Identification of Biokinetic Models using the Concept of Extents

Alma Mašić,[†] Sriniketh Srinivasan,[‡] Julien Billeter,[‡] Dominique Bonvin,[‡] and Kris Villez[*,†]

†*Eawag: Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, CH-8600 Dübendorf, Switzerland*

‡*Laboratoire d'Automatique, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland*

E-mail: Kris.Villez@eawag.ch

Manuscript and Supporting Information accepted for publication in Environmental Science & Technology

Abstract

The development of a wide array of process technologies to enable the shift from conventional biological wastewater treatment processes to resource recovery systems is matched by an increasing demand for predictive capabilities. Mathematical models are excellent tools to meet this demand. However, obtaining reliable and fit-for-purpose models remains a cumbersome task due to the inherent complexity of biological wastewater treatment processes. In this work, we present a first study in the context of environmental biotechnology that adopts and explores the use of extents as a way to simplify and streamline the dynamic process modeling task. In addition, the

extent-based modeling strategy is enhanced by optimal accounting for nonlinear algebraic equilibria and nonlinear measurement equations. Finally, a thorough discussion of our results explains the benefits of extent-based modeling and its potential to turn environmental process modeling into a highly automated task.

Introduction

Dynamic models are increasingly used to better understand, design, and operate environmental processes.[1,2] For biological wastewater treatment processes, the available activated sludge model family[3,4] has been used widely despite reported challenges in model identification. These challenges relate to *(i)* the information content and the quality of calibration data that limit practical identifiability,[5–10] *(ii)* the lack of mechanistic understanding,[11,12] and *(iii)* nonlinear and non-convex properties.[13–15] These issues are even more severe in the case of decentralized treatment processes that are proposed to address fast societal dynamics by providing straightforward upscaling of wastewater treatment operations.[16] In addition, both economical and political motives are driving a paradigm shift in objectives from environmental protection to a need to generate added-value products from wastewaters. To ensure both product quality and economically optimal operation, resource recovery from wastewater requires tight management and control of the involved processes. The urine nitrification process for fertilizer production developed at Eawag is an example of this.[17] Advanced control of such high-rate processes is not possible without detailed process understanding and predictive power. In addition, the diversity of the available technologies is rapidly increasing. For this reason, fast development of reliable models is paramount to attain sustainable urban water cycles.

In the past, model complexity has been been tackled by means of model identification protocols. Examples include *(i)* protocols that split model identification into steps corresponding to major fractions of the medium[18] and *(ii)* protocols based on iterative model building.[19] Despite these efforts, the aforementioned model identification challenges have

only been partly addressed. In this work, we focus on the development of a method that deals with the nonlinear and non-convex nature of kinetic identification in biological process modeling. In previous work,[20] a deterministic optimization method was found well suited to estimate parameters in a simple model for biological nitrite oxidation. This optimization method led to globally optimal parameter estimates. The same study demonstrated that a standard approach based on gradient-based optimization fails to find good parameter estimates. Unfortunately, deterministic global optimization is cumbersome when the number of parameters is large.

To deal with the model structure selection and parameter estimation challenge, we adopt an extent-based framework[21–24] to enable the application of deterministic optimization methods to biological process models involving multiple reactions. The concept of extents allows the transformation multivariate time series into a set of individual time series, each one reflecting the progress of a single reaction. This, in turn, enables the individual identification of the rate law and the corresponding parameters for each of the biological reactions. In addition, the use of extents facilitates model diagnosis. The proposed extent-based modeling methodology is demonstrated and benchmarked against a conventional approach by means of a simulated experiment with a urine nitrification process model.[25] All symbols used in this text are listed in Table 1.

Other factors complicating model identification include *(i)* the stochastic nature of environmental processes and *(ii)* the significant lack of identifiability of model structures and parameters, further leading to significant uncertainty and correlated parameter estimates. These issues are certainly important but not studied in this work. Instead, we focus on solving model identification problems to global optimality given experimental data. This also means that we assume that a proper experimental design has been executed.

Methods

Definitions

Species and Components

Consider a batch reactor with volume V containing S chemical species involved in R reactions. The numbers of moles are given as the S-dimensional vector $\mathbf{n}$. Among the R reactions, R_k reactions are kinetically controlled, and R_e reactions are considered to be at equilibrium, with $R = R_k + R_e$. The S species are split into S_k *kinetic species* that are only involved in kinetically controlled reactions (i.e., *not* in equilibrium reactions) and S_e *equilibrium species* that are involved in equilibrium (and possibly also in kinetically controlled) reactions ($S = S_k + S_e$). The corresponding numbers of moles are $\mathbf{n}_k$ and $\mathbf{n}_e$. *Equilibrium components* are defined as the S_c molecular constituents that are involved in equilibrium reactions and whose concentrations are conserved.[26] The $\bar{S} = S_k + S_c$ numbers of moles of the kinetic species $\mathbf{n}_k$ and the equilibrium components $\mathbf{n}_c$ can be written as:

$$\bar{\mathbf{n}} = \begin{bmatrix} \mathbf{n}_k \\ \\ \mathbf{n}_c \end{bmatrix} = \bar{\mathbf{E}}\ \mathbf{n} \tag{1}$$

where $\bar{\mathbf{E}}$ of dimension $\bar{S} \times S$ relates the numbers of moles of all species $\mathbf{n}$ to those of the kinetic species and equilibrium components $\bar{\mathbf{n}}$.

Example. Let us illustrate the notations through a simplified urine nitrification process model[25] that is used in this work to simulate experimental data. This model is selected because it is an excellent example of a biological process model based on the activated sludge model family and involving rate-controlling acid-base equilibria. There are $R = 6$ reactions involving $S = 10$ species dissolved in water. The kinetically controlled reactions are the biological nitritation and nitratation by ammonia oxidizing bacteria (AOB) and

nitrite oxidizing bacteria (NOB), respectively, that is, $R_k = 2$:

$$\mathrm{NH_3} + {}^3\!/_2\ \mathrm{O_2} \xrightarrow{AOB} \mathrm{HNO_2} + \mathrm{H_2O}$$
$$\mathrm{HNO_2} + {}^1\!/_2\ \mathrm{O_2} \xrightarrow{NOB} \mathrm{HNO_3}$$

The remaining reactions consist of $R_e = 4$ instantaneous acid-base equilibrium reactions:

$$\mathrm{NH_4^+} \rightleftharpoons \mathrm{H^+} + \mathrm{NH_3}$$
$$\mathrm{HNO_2} \rightleftharpoons \mathrm{H^+} + \mathrm{NO_2^-}$$
$$\mathrm{HNO_3} \rightleftharpoons \mathrm{H^+} + \mathrm{NO_3^-}$$
$$\mathrm{H_2O} \rightleftharpoons \mathrm{H^+} + \mathrm{OH^-}$$

The net growth of bacteria is assumed negligible. According to this reaction scheme, the $S = 10$ species are oxygen, ammonium, ammonia, nitrous acid, nitrite, nitric acid, nitrate, proton ions, hydroxyl ions, and water. Oxygen is *only* involved in the kinetically controlled reactions ($S_k = 1$). The remaining species are equilibrium species ($S_e = 9$). The numbers of moles are computed from the concentrations as follows:

$$\begin{aligned}\mathbf{n} &= V \left[\begin{array}{cccccccccc} [\mathrm{O_2}] & [\mathrm{NH_4^+}] & [\mathrm{NH_3}] & [\mathrm{HNO_2}] & [\mathrm{NO_2^-}] & [\mathrm{HNO_3}] & [\mathrm{NO_3^-}] & [\mathrm{H^+}] & [\mathrm{OH^-}] & [\mathrm{H_2O}] \end{array} \right]^{\mathrm{T}} \\ &= \left[\begin{array}{c} \mathbf{n}_k \\ \mathbf{n}_e \end{array} \right] = \left[\begin{array}{c} V\,[\mathrm{O_2}] \\ \mathbf{n}_e \end{array} \right]. \end{aligned} \tag{2}$$

The $S_c = 5$ molecular constituents that are conserved in the equilibrium reactions are total ammonia (total ammonia nitrogen, TAN), total nitrite (TNO2), total nitrate (TNO3), total proton (TH), and total hydroxyl (TOH). With $S_k = 1$ (oxygen), the 6×10 matrix $\bar{\mathbf{E}}$ reads:

$$\bar{\mathbf{E}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}. \tag{3}$$

Balance Equations

For a batch reactor, the differential mole balance equations are written as

$$\dot{\mathbf{n}}(t) = V\,\mathbf{N}^{\mathrm{T}}\,\mathbf{r}\big(\mathbf{n}(t)/V\big), \qquad \mathbf{n}(0) = \mathbf{n}_0 \tag{4}$$

with $\mathbf{N}$ the $R \times S$ stoichiometric matrix, V the volume (assumed to be constant), $\mathbf{r}$ the R-dimensional reaction rates, and $\mathbf{n}_0$ the S-dimensional initial numbers of moles. Upon pre-multiplying (4) by $\bar{\mathbf{E}}$, one obtains:

$$\dot{\bar{\mathbf{n}}}(t) = \bar{\mathbf{E}}\,\dot{\mathbf{n}}(t) = V\,\bar{\mathbf{E}}\,\mathbf{N}^{\mathrm{T}}\,\mathbf{r}\big(\mathbf{n}(t)/V\big) = V\,\bar{\mathbf{N}}^{\mathrm{T}}\,\mathbf{r}\big(\mathbf{n}(t)/V\big), \qquad \bar{\mathbf{n}}(0) = \bar{\mathbf{n}}_0 \tag{5}$$

with $\bar{\mathbf{N}}$ the corresponding stoichiometric matrix of dimension $R \times \bar{S}$ and $\bar{\mathbf{n}}_0 = \bar{\mathbf{E}}\,\mathbf{n}_0$. Given $\bar{\mathbf{n}}$, the vector $\mathbf{n}$ is obtained by solving the following system of $S = \bar{S} + R_e$ algebraic equations:[26]

$$\bar{\mathbf{E}}\,\mathbf{n}(t) = \bar{\mathbf{n}}(t) \tag{6}$$

$$\mathbf{g}\big(\mathbf{n}(t)/V\big) = \mathbf{0}_{R_e} \tag{7}$$

where $\mathbf{g}(\cdot)$ expresses the R_e instantaneous equilibria. The dynamics of the component concentrations are functions of the kinetically controlled reactions only, that is, the rows of $\bar{\mathbf{N}}$ corresponding to the equilibrium reactions contain only zeros.[26] Hence, a reduced stoichiometric matrix $\bar{\mathbf{N}}_k$ can be defined as the matrix consisting of the rows of $\bar{\mathbf{N}}$ with at least one non-zero element. Following this, (5) reduces to:

$$\dot{\bar{\mathbf{n}}}(t) = V\,\bar{\mathbf{N}}_k^{\mathrm{T}}\,\mathbf{r}_k\big(\mathbf{n}(t)/V\big), \qquad \bar{\mathbf{n}}(0) = \bar{\mathbf{n}}_0 \tag{8}$$

with $\mathbf{r}_k$ the kinetically controlled reaction rates.

