{T}$).

Parameter θ	Lower bound $\underline{\theta}$	Upper bound $\overline{\theta}$
θ_1	1×10^{-5}	10
θ_2	1×10^{-5}	10
θ_3	1×10^{-5}	10

Table 5: Initial executions of the branch-and-bound algorithm

Repetition	Number of iterations	Fraction [%]
1	35	9.38
2	99	3.08
3	181	2.69
4	196	2.36
5	191	2.36

The branch-and-bound algorithm is now executed starting with the contracted set as the root set and until all live sets are as small as $1/2^{11}$ times the range of the contracted bounding box in every time dimension (absolute resolution: $\theta_1 : 0.47 \times 10^{-3}$, $\theta_2 : 1.83 \times 10^{-3}$, $\theta_3 : 3.20 \times 10^{-3}$). In Fig. 3 the simulation according to the best-fitting values for $\boldsymbol{\theta}$ is given. One can see that the fit is reasonable, although systematic deviations can be observed. At the start of the experiment the TNN is underestimated and the OUR is overestimated and at the end of the experiment the OUR is overestimated. About halfway during the experiment TNN is overestimated and OUR is underestimated.

If Fig. 9, progress indicators for the branch-and-bound algorithm are given as a function of the iteration number. The algorithm terminated after 575845 iterations. This represents 0.067‰ of the maximal number of iterations ($2^{11^3} \sim 8.6 \times 10^9$), leading to a total of roughly 2.3×10^6 simulations. One can see that the number of live nodes exhibits a concave profile with an increasing trend until iteration 430667 and a decreasing trend afterwards. The maximal number of live sets reached during optimization is

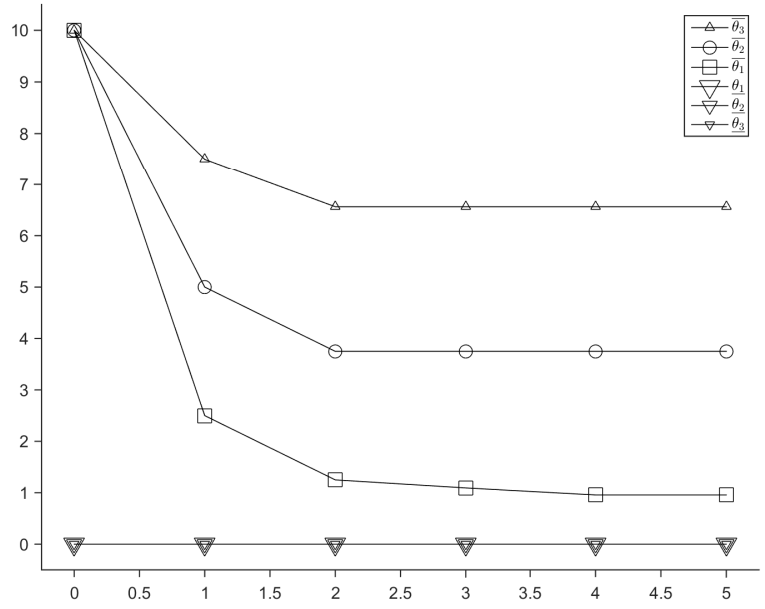

Figure 8: Bounds to the parameter values defining the bounding boxes around live sets following initial executions of the branch-and-bound algorithm.

3518700. At the end of the branch-and-bound optimization, 226910 live sets remain available. The volume represented by the live nodes (Fig. 9, middle panel) reduces monotonically during the algorithm execution. In log-scale, the profile has an inverse-sigmoid shape. At the end of the algorithm execution, the live sets represent 0.0264‰ of the contracted set (0.662×10^{-6} of the original set). In the bottom panel of Fig. 9, one can see the evolution of the lower and upper bounds to the weighted root mean squared residual (WRMSR= $\sqrt{h/(K_1 + K_2)}$). The lower bound increases monotonically while the upper bound decreases monotonically. The upper bound converges relatively fast and reaches its final value of 4.01 at iteration 25020, i.e. after less than 5% of the total number of iterations. The lower bound converges more slowly and reaches its final value of 3.91 at iteration 205770 (36% of total iterations). Most iterations are thus spent on reducing the volume of the sets containing the global optimum. At algorithm termination, a gap