Example. Following the aforementioned definitions, the stoichiometric matrix for all reactions is

$$\mathbf{N} = \begin{bmatrix} -3/2 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ -1/2 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & -1 \end{bmatrix} \tag{9}$$

with the reduced stoichiometric matrix

$$\bar{\mathbf{N}}_k = \begin{bmatrix} -3/2 & -1 & 1 & 0 & 2 & 1 \\ -1/2 & 0 & -1 & 1 & 0 & 0 \end{bmatrix} \tag{10}$$

describing the $R_k = 2$ kinetically controlled reactions in terms of the $S_k = 1$ kinetic species and the $S_c = 5$ equilibrium components.

The rate laws for the biological oxidation reactions are:

$$\mathbf{r}_k = \begin{bmatrix} r_{\mathrm{AOB}} \\ r_{\mathrm{NOB}} \end{bmatrix} = \begin{bmatrix} [\mathrm{NH_3}] \Big/ \left(\theta_{\mathrm{AOB},1} + \theta_{\mathrm{AOB},2}\,[\mathrm{NH_3}] + \theta_{\mathrm{AOB},3}\,[\mathrm{NH_3}]^2 \right) \\ [\mathrm{HNO_2}] \Big/ \left(\theta_{\mathrm{NOB},1} + \theta_{\mathrm{NOB},2}\,[\mathrm{HNO_2}] \right) \end{bmatrix} \tag{11}$$

with the kinetic parameters $\theta_{\mathrm{AOB},1}$, $\theta_{\mathrm{AOB},2}$, $\theta_{\mathrm{AOB},3}$, $\theta_{\mathrm{NOB},1}$, and $\theta_{\mathrm{NOB},2}$. The time dependence of rates and concentrations is omitted for the sake of conciseness. The two kinetic expressions correspond to Haldane and Monod kinetics, respectively. Since we assume that oxygen is sufficient for both oxidation processes, rate-limiting effects of oxygen can be safely ignored. The balance equations describing the equilibria cover four acid-base reactions so that (7) is

$$\mathbf{g}\left(\mathbf{n}/V\right) = \begin{bmatrix} \left([\mathrm{H^+}]\,[\mathrm{NH_3}] \right) \Big/ [\mathrm{NH_4{}^+}] - 10^{-pK_{a,\mathrm{NH_4{}^+}}} \\ \left([\mathrm{H^+}]\,[\mathrm{NO_2{}^-}] \right) \Big/ [\mathrm{HNO_2}] - 10^{-pK_{a,\mathrm{HNO_2}}} \\ \left([\mathrm{H^+}]\,[\mathrm{NO_3{}^-}] \right) \Big/ [\mathrm{HNO_3}] - 10^{-pK_{a,\mathrm{HNO_3}}} \\ [\mathrm{H^+}][\mathrm{OH^-}] - 10^{-pK_w} \end{bmatrix} = \mathbf{0}_4. \tag{12}$$

The initial numbers of moles are $\mathbf{n}_{k,0} = V \left[\begin{array}{cccccc} c_{\mathrm{O_2},0} & c_{\mathrm{TAN},0} & 0 & 0 & c_{\mathrm{TH},0} & c_{\mathrm{TOH},0} \end{array} \right]^{\mathrm{T}}$. As there is no liquid entering or leaving the reactor during the reaction, the simulated data correspond to a typical batch test[27–30] with a single pulse of ammonia dosed at the start

of the experiment. The initial concentrations of proton and hydroxyl component ($c_{\mathrm{TH},0}$ and $c_{\mathrm{TOH},0}$) are set to values that satisfy the equilibrium equations and deliver a zero ion balance. The initial oxygen and water concentrations can be set arbitrarily and do not affect the reaction rates nor the equilibria.

Measurement Equations

During the simulated batch experiment, M measurements are obtained at H distinct time instants t_h, with $h = 1, \ldots, H$ and $t_1 = 0$, as:

$$\tilde{\mathbf{y}}(t_h) = \mathbf{y}(t_h) + \boldsymbol{\epsilon}(t_h) = \mathbf{f}\big(\mathbf{n}(t_h)/V\big) + \boldsymbol{\epsilon}(t_h), \qquad \boldsymbol{\epsilon}(t_h) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_h) \tag{13}$$

with $\tilde{\mathbf{y}}(t_h)$ the M-dimensional vector of measurements, and $\mathbf{y}(t_h)$ the noise-free measured variables at time t_h. In words, the measurements are nonlinear functions of the species concentrations and are subject to additive Gaussian noise. The functions $\mathbf{f}(\cdot)$ are assumed continuous and differentiable.

Example. Measurements of the total ammonia, total nitrite and total nitrate concentrations and of pH are obtained. The noise-free measurements (13) are given as:

$$\mathbf{y} = \begin{bmatrix} y_{\mathrm{TAN}} \\ y_{\mathrm{TNO_2}} \\ y_{\mathrm{TNO_3}} \\ y_{\mathrm{pH}} \end{bmatrix} = \begin{bmatrix} [\mathrm{NH_4}^+] + [\mathrm{NH_3}] \\ [\mathrm{HNO_2}] + [\mathrm{NO_2}^-] \\ [\mathrm{HNO_3}] + [\mathrm{NO_3}^-] \\ -\log_{10}\big([\mathrm{H}^+]\big) \end{bmatrix} = \begin{bmatrix} \mathbf{G}\,\mathbf{n}/V \\ -\log_{10}\big([\mathrm{H}^+]\big) \end{bmatrix} \tag{14}$$

where $\mathbf{G}$ is the measurement matrix

$$\mathbf{G} = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}. \tag{15}$$

Clearly, the first three measurements are linear in the species concentrations. In contrast, the pH measurement depends nonlinearly on the proton concentration, which in turn depends nonlinearly on the component concentrations via the algebraic equilibrium relationships in (7). In our simulated experiment, the measurement error variance-covariance matrix is considered time-invariant and defined as follows:

$$\mathbf{\Sigma}_h = \mathbf{\Sigma} = \text{diag}\left(\begin{bmatrix} \sigma^2_{TAN} & \sigma^2_{TNO2} & \sigma^2_{TNO3} & \sigma^2_{pH} \end{bmatrix}^{\text{T}} \right), \quad h = 1, \ldots, H \tag{16}$$

where the diag$(\cdot)$ operator creates a diagonal matrix from a column vector argument.

Data Generation and Problem Formulation

Data Generation

The main objective of this paper is to compare a new method for model identification to a more conventional approach. To properly compare the two methods, simulated measurements are used. These measurements are obtained by solving the DAE system ((6)-(8)) with (3) and (10)-(12) from $t_1 = 0$ to $t_H = 10$ h. Measurements are obtained by means of (13)-(16) at regular intervals of 10 minutes so that $H = 61$. All parameter values used for simulation are given in Table 1.

Problem Formulation

The model identification problem consists in finding an appropriate model based on the measurements from a pulse experiment. For each kinetically controlled reaction, a set of five candidate rate laws are proposed. These are the zeroth-order, first-order, Monod, Tessier, and Haldane rate laws given in Table 2. The initial conditions $\bar{\mathbf{n}}_0$, the stoichiometric matrix $\mathbf{N}$, the equilibrium equations $\mathbf{g}(\cdot)$, the measurement equations $\mathbf{f}(\cdot)$, and the measurement error variance-covariance matrices $\boldsymbol{\Sigma}_h$ are assumed to be known. Hence, the aim is therefore to identify which of the candidate rate laws are appropriate for the two reactions, while also estimating the corresponding kinetic parameters. In this work, feasible values for the parameters are considered to be in the interval $[\,10^{-6},\, 10^{2}\,]$.

Notation. The j^{th} candidate rate law for the i^{th} kinetically controlled reaction is referred to as $r_{k,i}^{(j)}$. The corresponding parameter vectors are $\boldsymbol{\theta}_i^{(j)}$. The number of candidate rate laws for the i^{th} reaction is J_i, so that $j = 1, 2, \ldots, J_i$. For a given choice of rate laws for the kinetically controlled reactions, the parameter vector composed of the joint set of parameter vectors for all reactions is denoted as $\boldsymbol{\Theta}$.

Method 1: Simultaneous Model Identification

The simultaneous model identification procedure is an exhaustive method that consists in building a model for every possible combination of the candidate rate laws ($\mathbf{r}_k$) followed by the estimation of all kinetic parameters ($\boldsymbol{\Theta}$) for each model. As indicated above, we assume that the stoichiometry and equilibrium relations are known and the rate laws and their parameters are to be identified. For a given selection of candidate rate laws, parameter estimation is formulated mathematically as the following weighted least squares (WLS) estimation problem:

$$\hat{\boldsymbol{\Theta}} = \arg\min_{\boldsymbol{\Theta}} \quad \sum_{h=1}^{H} \left(\tilde{\mathbf{y}}(t_h) - \mathbf{y}(t_h)\right)^{\mathrm{T}} \boldsymbol{\Sigma}_h{}^{-1} \left(\tilde{\mathbf{y}}(t_h) - \mathbf{y}(t_h)\right) \tag{17}$$

$$\text{s.t.} \quad \mathbf{y}(t_h) = \mathbf{f}\left(\mathbf{n}(t_h)/V\right) \tag{18}$$

$$\mathbf{g}\left(\mathbf{n}(t)/V\right) = \mathbf{0}_{R_e} \tag{19}$$

$$\bar{\mathbf{E}}\,\mathbf{n}(t) = \bar{\mathbf{n}}(t) \tag{20}$$

$$\bar{\mathbf{n}}(t) = V \int_0^t \bar{\mathbf{N}}_k^{\mathrm{T}}\, \mathbf{r}_k\big(\mathbf{n}(\tau)/V, \boldsymbol{\Theta}\big) d\tau, \qquad \bar{\mathbf{n}}(0) = \bar{\mathbf{n}}_0 \tag{21}$$

$$\boldsymbol{\Theta} = \left[\boldsymbol{\theta}_1^{(j)^{\mathrm{T}}}, \ldots, \boldsymbol{\theta}_i^{(j)^{\mathrm{T}}}, \ldots, \boldsymbol{\theta}_{R_k}^{(j)^{\mathrm{T}}}\right]^{\mathrm{T}} \tag{22}$$

During this estimation, the simulated system (18)-(22) is the same as the data-generating process, except for the rate laws in $\mathbf{r}_k$ and the parameters therein. Because the measurement errors are assumed to be normally distributed according to (13), minimizing the WLS objective corresponds to a maximum-likelihood estimation (MLE).

The optimization problem (17)-(22) is solved by means of the Nelder-Mead simplex algorithm.[31] This algorithm is initiated with parameter values at the center of the feasible intervals considered above. The total number of models whose parameters need to be estimated equals the product of the numbers of candidate rate laws, $\prod_i J_i$. Following the parameter estimation for each of these models, a well-fitting model is selected from the complete set of models by trading off the WLS objective (17) against parsimony. To this end, the WLS objective is equivalently expressed as the weighted root mean squared residual (WRMSR):

$$WRMSR = \sqrt{\frac{1}{H \cdot M} \sum_{h=1}^{H} \left(\tilde{\mathbf{y}}(t_h) - \mathbf{y}(t_h)\right)^{\mathrm{T}} \boldsymbol{\Sigma}_h{}^{-1} \left(\tilde{\mathbf{y}}(t_h) - \mathbf{y}(t_h)\right)} \tag{23}$$

Example. To model the simulated process, five different candidate rate laws are consid-

ered for each of the two biological reactions ($J_1 = J_2 = 5$). The number of distinct models whose parameters are estimated is therefore $\prod_i J_i = J_1 \cdot J_2 = 5 \cdot 5 = 25$.

Method 2: Incremental Model Identification via Extents

This subsection introduces the concept of extents of reaction and shows how to compute them from the measured numbers of moles. The computed extents, named experimental extents, are then used to identify the kinetics of each reaction individually, thereby making the procedure incremental. Finally, the same measurements are used to fine-tune the kinetic parameters for the global model.

Definition of Extents

In batch reactors, the extents of reaction $\mathbf{x}(t)$ can be defined by means of the following integral:

$$\mathbf{n}(t) = \mathbf{n}_0 + V \int_0^t \mathbf{N}^{\mathrm{T}}\, \mathbf{r}\left(\mathbf{n}(\tau)/V\right) d\tau = \mathbf{n}_0 + \mathbf{N}^{\mathrm{T}}\, \mathbf{x}(t). \tag{24}$$

In words, an extent of reaction expresses the progress of the corresponding reaction in terms of the numbers of moles of the product it has produced since $t = 0$. This definition can be applied to multiphase systems as well.[21] In what follows, unless mentioned otherwise, the term extent refers specifically to the extent of a kinetically controlled reaction. Equation (8) can be integrated to give:

$$\bar{\mathbf{n}}(t) = \bar{\mathbf{n}}_0 + V \int_0^t \bar{\mathbf{N}}_k^{\mathrm{T}}\, \mathbf{r}_k\left(\mathbf{n}(\tau)/V\right) d\tau = \bar{\mathbf{n}}_0 + \bar{\mathbf{N}}_k^{\mathrm{T}}\, \mathbf{x}_k(t). \tag{25}$$

Reformulating the balance equations (21)-(20) in terms of extents allows the selection of rate laws and estimating parameters for each reaction individually. To do so, the available

measurements are first transformed into *experimental extents*. After this transformation, and for each reaction individually, selected rate laws can be fitted to the experimental extents. These steps are explained next.

Step 1 – Computation of Experimental Extents

The extents of reaction for the kinetically controlled reactions can be computed by solving the following WLS problem for each sampling instant t_h:

$$\tilde{\mathbf{x}}_k(t_h) = \arg \min_{\mathbf{x}_k(t_h)} \quad (\tilde{\mathbf{y}}(t_h) - \mathbf{y}(t_h))^{\mathrm{T}} \; {\boldsymbol{\Sigma}_h}^{-1} \; (\tilde{\mathbf{y}}(t_h) - \mathbf{y}(t_h)) \tag{26}$$

$$\text{s.t.} \quad \mathbf{y}(t_h) = \mathbf{f}\left(\mathbf{n}(t_h)/V\right) \tag{27}$$

$$\mathbf{g}\left(\mathbf{n}(t_h)/V\right) = \mathbf{0}_{R_e} \tag{28}$$

$$\bar{\mathbf{E}}\, \mathbf{n}(t_h) = \bar{\mathbf{n}}_0 + \bar{\mathbf{N}}_k^{\mathrm{T}}\, \mathbf{x}_k(t_h) \tag{29}$$

where (26) is the objective function, (27) expresses the expected measurements as functions of the numbers of moles of the species, (28) expresses the algebraic equilibria, and (29) relates the extents of the kinetically controlled reactions to the number of moles of the species. As above, minimizing the WLS objective to its global minimum corresponds to MLE. In general, the above problem is nonlinear, non-convex, and therefore solved numerically. In special cases, an analytic solution can be provided.[26]

The initial numbers of moles $\bar{\mathbf{n}}_0$, and volume V are assumed to be known. Hence, one can compute the point-wise approximation $\boldsymbol{\Lambda}_h$ to the variance-covariance matrix of the experimental extents as the inverse of the Fisher information matrix $\mathbf{J}_h^{\mathrm{T}}\, {\boldsymbol{\Sigma}_h}^{-1}\, \mathbf{J}_h$, where $\mathbf{J}_h$ is the Jacobian matrix, with $\mathbf{J}_h(m,i) = \partial y_m / \partial x_{k,i}|_{\tilde{\mathbf{x}}_k(t_h)}$. The elements of $\mathbf{J}_h$ are computed by numerical differentiation unless analytical derivatives are available. This procedure allows writing the following approximate distribution for the extent estimation errors (i.e., the difference between the experimental extents $\tilde{\mathbf{x}}_k$ and the true extents $\mathbf{x}_k$):

$$\tilde{\mathbf{x}}_k(t_h) - \mathbf{x}_k(t_h) \sim \mathcal{N}(\mathbf{0}_{R_k}, \mathbf{\Lambda}_h). \tag{30}$$

Example. In the simulated experiment, it follows from (1), (14), and (25) that the Jacobian consists of three rows that are computed analytically and a fourth row that is evaluated numerically:

$$\mathbf{J}_h = \begin{bmatrix} \frac{1}{V}\,\mathbf{G}\,\bar{\mathbf{E}}^{+}\,\bar{\mathbf{N}}_k^{\mathrm{T}} \\ \left.\frac{\partial pH}{\partial \tilde{\mathbf{x}}_k}\right|_{\tilde{\mathbf{x}}_k(t_h)} \end{bmatrix} \tag{31}$$

with the superscript $(\cdot)^{+}$ indicating the Moore-Penrose pseudo-inverse. The i^{th} element of the last row is computed as $\frac{pH_1 - pH_0}{\delta}$, with pH_0 and pH_1 the pH values obtained by solving (27)-(29) at $\tilde{\mathbf{x}}_k(t_h)$ and $\tilde{\mathbf{x}}_k(t_h) + \Delta_i$, with Δ_i a vector with the small number δ in its i^{th} position and zeros elsewhere.

Step 2 – Extent Modeling

The original identification problem (17)-(22) is now simplified by fitting the rate laws to the experimental extents instead of to the original measurements and by estimating the parameters of a single reaction at the time. The idea is to model each reaction by optimizing the fit to the corresponding experimental extent, $\tilde{x}_{k,i}$, with $i = 1, \ldots, R_k$. However, since the reaction rate $r_{k,i}$ is a function of concentrations that might depend on the progress of several reactions, one estimates the contribution of the *other* reactions from measurements.[21] This results in the following optimization problem for the jth candidate rate law for the ith kinetically controlled reaction:

$$\hat{\boldsymbol{\theta}}_i^{(j)} = \arg\min_{\boldsymbol{\theta}_i^{(j)}} \quad ssq_i := \sum_{h=1}^{H} \frac{\left(\tilde{x}_{k,i}(t_h) - x_{k,i}(t_h)\right)^2}{\lambda_{i,h}} \tag{32}$$

$$\text{s.t.} \quad \mathbf{g}\left(\mathbf{n}(t)/V\right) = \mathbf{0}_{R_e} \tag{33}$$

$$\bar{\mathbf{E}}\,\mathbf{n}(t) = \bar{\mathbf{n}}_0 + \bar{\mathbf{N}}_k^{\mathrm{T}}\,\mathbf{x}_k(t) \tag{34}$$

$$\forall r = 1, \ldots, R_k : \tag{35}$$

$$x_{k,r}(t) = \begin{cases} V \int_0^t r_{k,i}^{(j)}\left(\frac{\mathbf{n}(\tau)}{V}, \boldsymbol{\theta}_i^{(j)}\right) d\tau, \quad x_{k,i}(0) = 0 & \text{if } r = i \\ \mathcal{I}(\mathbf{t}, \tilde{\mathbf{x}}_{k,r}, t), & \text{if } r \neq i \end{cases}$$

where $\lambda_{i,h} := \boldsymbol{\Lambda}_h(i,i)$, $\mathbf{t} = [t_1, t_2, \ldots, t_h, \ldots, t_H]$ and with the operator $\mathcal{I}(\cdot)$ defined as

$$\forall t \in \{t : t_l \leq t \leq t_{l+1}\} : \quad \mathcal{I}(\mathbf{t}, \tilde{\mathbf{x}}_{k,r}, t) := \tilde{x}_{k,r}(t_l) + \Big(\tilde{x}_{k,r}(t_{l+1}) - \tilde{x}_{k,r}(t_l)\Big)\frac{t - t_l}{t_{l+1} - t_l}. \tag{36}$$

In the above problem, (32) is the objective function expressing that the ith predicted extent should be as close as possible to the corresponding experimental extents in the WLS sense. As before, (33) and (34) express the algebraic equilibria and the relationships between the extents of the kinetically controlled reactions and the number of moles of all species. Equation (35) indicates that the predicted extents stem from *(i)* the simulated ith reaction, and *(ii)* piecewise linear interpolation of the experimental extents for the other kinetically controlled reactions. The most important consequence of this method is that only the kinetic parameters of the ith candidate rate law appear in the optimization problem. Indeed, the interpolation of the experimental extents (36) implies that the kinetic parameters of the corresponding reactions are not needed. The original optimization problem is thereby replaced by multiple optimization problems involving a univariate system including only one reaction. Furthermore, the modification also means that one does not need to know the structure of the rate laws corresponding to the interpolated experimental extents, that

is, the best candidate rate law for each reaction can be found independently of the rate laws for the other reactions.