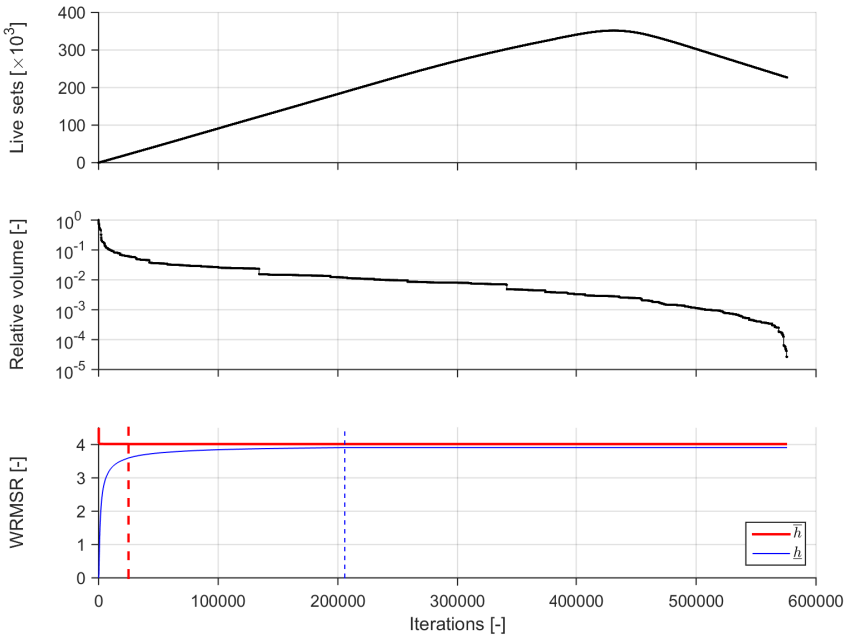

Figure 9: Progress of the branch-and-bound algorithm. Top: Number of live sets. Middle: Relative volume of live sets. Bottom: Lowest values for the bounds among all live sets. Dashed lines indicate the iteration at which the final value for the lowest upper and lower bound is obtained.

of 0.106 between the lowest lower bound and lowest upper bound remains (relative gap: 5.5%). The algorithmically proven lower bound of 3.91 for the WRMSR signifies that the model exhibits a significant and irreducible lack-of-fit. This is true since the WRMSR corresponds to a χ^2-statistic with mean equal to one if the proposed model and assumptions are correct. Importantly, convergence to a local optimum or lack of convergence can be excluded as an explanation. As a consequence, one necessary concludes that at least one of the model assumptions, including model structure, are causing the observed lack-of-fit.

Fig. 10 shows bounds to the location of the optimal parameter vector. Topologically speaking, the complete set of retained live leaf nodes is a ball (body with genus 0, i.e. the body has no holes or internal empty spaces). The live sets represent $0.0622 \times 10^{-3}\%$ of the original set or $2.64 \times 10^{-3}\%$ of the

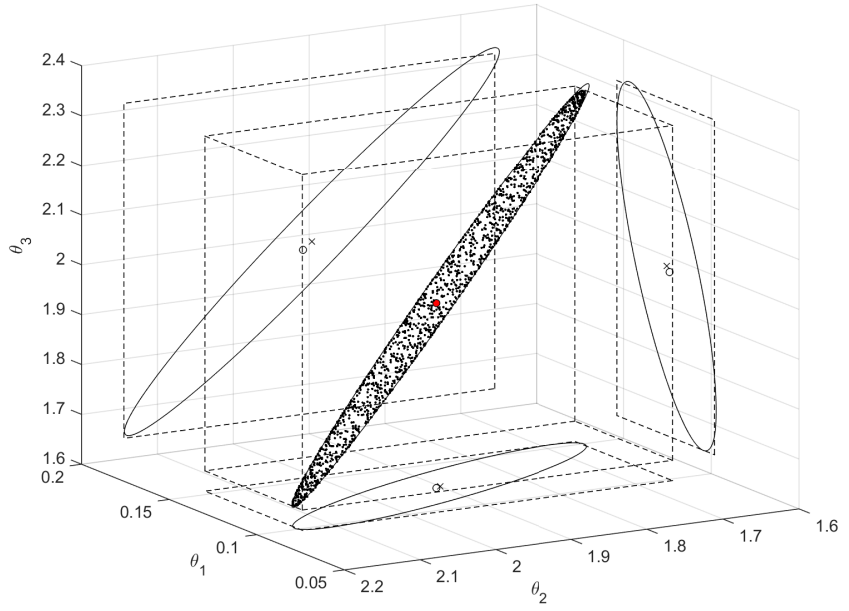

Figure 10: Location of the optimal parameter vector. Circles: location (filled) and projections (not filled) of the best-known parameter vector. Black dots: Points describing the convex hull to the live sets. Black dashed lines: bounding box. Black full lines: minimum volume enclosing ellipsoid. Black crosses: Projections of the center of minimum volume enclosing ellipsoid.

contracted set. The convex hull around all live sets at algorithm termination has a volume of $0.0714 \times 10^{-3}\%$ of the original set ($3.0328 \times 10^{-3}\%$ of the contracted set. The bounding box and minimal volume enclosing ellipsoid that contain all live leaf nodes are shown in Fig. 10. One can easily see that the volume of the bounding box is much larger than the volume of the enclosing ellipsoid. This is a consequence of the correlation between the parameters, causing the ellipsoid to circumscribe the obliquely oriented set of live sets better. The bounding box represents $1.8294 \times 10^{-3}\%$ of the original set (Table 5). The enclosing ellipsoid represents $0.0734 \times 10^{-3}\%$ of the original set ($3.12 \times 10^{-3}\%$ of contracted set). Thus, the enclosing minimum-volume ellipsoid is a reasonable and simple body to enclose the solution sets. It is oblong and flattened in appearance (semi-axes lengths: 0.86, 0.10, and

0.017). This is visible in Fig. 10 thanks to alignment of the line of eye-sight (orthogonal to the 2D-rendered image) with the mid-length axis.