The second method allows solving each individual parameter estimation problem to global optimality by means of the branch-and-bound algorithm proposed earlier.[20] This way, the best parameter values are guaranteed to be found within the considered feasible intervals. The bounding procedures required for this algorithm are given in the Supporting Information. With each candidate rate law and the associated optimal parameters $\hat{\boldsymbol{\theta}}_i^{(j)}$, one obtains the modeled extent $\hat{x}_{k,i}^{(j)}$ and the following extent-specific WRMSR:

$$WRMSR_i^{(j)} = \sqrt{\frac{1}{H} \sum_{h=1}^{H} \frac{\left(\tilde{x}_{k,i}(t_h) - \hat{x}_{k,i}^{(j)}(t_h)\right)^2}{\lambda_{i,h}}}. \tag{37}$$

The rate law $\hat{r}_{k,i}$ is selected by trading off the WRMSR against parsimony. This is repeated for every reaction, which means that the number of parameter estimation problems to be solved now equals the sum of the numbers of candidate rate laws, $\sum_i J_i$. In addition, the number of parameters that are estimated in each problem is generally lower than the number of parameters estimated with the first method (17)-(22).

Example. With 5 candidate rate laws considered for each reaction, $\sum_i J_i = J_1 + J_2 = 10$ instances of the parameter estimation problem need to be solved. The number of parameters that are estimated in each problem ranges from 1 (e.g. zeroth-order rate law) to 3 (Haldane). In comparison, the first method requires the estimation of 2 (zeroth-order rate law for both reactions) up to 6 (Haldane rate law for both reactions) parameters at once.

Step 3 – Model Fine-Tuning

Following the rate-law selection, the model parameters are fine-tuned by simultaneously estimating all kinetic parameters via (17)-(22). As in the first method, this is done using

the Nelder-Mead simplex algorithm. In contrast to the first method, this algorithm is now executed for only one model containing the rate laws selected in Step 2 and is initiated with the corresponding parameter estimates obtained in Step 2.

Results

Process Simulation

The nitrification model (6)-(15) is used to generate concentration and pH measurement series. The results are shown in Fig. 1. One can see a fairly distinct separation in time of the two reactions with the TNO2 concentration rising to 50% of the original TAN concentration at about 4 h. Before (after) this time, a net production (consumption) of TNO2 is observed. The figure also shows the free ammonia concentration [NH_3]. The ammonia oxidation stops when this concentration reaches zero. The nitrite and nitrate ion concentrations are nearly indistinguishable from the total nitrite and total nitrate concentrations (not shown). At the end of the experiment, about half of the available TAN is converted via nitrite to nitrate. The limited buffering capacity in the simulated system causes fairly large changes in pH. Additive Gaussian noise is simulated added to generate realistic measurements.

Method 1: Simultaneous Model Identification

The kinetic parameters of 25 different models, each with a unique pair of rate laws for the first and second reactions, are estimated by solving (17)-(22). The resulting WRMSR values shown in Fig. 2 range from 6.57 to 37.86. These values indicate the model prediction error standard deviation relative to the measurement error standard deviation. Assuming the correct model, it exhibits a χ^2-distribution with a mean value of 1 and a right-sided 99% confidence limit of 1.11. The graph also shows the WRMSR value of 1.01 obtained with the noisy measurements and the true model including its parameters. This WRMSR is very close to the expected value of 1. Note that the best model gives a WRMSR value that is

6.57 times larger than the WRMSR value obtained with method 2 (see below). Clearly, this method is unfit to find a good model. In all cases, including the case involving the true rate laws used for simulation, only a locally optimal parameter set could be found. In addition, the best model (Model 16) includes the Tessier rate law for the first reaction and the zeroth-order rate law for the second reaction, which does not correspond to the true rate laws. Additional results, including simulations using each of the 25 models after parameter estimation, are included in the Supporting Information.

Method 2: Incremental Model Identification via Extents

Step 1 – Extent Computation

The extents computed by solving (26)-(29) using the TAN, TNO2, TNO3, and pH measurements are shown in Fig. 3(a). The confidence bands for the experimental extents vary with time, in particular for the first extent. High precision is obtained at the beginning and during most of the second half of the experiment. However, during the first half, the uncertainty first increases and then decreases. At the end of the experiment, the uncertainty increases again. These effects are due to the nonlinear propagation of the pH measurement error through the measurement and algebraic equations. The ellipsoidal confidence regions at 0.5, 1.5, 2.5, 3.5, and 4.5 h are shown in Fig. 3(b). The orientation of the confidence region becomes more oblique with increased uncertainty in the first extent.

Step 2 – Extent Modeling

The global solutions to (32)-(35), obtained for every reaction and every candidate rate law, are discussed next.

Modeling the First Extent. The best fits of the first extent obtained with the various candidate rate laws are shown in Fig. 4(a). It is clear that the zeroth- and first-order models do not fit the experimental extents well. The Monod and Tessier models fit better, yet they

over-estimate the experimental extent. This is clearly visible in Fig. 4(b-c), where the model errors are shown. In contrast, the Haldane rate law fits the extent profile well. As such, the Haldane model is easily selected as the best among the model candidates. In Fig. 5, the WRMSR values (37) are given with a 95% upper control limit based on the corresponding χ^2-statistics. Based on this statistic, all models except for the Haldane model are rejected for the first extent.

Modeling the Second Extent. The best fits of the second extent are visualized in Fig. 6(a). Here, all rate models fit the experimental extents reasonably well, except for the zeroth-order model. The first-order model leads to visibly auto-correlated residuals Fig. 6(b-c). This is also evident from the WRMSR values (Fig. 5), on the basis of which the zeroth-order and first-order models are rejected. In this case, a parsimonious model is chosen among the three remaining candidates. The Monod model delivers the best fit among the simplest candidates (Monod and Tessier). An alternative approach may consist in designing an experiment that enables better discrimination of the remaining rate laws. This is not explored in this work.

Step 3 – Model Fine-Tuning

The model structure consisting of the two selected rate laws, namely, Haldane and Monod, is used next to fine-tune the model parameter via the simultaneous approach (17)-(22). Fig. 7 compares the simulated concentration and pH values with the predictions of the identified models prior and after fine-tuning. These three simulations are hard to distinguish from each other. The resulting overall WRMSR (23) equals 1.0013 and is shown in Fig. 2. Most importantly, the extent-based model identification procedure has delivered a well-fitting set of rate laws and kinetic parameter estimates. Furthermore, the selected rate laws are exactly those used to generate the simulated experimental measurements. The parameter estimates deviate at most 10% from their true values, except $\theta_{AOB,2}$ which deviates by about 30%. Such

deviations are typical for biokinetic wastewater treatment models and are in part explained by correlation between parameter estimates.

Discussion

The results presented above are now interpreted in a broader biokinetic modeling context.

Interpretation of the Results. In this study, the concept of extents is introduced for the first time for the purpose of dynamic modeling of an environmental biochemical process. By means of a simplified biokinetic model of the urine nitrification process and simulated batch experiments, several benefits of the extent-based modeling approach have been demonstrated. Concretely, the identification of biokinetic models via extents:

- allows using deterministic optimization methods to obtain excellent parameter estimates. Despite the fact that the individual extent modeling steps only approximate the original model identification problem, one can obtain a well-fitting model. Most importantly, the convergence to local optima as observed with a conventional parameter estimation method can be avoided.
- provides an intuitive diagnostic tool for modeling. Indeed, extent-modeling indicates whether a reaction can be modeled appropriately with a given candidate rate law, thus allowing modelers to pay more attention to reactions that are more difficult to model. Similarly, this approach indicates whether sufficient information is available within a given experimental data set to discriminate between candidate rate laws.
- reduces a model selection problem that is polynomial in the number of candidate rate laws to a model selection problem that is linear in this number. In this study, the extent-based modeling method required solving 10 parameter estimation problems involving 1 to 3 parameters, whereas the conventional simultaneous approach required solving 25 parameter estimation problems involving 2 to 6 parameters.

It is of special importance that the extent-based model identification method is the only method delivering an acceptable model. Indeed, the conventional model identification method did not result in an acceptable model, despite the apparent simplicity of the studied process and simulated experiment.

Links in Prior Work. While the concept of extents is new in the context of dynamic modeling of environmental processes, it is important to note that a number of important concepts in use today are somewhat similar. For instance, the integral defined by the area under the oxygen uptake rate curve, a.k.a. *respirogram*, is matched to the total accumulated oxygen uptake in typical respirometric experiments.[32] Similar concepts include *accumulated methane production*[33] and *number of base pulses*.[34] It is also interesting to note that the accumulated cellulose solubilisation has been described as the *extent of solubilisation*,.[35] However, this is without links to the general concept of extents. The most important difference between extents and the concepts already in use is that extents reflect individual processes rather than several simultaneous processes. So far, model reduction on the basis of the concept of reaction invariants[36] is the only related application known in the environmental engineering sciences. We expect tangible benefits from a broader and systematic use of extents, including those mentioned above.