To further inspect the obtained solutions, the parameter vectors representing the convex hull of the live sets retained at algorithm termination are used to simulate the extremal relative concentration profiles. These are shown in Fig. 11 together with the empirical bounds derived from them as well as the bounds corresponding to the (larger) set of parameters described by the bounding box set. One can see that the bounds for the relative concentration profiles are fairly narrow, despite the bounding gap discussed above. In addition, these simulations clearly indicate that the bounding box is ill-fitted to describe the region within which the globally optimal parameter vector lies.

3.4. Parameter values and derived results

The best-fit parameter values for the non-dimensional parameters (θ_1, θ_2, θ_3, $\beta_{1,1}$, $\beta_{2,0}$, $\beta_{2,2}$) and the original model parameters (a_{max}, K_S, K_I, $S_{N,0}$, r_{endo}, i_{growth}, according to (74)) are listed in Table 6. Values for the biomass yield, biomass growth rate, and biomass decay rate were taken from pre-existing works. These allow to compute a ballpark estimate of the biomass concentration, and changes of the biomass concentration through growth and decay. These aditional parameters are also given in Table 6. Growth amounts to roughly 4.3% increase in biomass while decay amounts to about 4.1% decrease of the biomass. The assumption of a negligible net biomass growth is therefore considered acceptable. The limited options to completely avoid such approximation errors while enabling global optimization methods are discussed below.

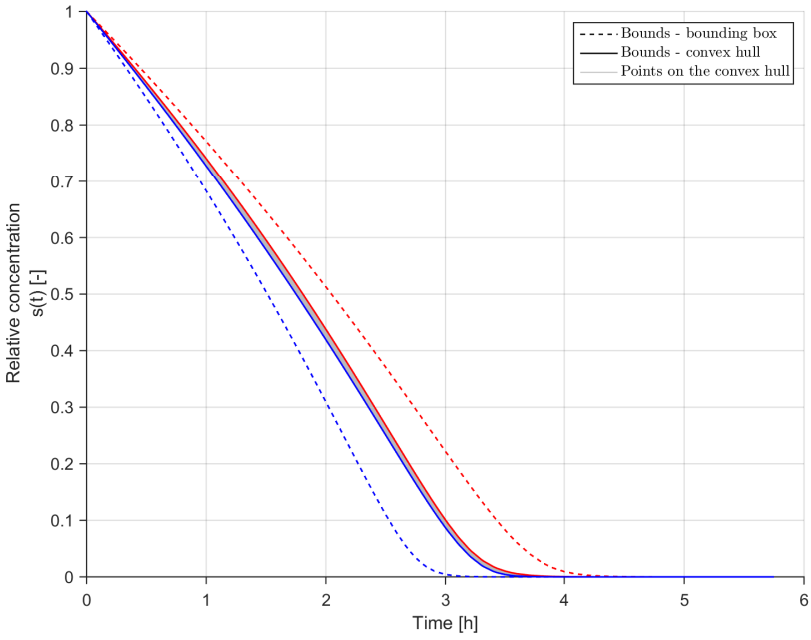

Figure 11: Simulations of the relative substrate concentration corresponding to *(i)* every parameter set defining the convex hull of the live sets (full thin lines), *(ii)* maximum and minimum over all parameter sets (full thick lines), and *(iii)* bounds corresponding to the minimum volume box enclosing all live sets (dashed lines). The simulations with the convex hull of the live sets form a patch rather than distinct lines. The obtained bounds corresponding to the bounding box around the parameter sets are not tight.

Table 6: Best-fit parameter values.

Parameter	Source/derivation	Value	Unit
θ_1	Optimization	0.124	$1/h$
θ_2	Optimization	1.91	$1/h$
θ_3	Optimization	1.97	$1/h$
$\beta_{1,1}$	Optimization	18.9	$mg\mathrm{N} \cdot L^{-1}$
$\beta_{2,0}$	Optimization	61.5	$mg\mathrm{O_2} \cdot L^{-1} \cdot h^{-1}$
$\beta_{2,2}$	Optimization	392	$mg\mathrm{O_2} \cdot L^{-1}$
a_{max}	$= \beta_{1,1}/\theta_2$	9.92	$1/h$
K_S	$= \beta_{1,1}\ \theta_1/\theta_2$	1.23	$1/h$
K_I	$= \beta_{1,1}\ \theta_2/\theta_3$	18.3	$1/h$
$S_{\mathrm{TNN,0}}$	$= \beta_{1,1}$	18.9	$mg\mathrm{N}$
r_{endo}	$= \beta_{2,0}$	61.5	$mg\mathrm{O_2} \cdot L^{-1} \cdot h^{-1}$
i_{growth}	$= \beta_{2,2}/\beta_{1,1}$	20.7	$mg\mathrm{O_2} \cdot (mgN)^{-1} \cdot h^{-1}$
b_{NOB}	[a]	0.17	d^{-1}
Y_{NOB}	[b]	1.12	$g\mathrm{COD} \cdot molN^{-1}$
μ_{max}	[b]	0.55	d^{-1}
X_{NOB}	$= a_{max}\ Y_{NOB}/\mu_{\mathrm{max}}/24h \cdot d^{-1}$	0.0346	$gCOD \cdot L^{-1}$
$\Delta X_{\mathrm{NOB,growth}}$	$= S_{N,0}\ Y_{NOB}/14\ 10^3\ mg\mathrm{N} \cdot molN^{-1}$	1.51×10^{-3}	$gCOD$
$\Delta X_{\mathrm{NOB,decay}}$	$= b_{NOB}\ X_{NOB}\ (t_{K_2} - t_1)$	1.41×10^{-3}	$gCOD/L$