Analysis of the Extent-based Modeling Method. In the general case, the extent-based modeling method does not solve the exact same problem as the conventional simultaneous modeling method. Extent-based modeling solves the same problem (17)-(22) only if *(i)* the measurement equations are linear and there are no nonlinear algebraic equations involved in the extent computations (26)-(29), *(ii)* the off-diagonal elements of the matrices $\mathbf{\Lambda}_h$ are equal to zero, that is, in absence of correlation between experimental extents, and *(iii)* the reaction rates can be expressed as functions of the modeled extents. These requirements are rarely satisfied so that the resulting parameter estimates likely deviate from those obtained by solving (17)-(22). However, the extent-based modeling framework is particularly

useful when solving (17)-(22) to global optimality is difficult or computationally prohibitive.

In computing the solution to (26)-(29), one can encounter different situations:

1. The first situation occurs when the available measurements are linear in the extents of the kinetically controlled reactions and do not depend on the equilibrium species concentrations. In this case, one can discard all nonlinear (equilibrium) equations and an analytic solution for the extents can be found.

2. The second situation occurs when the number of measured variables matches the number of computed extents exactly, thereby resulting in a fully determined system (hence no need for optimization). In this situation, one can find extents that make the objective function (26) equal to zero, while satisfying (27)-(29). The solution can therefore be obtained by solving the equation system (27)-(29) numerically. In the process considered in this work, this situation would occur if the pH and one of the remaining variables (TAN, TNO2, TNO3) were measured (not demonstrated).

3. The third situation occurs when the number of measured variables exceeds the number of computed extents (overdetermined system). This corresponds to the case studied in this work (TAN, TNO2, TNO3, and pH measured). One approach consists of discarding (26) and solving (27)-(29) in a least-squares sense.[26] When doing so, the experimental extents are not the solution to (26)-(29). We recommend solving (26)-(29) exactly, as in this work, to obtain experimental extents that are WLS-optimal.

In its current form, the proposed extent-based modeling method assumes a closed batch process whose stoichiometric matrix and the algebraic equilibrium equations are known or estimated precisely. However, this is not true in general. The method presented here can easily be expanded to account for mass transfer as well as gas-liquid transfer as demonstrated already.[26,37] The main reason this has not been included here is to maintain a clear presentation of the developed method. Not knowing the stoichiometric matrix or the algebraic equilibrium equations means that, prior to modeling via extents, one may use target factor

analysis[38] to identify the stoichiometric matrix or detailed physico-chemical analysis to obtain a model for acid-base and salt speciation. However, certain situations allow using the extent-based modeling framework to estimate equilibrium parameters[26] as well as stoichiometric parameters.[39] Even more critical is the fact that extent-based model identification requires at least as many measured variables as there are kinetically controlled reactions. When this requirement is not met, one can opt to partition the model identification problem into smaller problems which include more than one reaction.[39]

Methodological Improvements. Methodologically speaking, this work adds four elements to the extent-based modeling framework, namely:

- Extent computation with measurements that are nonlinear in the species concentrations.
- Optimal estimation of the experimental extents when (27)-(29) involves more measured variables than extents, that is, in the overdetermined case.
- Accounting for nonlinear effects during experimental extent computation by means of a Laplacian approximation of their distribution.
- Extent-based modeling and deterministic global optimization are combined for the first time into a single model identification framework.

Future Work. The developments in this study are considered critical steps towards a first real-world application of the extent-based modeling of environmental processes. However, the following aspects call for further development and testing of the method prior to experimental validation in full-scale wastewater treatment systems:

- Realistic sensor data. So far, measurement devices are considered to exhibit an instantaneous response within the extent-based modeling framework. However, typical devices respond dynamically to the measured variable.[40,41] Explicit accounting of sensor dynamics is not feasible yet in the extent-based modeling framework.

- Laboratory validation. Several aspects of real biological processes have been ignored to facilitate the introduction of extent-based modeling. The ignored elements include *(i)* bacterial growth and decay processes, *(ii)* complex composition of actual wastewater, and *(iii)* complex physico-chemical reaction systems in high-strength wastewater such as source-separated urine. The first element only affects the extent-based methodology due to a lack of extent observability. This can be accounted for in special cases[39] but may prove difficult to address in general.[42] The second and third element affect both modeling methods used in this study and are being addressed currently by adopting a more realistic physico-chemical urine composition and associated reaction system in view of a lab-scale validation.
- Prior knowledge. In this work, the reactor volume V, the initial conditions $\bar{\mathbf{n}}_0$, and the stoichiometric matrix $\mathbf{N}$ are considered known. Methods permitting the estimation of these variables and parameters remain to be investigated.
- Completeness of the candidate rate laws. In this work, we have assumed that the set of candidate rate laws includes the true rate laws in the data-generating process. This is not true in general. An alternative model structure based on shape constrained splines can address this problem.[43] So far, this type of models has only been applied to monoculture processes. Its use in connection with the extent-based model identification remains to be evaluated.

Acknowledgement

This study was made possible by Eawag Discretionary Funds (grant no.: 5221.00492.009.03, project: DF2015/EMISSUN). All results were obtained by use of Matlab[44] and the Spike_O toolbox for optimization.[45]

Table 1: List of symbols and parameter values used for simulation. Values in parentheses refer to the best-available estimates.

Symbol	Description	Value	Unit
c_S	Substrate concentration	–	$\mathrm{mol{\cdot}L^{-1}}$
$c_{\mathrm{TAN},0}$	Initial TAN concentration	0.35	$\mathrm{mol{\cdot}L^{-1}}$
$c_{\mathrm{TH},0}$, $c_{\mathrm{TOH},0}$, $c_{\mathrm{O_2},0}$	Initial concentrations	–	$\mathrm{mol{\cdot}L^{-1}}$
$\bar{\mathbf{E}}$	Matrix defining the set of kinetic species and conserved molecular constituents	–	–
$\mathbf{f}(\cdot)$	Measurement expressions	–	
$\mathbf{G}$	Measurement gain matrix	–	–
$\mathbf{g}(\cdot)$	Algebraic equilibrium expressions	–	
H	Number of samples	61	–
h	Measurement sample index	–	–
i	Reaction index	–	–
$\mathbf{J}_h$	Jacobian matrix at the hth sample time	–	
J_i	Number of rate law candidates for the ith reaction	–	–
j	Rate law candidate index	–	–
M	Number of measurements	–	–
m	Measured variable index	–	–
$\mathbf{N}$, $\bar{\mathbf{N}}$, $\bar{\mathbf{N}}_k$	Stoichiometric matrices	–	–
$\mathbf{n}$	Numbers of moles of all species	–	mol
$\bar{\mathbf{n}}$	Numbers of moles of kinetic species and conserved molecular constituents	–	mol
$\mathbf{n}_0$, $\mathbf{n}_{k,0}$, $\bar{\mathbf{n}}_0$	Initial numbers of moles	–	mol
$\mathbf{n}_c$	Numbers of moles of conserved molecular constituents	–	mol
$\mathbf{n}_e$	Numbers of moles of equilibrium species	–	mol
$\mathbf{n}_k$	Numbers of moles of kinetic species	–	mol
$pK_{a,\mathrm{NH_4^+}}$	Logarithmic acid dissociation constant of $\mathrm{NH_4^+}$	+9.24	–
$pK_{a,\mathrm{HNO_2}}$	Logarithmic acid dissociation constant of $\mathrm{HNO_2}$	+3.25	–
$pK_{a,\mathrm{HNO_3}}$	Logarithmic acid dissociation constant of $\mathrm{HNO_3}$	−1.40	–

Table 1: List of symbols and parameter values used for simulation. Values in parentheses refer to the best-available estimates.

Symbol	Description	Value	Unit
pK_w	Logarithmic water dissociation constant	14	–
R	Number of reactions	6	–
R_e	Number of equilibrium reactions	4	–
R_k	Number of kinetic reactions	2	–
r	Reaction index	–	–
$\mathbf{r}$	Reaction rates	–	$\mathrm{mol{\cdot}L^{-1}{\cdot}h^{-1}}$
$\mathbf{r}_k$ ($r_{k,i}$)	Reaction rates of the kinetically controlled reactions (of the ith kinetically controlled reaction)	–	$\mathrm{mol{\cdot}L^{-1}{\cdot}h^{-1}}$
$\hat{r}_{k,i}$	Selected rate law for the ith kinetically controlled reaction	–	–
S	Number of chemical species	10	–
$\bar{S}$	Number of kinetic species and conserved molecular constitutants	6	–
S_c	Number of conserved molecular constitutants	5	–
S_e	Number of equilibrium species	9	–
S_k	Number of kinetic species	1	–
t, t_h	Time (of measurement)	–	h
V	Volume	1	L
$WRMSR$	Weighted root mean squared residual	–	–
$WRMSR_i^{(j)}$	WRMSR of the ith reaction with the jth candidate rate law	–	–
$\mathbf{x}_k$ ($x_{k,i}$, $x_{k,r}$)	Extents of the kinetically controlled reactions (of the i/rth kinetically controlled reaction)	–	mol
$\tilde{\mathbf{x}}_k$ ($\tilde{x}_{k,i}$, $\tilde{\mathbf{x}}_{k,r}$, $\tilde{x}_{k,r}$)	Experimental extents of kinetically controlled reactions (of the i/rth kinetically controlled reaction)	–	mol
$\hat{x}_{k,i}^{(j)}$	Extent estimate for the ith kinetically controlled reaction with the jth rate law candidate	–	mol
$\mathbf{y}$	Measured variables	–	
$\tilde{\mathbf{y}}$	Measurements	–	

Table 1: List of symbols and parameter values used for simulation. Values in parentheses refer to the best-available estimates.