[a] (Jubany, 2007)

[b] based on the following excerpt taken from (Fumasoli, 2016): *estimated from Hunik et al. (1994) for a temperature of 25°C and corrected for the salt concentration according to Moussa et al. (2006)*

3.5. A comparison with a conventional optimization strategy

The performance of the deterministic optimization method is now compared with the optimization scheme based on gradient-based search discussed above. To this end, the gradient-based quasi-Newton optimization method was run with initial values for the parameter vector $\boldsymbol{\theta}$ arranged in a uniformly spaced 20x20x20 grid spanning $\mathcal{T}$. The total number of simulations needed for this was just under 1.14×10^6. In Fig. 12, one can see the final objective function value obtained after convergence of each optimization as well as the best and final bounds by the deterministic optimization method. First of all, one can see that many of the gradient-based searchers fail to find an objective value that is close to the best-known objective value. Only 2% of the searches lead to a value within 1% above the obtained upper bound for the WRMSR and only 3.75% deliver a value within 10% above the same upper bound. If one assumes *(i)* that a 1% margin is acceptable and *(ii)* that one aims to run the gradient-based search by sampling initial values randomly until this margin is reached with a 99.9% success rate, then the minimum number of gradient-based searches delivering the expected success rate is 342 ($\sum_j^{341} 0.02 \times (1-0.02)^{j-1} \leq 0.999 \leq \sum_j^{342} 0.02 \times (1-0.02)^{j-1}$). The average number of simulations for a single gradient-based search is 142. Therefore, a sheme using 342 random starting values, requires under 50000 simulations on average.

4. Discussion

4.1. Deterministic optimization for biokinetic modeling

In this work, a deterministic optimization scheme for global optimization is proposed which enables the identification of the best parameter values in the WLS sense for a given biokinetic model. Thanks to a well-chosen parameterization, the original six-dimensional parameter optimization problem is reduced to an optimization problem that is nonlinear in three parameters only. Furthermore, the computation of a lower bound to the objective function was made feasible through a combination of interval arithmetic and a relaxation of WLS regression problems for bounded inputs. The combined method makes deterministic nonlinear parameter estimation feasible. A further benefit of this approach is that derivatives of the objective function are not required. This makes the implementation of the applied bounds fairly straightforward.

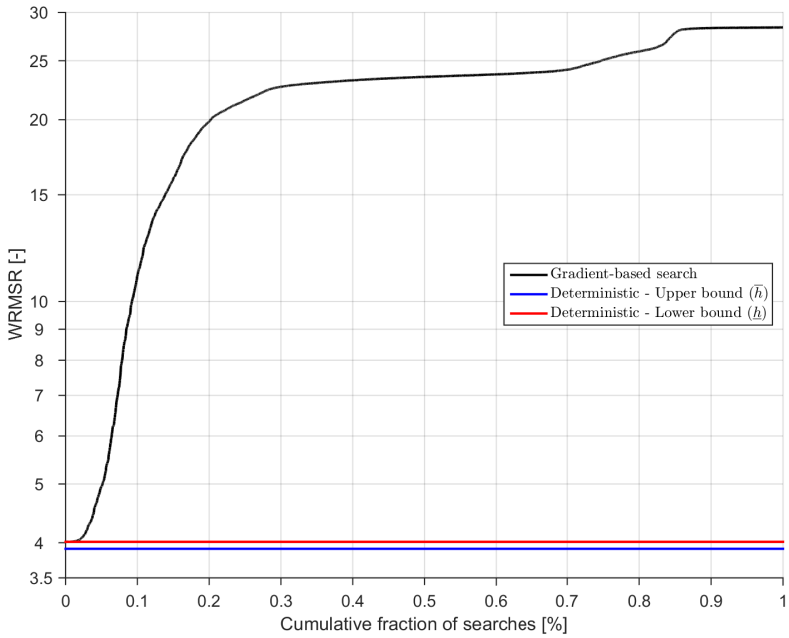

Figure 12: Sorted values for the objective function (WRMSR) after convergence of gradient-based search as a function of the fraction of executed searches. Clearly, the global minimum is only reached from a small fraction of initial parameter values.

Importantly, the use of deterministic optimization scheme means that the model fit cannot be improved beyond the lower bound by manipulation of the parameter values upon termination of the optimization algorithm. This facilitates the diagnosis task during model building as one can exclude lack of convergence or convergence to a local minimum as the cause for an observed lack-of-fit. Instead, any lack-of-fit is explained by means of a model structure deficit, ill-chosen assumptions, and/or unaccounted errors (e.g., input disturbances, measurement errors). This is expected to facilitate a more straightforward model diagnosis.

A comparison with a gradient-based search algorithm with uniform grid-based sampling of initial values shows that the number of simulations required to find an objective function value within 1% of the obtained upper bound is much lower (50000 vs. 2.3×10^6). However, to obtain this number of required simulations, about 1.14×10^6 model simulations were required. This number

is in the same scale as the number of simulations required for the deterministic optimization. In addition, such grid-based sampling or random sampling cannot guarantee that the global optimum is actually found. If not, empirical success rates based on the best-found values will overestimate the actual success rate. In this light, the deterministic optimization approach offers significant benefits, avoiding any uncertainty associated with the possible lack of convergence of gradient-based or otherwise local search algorithms.

In (2), a seemingly conservative set of admissibility requirements are specified for the rate expression and the measurement equations in (1)-(5). However, these requirements are not as conservative as one may think. To support this argument, consider that most reaction rate expressions used for environmental process modeling are nonnegative for any set of substrate, product, and inhibitor concentrations (i.e., irreversible processes). In contrast, the non-increasing property is usually not met for rate expressions given in their conventional form. However, this requirement can often be met by replacing the original parameters with their opposite or their inverse. Importantly, this is possible for a wide array of rate expressions, including but not limited to all expressions *(i)* that are exponential (with fixed base) or posynomial (with fixed exponents) in the substrate concentration and *(ii)* that can be written so that each parameter appears only once. The Haldane reaction rate expression used in this work is an example of this. Furthermore, admissible expressions included any expression that can be formulated as a sum of products of and divisions by admissible expressions which do not share any parameters. This includes a large fraction of the affinity and inhibition switching functions found in Bastin and Dochain (1990) (e.g., Blackman, 1905; Tessier, 1942; Monod, 1949; Haldane, 1965; Andrews, 1968; Sokol and Howell, 1981; Ming et al., 1989), and sum-of-product combinations thereof (e.g., Shehata and Marr, 1971; Jost et al., 1973; Chen and Hashimoto, 1978; Hoppe and Hansford, 1982; Hellinga et al., 1999). Rate expressions with multiple appearances of the same parameter may still satisfy the requirements, however pending detailed analysis (e.g., Steele, 1965). If the non-negativity requirement for the vectors $\boldsymbol{\beta}_j$ is not met initially, then linear transformation of the obtained measurements is sufficient in most cases to satisfy the requirement.

Another apparent complication may arise from the appearence of product concentrations in the rate expression. However, in batch experiments the product concentrations can be written as a function of the substrate concentration based on stoichiometric balances. As a result, rate expres-

sions including product inhibition may easily be reformulated again so to satisfy the form of (1)-(2) (not shown, e.g., Aiba et al., 1968; Aborhey and Williamson, 1977).

4.2. Limitations of the study

Our study is limited in a few ways. First of all, the reformulation of the problem as a three-dimensional nonlinear problem is only shown possible for a single pulse experiment. When multiple pulses are used in series, the possible dimension reduction is not as dramatic unless the initial concentration and pulse additions are known exactly (not demonstrated).

A second limitation is that the results are demonstrated for experiments involving a univariate process and by assuming negligible net biomass growth. More specifically, the proposed bounding procedures are currently limited to irreversible univariate processes. Similar bounding procedures are likely applicable to multivariate systems, in particular when they are cooperative. This includes monotone non-reversible reaction systems (De Leenheer et al., 2007). This is not demonstrated yet. However, even if it is applicable, convergence may become prohibitively slow as more parameters are included in the optimization problem. Therefore, our current efforts are instead focused on fusing the global optimization strategy with methods that transform experimental data into extents, i.e. data series reflecting the dynamics of each dynamic phenomenon separately (Bhatt et al., 2011). On the plus side, most batch/pulse experiments in wastewater engineering involve processes modeled in a univariate fashion thanks to addition of enzyme inhibiting products. This prevents confounding phenomena from occurring and is typically motivated as a way to improve practical identifiability.

Thirdly, there are notable rate expressions found in the literature which could not (yet) be reformulated by the authors to satisfy the requirements in (1)-(2). These include switching functions for substrate affinity, substrate inhibition, and product inhibition (e.g., Moser, 1958; Konak, 1974; Levenspiel, 1980; Luong, 1987). To make the optimization method applicable in this case, the procedure to bound the state estimates (18)-(20) can be replaced with more general methods (e.g., Berleant and Kuipers, 1997; Sahlodin and Chachuat, 2011).