Symbol	Description	Value	Unit
y_{TAN}, $y_{\mathrm{TNO_2}}$, $y_{\mathrm{TNO_3}}$, y_{pH}	Measured variables	–	
Δ (Δ_i)	Perturbation vector (for the ith reaction)	–	mol
δ	Perturbation parameter	$1 \cdot 10^{-12}$	mol
$\boldsymbol{\Theta}$	Kinetic parameters for all rate laws	–	
$\hat{\boldsymbol{\Theta}}$	Kinetic parameter estimates for all rate laws	–	
$\boldsymbol{\theta}$ ($\boldsymbol{\theta}_i^{(j)}$)	Kinetic parameters (for the jth rate law candidate of ith reaction)	–	
$\hat{\boldsymbol{\theta}}_i^{(j)}$	Parameter estimates for jth kinetic rate law candidate for the ith kinetically controlled reaction	–	
$\theta_{AOB,1}$	Kinetic parameter for AOB activity	0.025 (0.024)	h
$\theta_{AOB,2}$	Kinetic parameter for AOB activity	0.1 (0.13)	$\mathrm{h{\cdot}L{\cdot}mol^{-1}}$
$\theta_{AOB,3}$	Kinetic parameter for AOB activity	2.5 (2.4178)	$\mathrm{h{\cdot}L^2{\cdot}mol^{-2}}$
$\theta_{NOB,1}$	Kinetic parameter for NOB activity	$0.11 \cdot 10^{-3}$ ($0.1 \cdot 10^{-3}$)	h
$\theta_{NOB,2}$	Kinetic parameter for NOB activity	1.1 (1.13)	$\mathrm{h{\cdot}L{\cdot}mol^{-1}}$
$\boldsymbol{\Lambda}$ ($\boldsymbol{\Lambda}_h$)	Extent variance-covariance matrix (for the hth sample)	–	$\mathrm{mol^2}$
λ ($\lambda_{i,h}$)	Extent variance (for the ith reaction and the hth sample)	–	$\mathrm{mol^2}$
$\boldsymbol{\Sigma}$ ($\boldsymbol{\Sigma}_h$)	Measurement error variance-covariance matrix (for the hth sample)	–	
σ_{TAN}	Measurement standard deviation for TAN	0.01	$\mathrm{mol \cdot L^{-1}}$
σ_{TNO2}	Measurement standard deviation for TNO2	0.01	$\mathrm{mol \cdot L^{-1}}$
σ_{TNO3}	Measurement standard deviation for TNO3	0.01	$\mathrm{mol \cdot L^{-1}}$
σ_{pH}	Measurement standard deviation for pH	0.05	–
τ	Integrand (time)	–	h
$[\cdot]$	concentration symbol equivalent to c	–	$\mathrm{mol \cdot L^{-1}}$

Table 2: List of candidate rate laws used for both nitritation and nitratation reactions. The substrate concentration c_S is the free ammonia concentration [NH_3] for the nitritation and the free nitrous acid concentration [HNO_2] for the nitratation.

Name	Index j	Candidate rate law $r_{k,i}^{(j)}(c_S, \boldsymbol{\theta}_i^{(j)})$, $i \in \{1,2\}$	Parameter vector $\boldsymbol{\theta}_i^{(j)}$
Zeroth order	1	$\begin{cases} 1/\theta_{i,1}^{(1)} & \text{if } c_S \geq 0 \\ 0 & \text{otherwise} \end{cases}$	$[\,\theta_{i,1}^{(1)}\,]$
First order	2	$\frac{c_S}{\theta_{i,1}^{(2)}}$	$[\,\theta_{i,1}^{(2)}\,]$
Monod	3	$\frac{c_S}{\theta_{i,1}^{(3)}+\theta_{i,2}^{(3)}\ c_S}$	$[\ \theta_{i,1}^{(3)} \quad \theta_{i,2}^{(3)}\ ]^{\mathrm{T}}$
Tessier	4	$\frac{1-\exp\left(-c_S\ \theta_{i,2}^{(4)}/\theta_{i,1}^{(4)}\right)}{\theta_{i,2}^{(4)}}$	$[\ \theta_{i,1}^{(4)} \quad \theta_{i,2}^{(4)}\ ]^{\mathrm{T}}$
Haldane	5	$\frac{c_S}{\theta_{i,1}^{(5)}+\theta_{i,2}^{(5)}\ c_S+\theta_{i,3}^{(5)}\ c_S^2}$	$[\ \theta_{i,1}^{(5)} \quad \theta_{i,2}^{(5)} \quad \theta_{i,3}^{(5)}\ ]^{\mathrm{T}}$

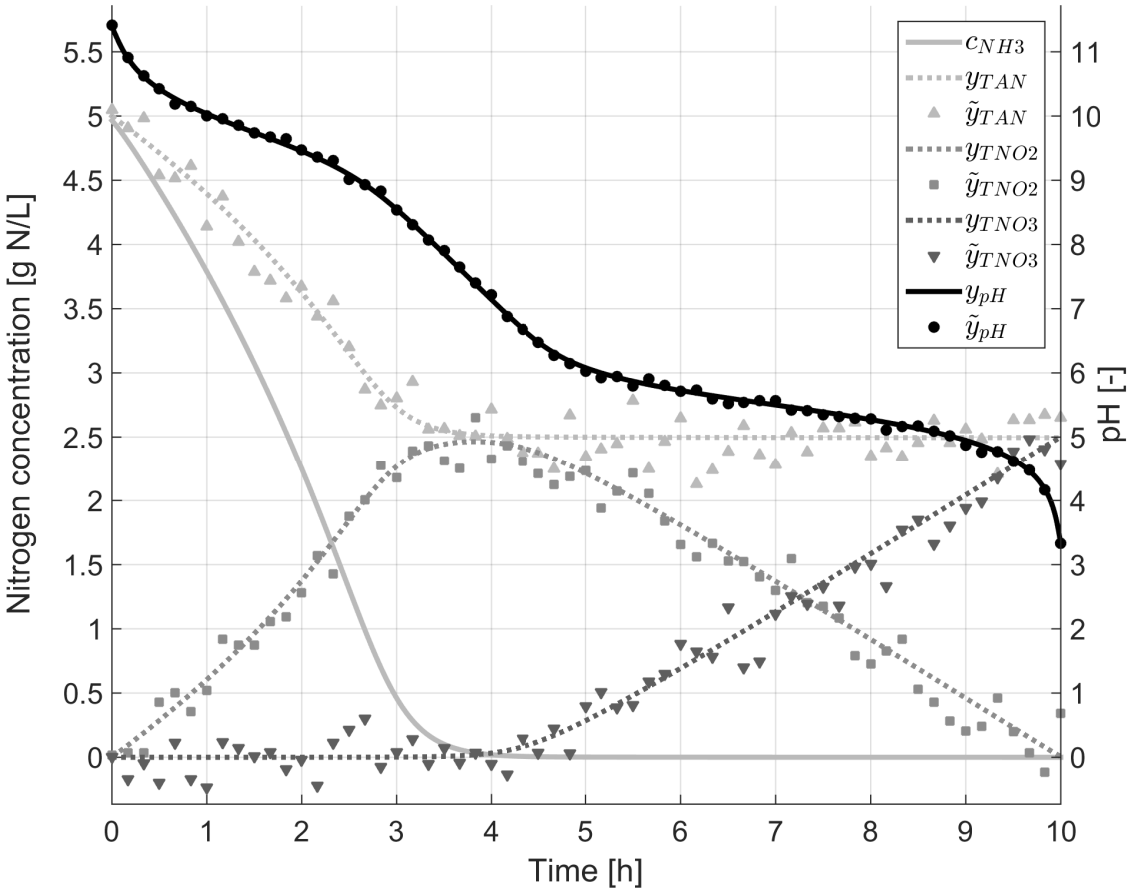

Figure 1: **Data generation:** Simulated concentration (in gN/L) and pH (continuous and dashed lines) with corresponding measurements (dots, squares, triangles).

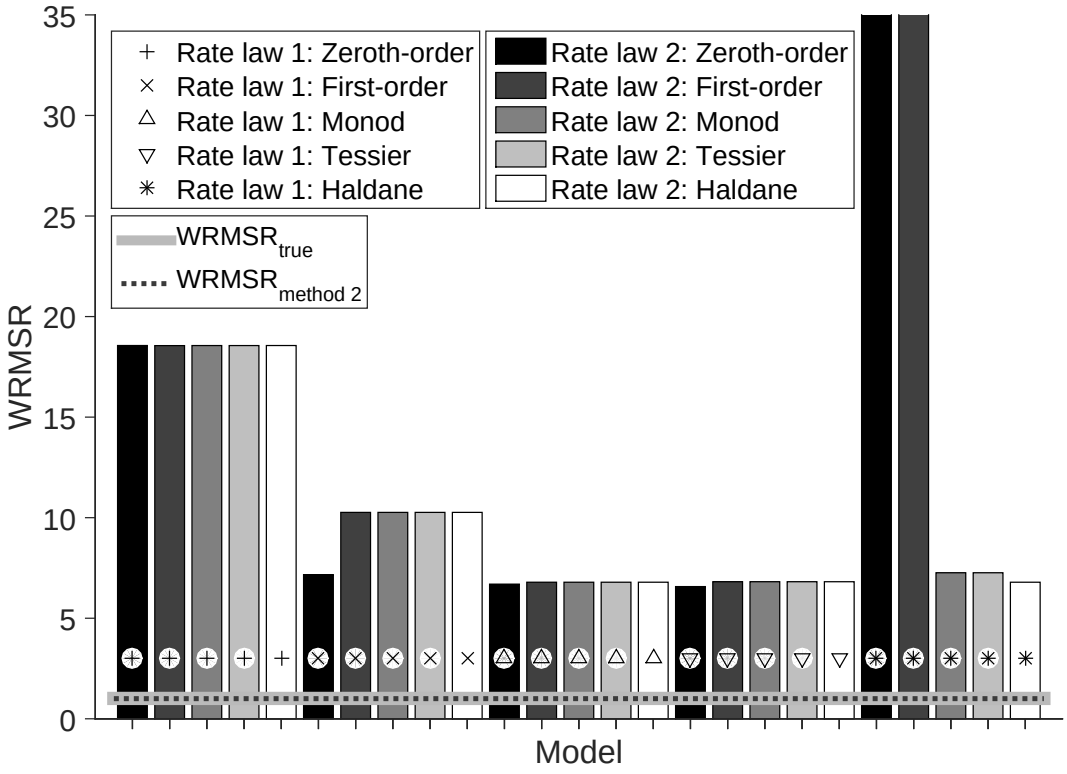

Figure 2: **Method 1: simultaneous model identification.** WRMSR values for 25 models. The markers indicate the selected rate law for the first rate law. Shading of the bars indicates the selected rate law for the second reaction. The WRMSR values for the true model and for the best model obtained with Method 2 are indicated by a full and a dashed line, respectively.