Lastly, it was shown that the provided bounds result in fairly large gaps between the lower and upper bounds, further leading to relatively slow convergence and fairly large region around the global optimum. The gap is believed to be caused mainly by the necessary coupling of lower and upper

bound values for s with lower and upper bound values for $q(s)$. The applied rules for interval arithmetic lead to expressions within which $\boldsymbol{\theta}$ appears multiple times. This, in turn, leads to a *dependency* problem as the complete polyhedron for $\boldsymbol{\theta}$ is considered independently for every instance of $\boldsymbol{\theta}$ in the expression (Moore et al., 2009). As a consequence, the bounds for $q(s)$ are not tight. An analytical expression for $q(s, \boldsymbol{\theta})$ within which $\boldsymbol{\theta}$ appears only once is not known to the authors and for this reason, it is unclear how the bounding gap can be reduced for the given setup. The problem may be avoided by fitting a model to the oxygen measurements, rather than to the derived OUR signal. Such an approach requires however that the oxygen gas-transfer and sensor dynamics are explicitly accounted for, either by assuming these dynamic processes are known or by modeling them as well. To this end, the extent-based approach discussed above is again considered promising.

It is noted that a slightly better lower bound was obtained by incorporating shape constraints for the relative concentration profiles into account (see 2.3.6). However, this improved bound does not solve the dependency problem discussed above. Furthermore, no significant improvements were observed in terms of rate of convergence or the volume of the live sets at termination, despite a considerable increase in computational complexity. Initial results (not shown) suggest that the obtained improvement is marginal compared to the gap caused by dependency problem. This suggests that the usefulness of this lower bound depends on whether the dependency problem can be solved or otherwise avoided.

The efficiency of deterministic optimization tools for the posed model-fitting problem could increase dramatically if the fitting to rate measurements is avoided to eliminate the dependency problem. One may also attempt to solve the problem in its original nonlinear form to avoid using the relaxed regression for the lower bound, then however requiring branch-and-bound optimization in six dimensions. Despite the larger dimensionality, tighter bounds may still lead to faster convergence. In addition, the algorithm efficiency is also affected by the information content of the data and the model parameterization. The interactive effects of *(i)* experimental design and data information content, *(ii)* available model outputs (state, rate, or mixed type measurements), *(iii)* appearance of the dependency problem, *(iv)* dimensionality reducing schemes, and *(v)* applied bounding procedures on algorithm performance remains open for further study.

4.3. Opportunities in kinetic modeling

An interesting observation is that the set of live parameter vector sets retained at termination of the branch-and-bound algorithm can be approximated well by an ellipsoidal body. This is believed to be in part due to the particular parameterization of the reaction rate (75). Indeed, such a reformulation can facilitate optimization as least-squares estimators of the parameters tend to be unbiased and normally distributed (Ratkowsky, 1983, 1986). As a consequence, equiprobable surfaces are approximated well with ellipsoids. However, the strength of the link between the close-to-Gaussian behavior of parameter estimates and the close-to-ellipsoidal nature of the region including the global optimum remains open for study. In addition, it was observed that the upper and lower bounds to the objective function converge at an early stage of the algorithm. This means early-stopping criteria could be useful to find acceptable parameter estimates before actual algorithm termination. This means that the global optimum is then enclosed in a significantly larger region of the parameter space.

Having found a globally optimal parameter vector and an enclosing ellipsoid also presents an opportunity for uncertainty analysis. The obtained mode and enclosing ellipsoid can be used to describe the uncertainty in a qualitative manner, can be used to populate stochastic sampling methods during initialization, or could assist in finding a good proposal distribution for Markovian sampling techniques. Importantly, the obtained results suggest that the instance of the optimization problem solved in this study exhibits only one optimum which, if so, is the global optimum. This is not generally true for the studied model and depends on the experimental design and the obtained measurements. Enumerating all (local) minima is non-trivial and would also require a deterministic numerical optimization scheme for guaranteed results. As such, our optimization method is the only method known to the authors which guarantees global optimality of the parameter estimates for the studied model structure.

5. Conclusions

A deterministic global optimization method has been proposed and applied for parameter estimation. The method has been tested to describe a batch pulse experiment executed for modeling of nitrite oxidation by autotrophic bacteria in a urine nitrification reactor. The obtained results show

that this is indeed possible. The reported success is attributed to a combination of model reformulation, interval arithmetic, and problem relaxation. Importantly, these tools are generally applicable so that the optimization technique is not limited to a single experiment, reaction, or process. The current form of the optimization method is however limited to the modeling of batch experiments involving a single reaction only. Suggestions to improve the scope of applicability as well as increasing the speed of convergence have been discussed in the text. Most interestingly, the results suggest that the estimation problem is actually pseudo-convex in nature and could therefore be solved to its unique local optimum by means of fast algorithms developed for this kind of problems.

6. Acknowledgments

The authors thank Dr. Dominique Bonvin (EPFL) for critical reading of our manuscript, Gabriel Kämpf for collecting the experimental data, the authors of the Spike_O toolbox (Villez et al., 2013; Villez and Habermacher, 2016; Villez et al., 2016), Johannes Korsawe for the bounding box computation and visualization code and Nima Moshtagh for the minimal volume ellipsoid computation and visualization code. This research is made possible by Eawag Discretionary Funds (grant no.: 5221.00492.009.03, project: DF2015/EMISSUN).