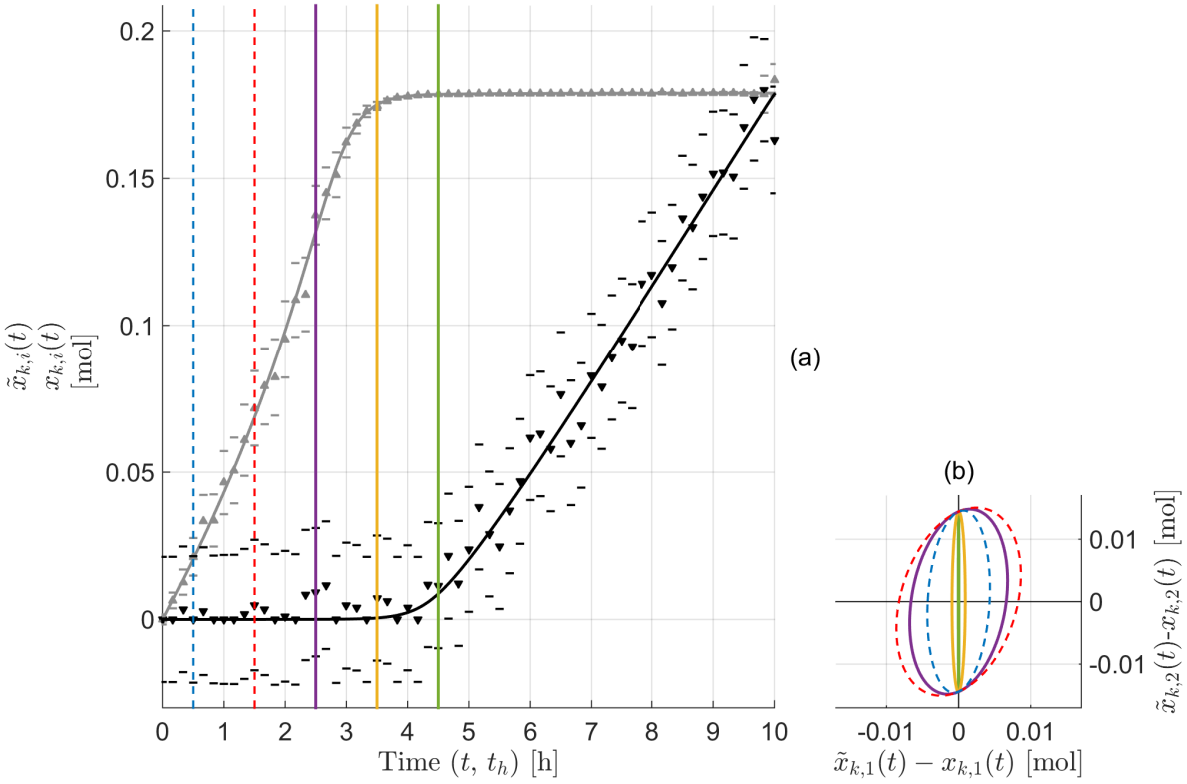

Figure 3: **Method 2: extent-based modeling – Step 1: Computation of experimental extents.** (a) True (lines) and experimental (dots) extents with 3σ confidence intervals for the nitritation (gray) and nitratation (black) reactions. (b) Variance-covariance matrix as 3σ confidence region (ellipsoid) for the experimental extent errors around $(0, 0)$; colored lines corresponding to $t_h = 0.5, 1.5, 2.5, 3.5$ and 4.5 h indicated with matching colors and styles in (a).

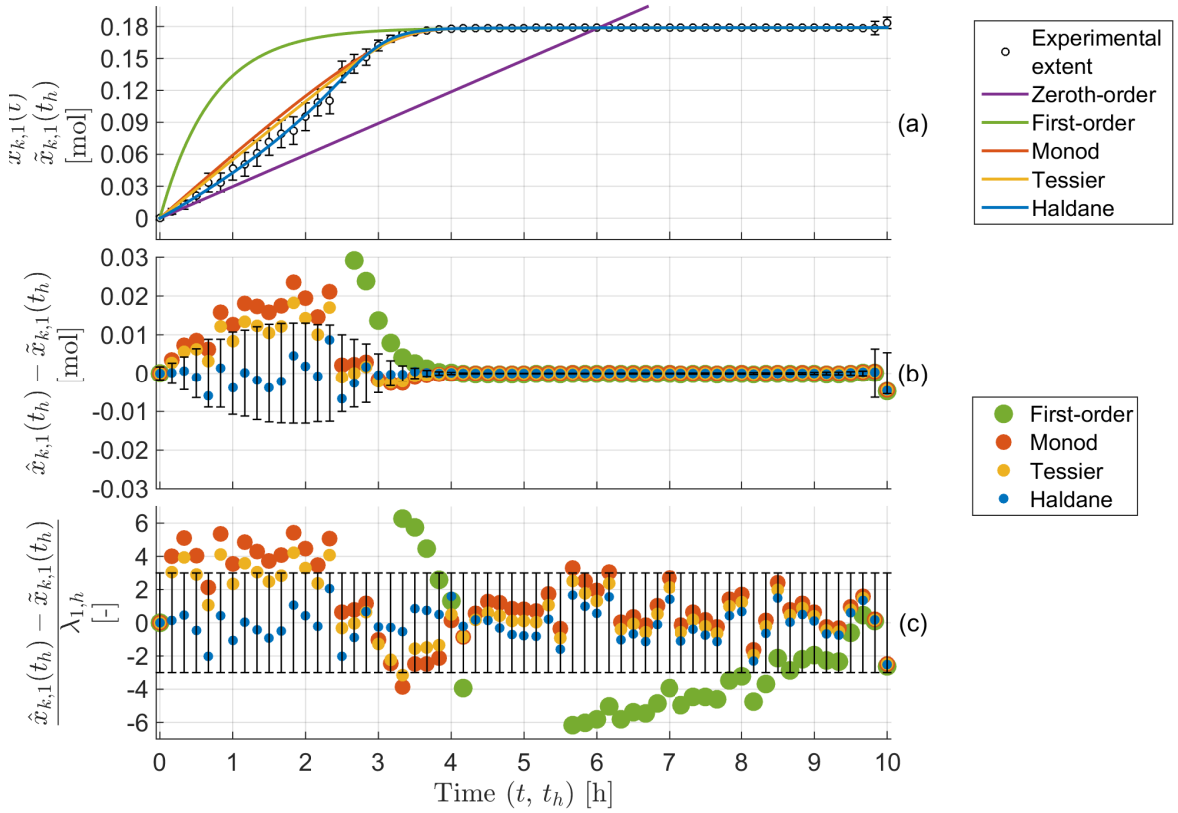

Figure 4: **Method 2: Extent-based modeling – Step 2: Modeling of the 1st extent.** (a) Experimental (circles, with error bars) and modeled (continuous lines) extents as functions of time; (b) Residuals between modeled and experimental extents as functions of time; (c) Normalized residuals between modeled and experimental extents as functions of time.

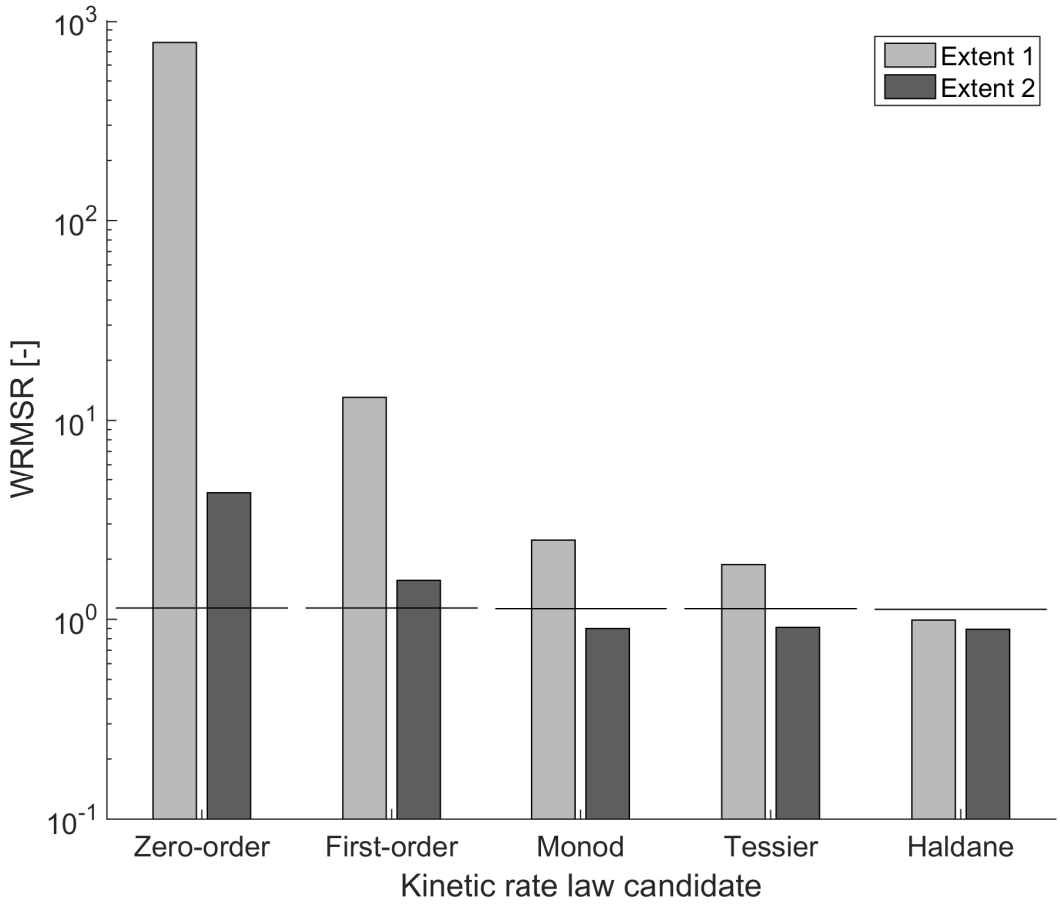

Figure 5: **Method 2: extent-based modeling – Step 2: Modeling of extents – Lack-of-Fit.** WRMSR for all extents and all candidate rate laws (bars) and 95% upper control limits of the associated χ^2-distribution (lines).