7. Supplementary Information

The Supplementary Information includes the data and software necessary to reproduce our results.

8. References

Aborhey, S., Williamson, D., 1977. Modelling of lactic acid production by *Streptococcus cremoris* hp. The Journal of General and Applied Microbiology 23 (1), 7–21.

Aiba, S., Shoda, M., Nagatani, M., 1968. Kinetics of product inhibition in alcohol fermentation. Biotechnology and Bioengineering 10 (6), 845–864.

Andrews, J. F., 1968. A mathematical model for the continuous culture of microorganisms utilizing inhibitory substrates. Biotechnology and Bioengineering 10 (6), 707–723.

Bastin, G., Dochain, D., 1990. On-line estimation and adaptive control of bioreactors. Elsevier New York, Amsterdam.

Benedetti, L., Claeys, F., Nopens, I., Vanrolleghem, P. A., 2011. Assessing the convergence of LHS Monte Carlo simulations of wastewater treatment models. Water Science and Technology 63 (10), 2219–2224.

Berleant, D., Kuipers, B., 1997. Qualitative and quantitative simulation: bridging the gap. Artif. Intell. 95, 215–255.

Bhatt, N., Amrhein, M., Bonvin, D., 2011. Incremental identification of reaction and mass-transfer kinetics using the concept of extents. Industrial & Engineering Chemistry Research 50 (23), 12960–12974.

Blackman, F. F., 1905. Optima and limiting factors. Annals of Botany 19 (74), 281–295.

Checchi, N., Giusti, E., Marsili-Libelli, S., 2007. PEAS: A toolbox to assess the accuracy of estimated parameters in environmental models. Environmental Modelling and Software 22 (6), 899–913.

Chen, Y. R., Hashimoto, A. G., 1978. Kinetics of methane fermentation (no. conf-780549-8). Science and Education Administration, Clay Center, NE (USA). Meat Animal Research Center.

Dauphin, Y. N., Pascanu, R., Gulcehre, C., Cho, K., Ganguli, S., Bengio, Y., 2014. Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. Advances in neural information processing systems, 2933–2941.

De Leenheer, P., Angeli, D., Sontag, E. D., 2007. Monotone chemical reaction networks. Journal of Mathematical Chemistry 41 (3), 295–314.

Floudas, C., 1999. Deterministic Global Optimization: Theory, Algorithms and Applications. Kluwer Academic, Dordrecht.

Forst, W., Hoffmann, D., 2010. Optimization - Theory and Practice. Springer.

Fumasoli, A., 2016. Nitrification of urine as pretreatment for nutrient recovery. Ph.D. thesis, ETH Zürich.

Fumasoli, A., Morgenroth, E., Udert, K., 2015. Modeling the low pH limit of nitrosomonas eutropha in high-strength nitrogen wastewaters. Water Research 83, 161–170.

Haldane, J., 1965. Enzymes, 1930. Reprint by MIT, Cambridge.

Hansen, E., Walster, G. W. E. (Eds.), 2003. Global optimization using interval analysis: Revised and expanded (Vol. 264). CRC Press.

Hellinga, C., Van Loosdrecht, M. C. M., Heijnen, J. J., 1999. Model based design of a novel process for nitrogen removal from concentrated flows. Mathematical and Computer Modelling of Dynamical Systems 5 (4), 351–371.

Henze, M., Gujer, W., Mino, T., van Loosdrecht, M., 2000. Activated sludge models ASM1, ASM2, ASM2d and ASM3. IWA Scientific and Technical Report. IWA Publishing, London, UK.

Hoppe, G. K., Hansford, G. S., 1982. Ethanol inhibition of continuous anaerobic yeast growth. Biotechnology Letters 4 (1), 39–44.

Hunik, J. H., Tramper, J., Wijffels, R. H., 1994. A strategy to scale up nitrification processes with immobilized cells of *Nitrosomonas europaea* and *Nitrobacter agilis*. Bioprocess Engineering 11 (2), 73–82.

Inuiguchi, M., Tanino, T., 2006. Interval linear regression analysis based on Minkowski difference – A bridge between traditional and interval linear regression models. Kybernetika 42 (4), 423–440.

Jakeman, A. J., Letcher, R. A., Norton, J. P., 2006. Ten iterative steps in development and evaluation of environmental models. Environmental Modelling and Software 21 (5), 602–614.

Jiang, F., Chen, Y., Mackey, H. R., Chen, G. H., van Loosdrecht, M. C. M., 2011. Urine nitrification and sewer discharge to realize in-sewer denitrification to simplify sewage treatment in Hong Kong. Water Science and Technology 64 (3), 618–626.

Jost, J. L., Drake, J. F., Fredrickson, A. G., Tsuchiya, H. M., 1973. Interactions of *Tetrahymena pyriformis, Escherichia coli, Azotobacter vinelandii,* and glucose in a minimal medium. Journal of Bacteriology 113 (2), 834–840.

Jubany, I., 2007. Operation, modelling and automatic control of complete and partial nitrification of highly concentrated ammonium wastewater. Ph.D. thesis, Universitat Autonoma de Barcelona, Barcelona, Spain.

Jubany, I., Baeza, J. A., Carrera, J., Lafuente, J., 2005. Respirometric calibration and validation of a biological nitrite oxidation model including biomass growth and substrate inhibition. Water research 39 (18), 4574–4584.

Konak, A. R., 1974. Derivation of a generalised Monod equation and its applications. Journal of Applied Chemistry and Biotechnology.

Levenspiel, O., 1980. The Monod equation: a revisit and a generalization to product inhibition situations. Biotechnology and Bioengineering 22 (8), 1671–1687.

Levine, W. S. (Ed.), 1996. The control handbook. Boca Raton, New York: CRC Press, IEEE Press.

Luong, J. H. T., 1987. Generalization of Monod kinetics for analysis of growth data with substrate inhibition. Biotechnology and Bioengineering 29 (2), 242–248.

Marsili-Libelli, S., 2010. Modelling and automation of water and wastewater treatment processes. Environmental Modelling and Software 25 (5), 613–615.

Ming, F., Howell, J. A., Canovas-Diaz, M., 1989. Computer Applications in Fermentation Technology: Modelling and Control of Biotechnological Processes. Springer Netherlands, Ch. Mathematical simulation of anaerobic stratified biofilm processes, pp. 69–77.

Monod, J., 1949. The growth of bacterial cultures. Annual Reviews in Microbiology 470 (3), 371–394.

Moore, R. E., Kearfott, R. B., Cloud, M. J., 2009. Introduction to Interval Analysis. SIAM.

Moser, H., 1958. The dynamics of bacterial populations in the chemostat. Carnegie Inst Publication, n614, Washington.

Moussa, M. S., Sumanasekera, D. U., Ibrahim, S. H., Lubberding, H. J., Hooijmans, C. M., Gijzen, H. J., van Loosdrecht, M. C. M., 2006. Long term effects of salt on activity, population structure and floc characteristics in enriched bacterial cultures of nitrifiers. Water Research 40 (7), 1377–1388.

Nocedal, J., Wright, S., 2006. Numerical optimization. Springer Science & Business Media.

Oosterhuis, M., van Loosdrecht, M. C. M., 2009. Nitrification of urine for H_2S control in pressure sewers. Water Practice and Technology 4(3).

Ratkowsky, D. A., 1983. Nonlinear Regression Modeling – A Unified Practical Approach. Vol. 48 of Statistics: Textbooks and Monographs. Dekker, New York.

Ratkowsky, D. A., 1986. A suitable parameterization of the Michaelis-Menten enzyme reaction. Biochemical Journal 240, 357–360.

Rieger, L., Gillot, S., Langergraber, G., Ohtsuki, T., Shaw, A., Takács, I., Winkler, S., 2012. Guidelines for using activated sludge models. IWA Task Group on Good Modelling Practice. IWA Scientific and Technical Report. IWA Publishing.

Rump, S. M., 1999. Developments in Reliable Computing. Kluwer Academic Publishers, Dordrecht., Ch. INTLAB - INTerval LABoratory., pp. 77–104.

Sahlodin, A. M., Chachuat, B., 2011. Convex/concave relaxations of parametric ODEs using Taylor models. Computers and Chemical Engineering 35, no. (5 (): 844-857.).

Shehata, T. E., Marr, A. G., 1971. Effect of nutrient concentration on the growth of *Escherichia coli*. Journal of Bacteriology 107 (1), 210–216.

Sin, G., De Pauw, D. J., Weijers, S., Vanrolleghem, P. A., 2008. An efficient approach to automate the manual trial and error calibration of activated sludge models. Biotechnology and bioengineering 100 (3), 516–528.

Sokol, W., Howell, J., 1981. Kinetics of phenol oxidation by washed cells. Biotechnology and Bioengineering 23, 2039–2049.

Steele, J. H., 1965. Primary Productivity in Aquatic Environments (Mem. Ist. Ital. Idrobiol., 18 Suppl.). University of California Press, Berkeley, Ch. Notes on some theoretical problems in production ecology, p. 383398.

Tessier, G., 1942. Croissance des populations bactériennes et quantité daliment disponible. Revue Scientifique, Paris 3208, 209–214.

Udert, K. M., Wächter, M., 2012. Complete nutrient recovery from source-separated urine by nitrification and distillation. Water research 46 (2), 453–464.

Villez, K., Habermacher, J., 2016. Shape anomaly detection for process monitoring of a sequencing batch reactor. Computers and Chemical Engineering 91, 365–379.

Villez, K., Vanrolleghem, P. A., Corominas, L., 2016. Optimal flow sensor placement on wastewater treatment plants. Water Research 101, 75–83.

Villez, K., Venkatasubramanian, V., Rengaswamy, R., 2013. Generalized shape constrained spline fitting for qualitative analysis of trends. Computers and Chemical Engineering 58, 116–134.