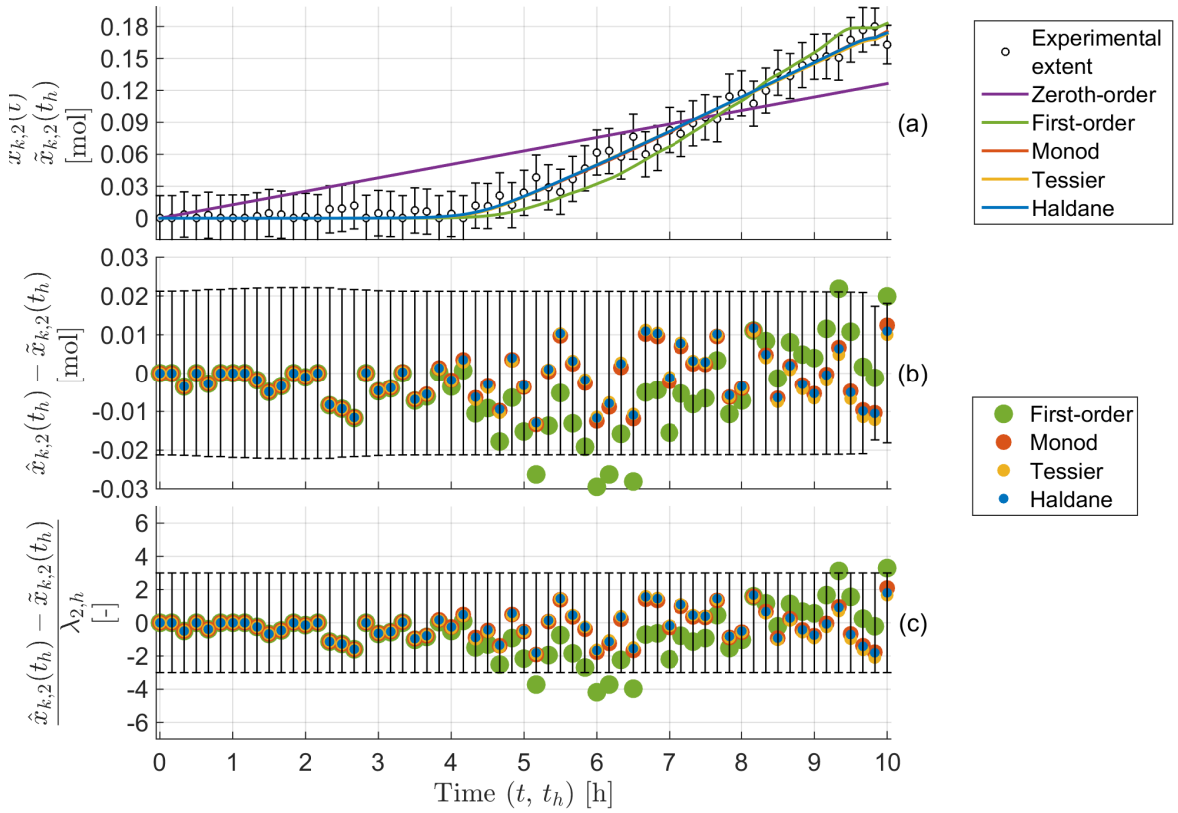

Figure 6: **Method 2: extent-based modeling – Step 2: Modeling of the 2nd extent.** (a) Experimental (circles, with error bars) and modeled (continuous lines) extents as functions of time; (b) Residuals between modeled and experimental extents as functions of time; (c) Normalized residuals between modeled and experimental extents as functions of time.

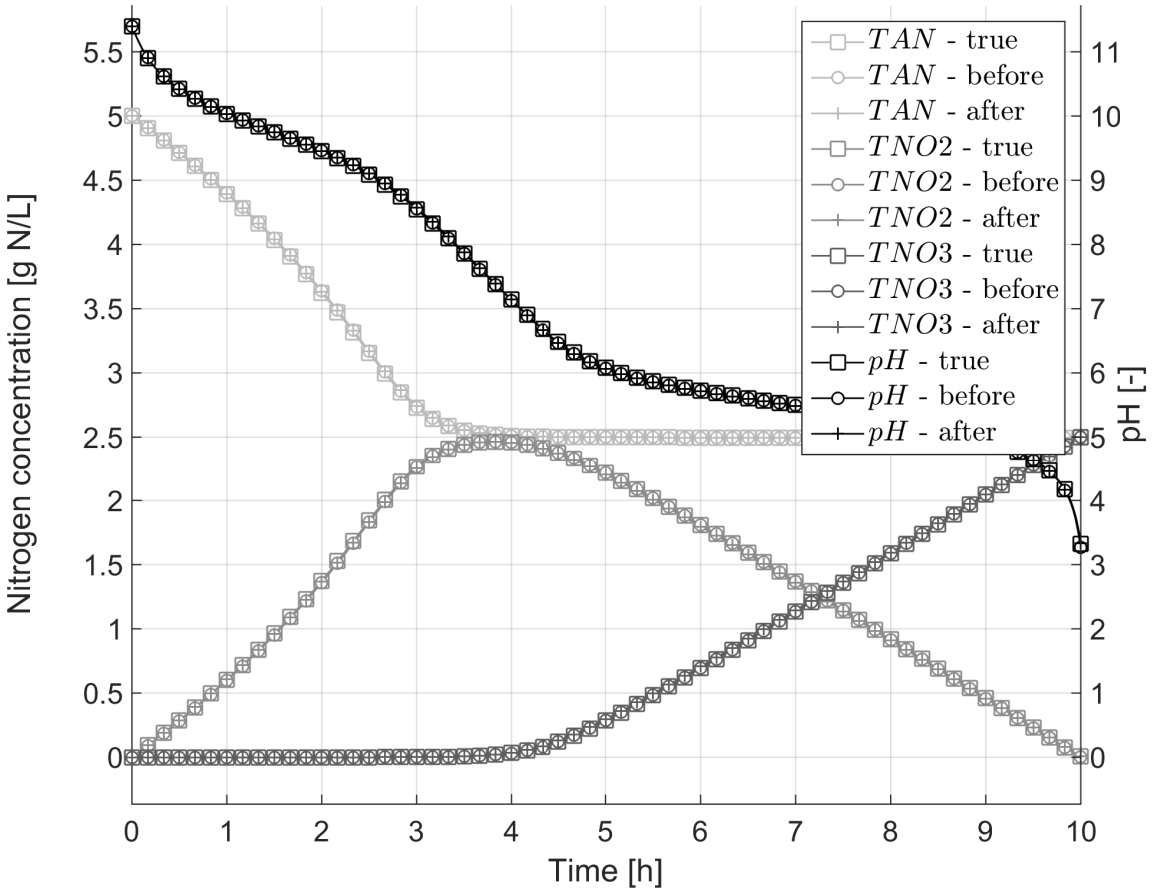

Figure 7: **Method 2: extent-based modeling – Step 3: Model fine-tuning.** Simulation of the TAN, TNO2, and TNO3 concentrations and pH for *(i)* the true data-generating model, *(ii)* the model obtained before fine-tuning, and *(iii)* the model obtained after fine-tuning. Differences between these simulations are barely noticable.

Supporting Information Available

Supporting Information includes bounding procedures, additional figures, and all code to produce our results.

This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Ni, B. J.; Peng, L.; Law, Y.; Guo, J.; Yuan, Z. Modeling of nitrous oxide production by autotrophic ammonia-oxidizing bacteria with multiple production pathways. *Environmental Science & Technology* **2014**, *48*, 3916–3924.

(2) Liu, L.; Binning, P. J.; Smets, B. F. Evaluating alternate biokinetic models for trace pollutant cometabolism. *Environmental Science & Technology* **2015**, *49*, 2230–2236.

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

(4) Jenkins, D.; Wanner, J. Activated Sludge – 100 Years and Counting. *Water Intelligence Online* **2014**, *13*.

(5) Brun, R.; Reichert, P.; Künsch, H. R. Practical identifiability analysis of large environmental simulation models. *Water Resources Research* **2001**, *37*, 1015–1030.

(6) Liu, C.; Zachara, J. M. Uncertainties of Monod kinetic parameters nonlinearly estimated from batch experiments. *Environmental Science & Technology* **2001**, *35*, 133–141.

(7) Nelly, N.; Müller, T. G.; Gyllenberg, M.; Timmer, J. Quantitative analyses of anaerobic wastewater treatment processes: identifiability and parameter estimation. *Biotechnology and Bioengineering* **2002**, *78*, 89–103.

(8) Brockmann, D.; Rosenwinkel, K.-H.; Morgenroth, E. Practical identifiability of biokinetic parameters of a model describing two-step nitrification in biofilms. *Biotechnology and Bioengineering* **2008**, *101*, 497–514.

(9) Neumann, M. B.; Gujer, W. Underestimation of uncertainty in statistical regression of environmental models: influence of model structure uncertainty. *Environmental Science & Technology* **2008**, *42*, 4037–4043.

(10) Bennett, N. D.; Croke, B. F.; Guariso, G.; Guillaume, J. H.; Hamilton, S. H.; Jakeman, A. J.; Marsili-Libelli, S.; Newham, L. T.; Norton, J. P.; Perrin, C.; Pierce, S. A. Characterising performance of environmental models. *Environmental Modelling & Software,* **2013**, *40*, 1–20.

(11) Brugnach, M.; Pahl-Wostl, C.; Lindenschmidt, K.; Janssen, J.; Filatova, T.; Mouton, A.; Holtz, G.; Van der Keur, P.; Gaber, N. *Developments in Integrated Environmental Assessment, Vol.3*; Elsevier, 2008; Chapter 4 - Complexity and uncertainty: Rethinking the modelling activity, pp 49–68.

(12) Sin, G.; Ödman, P.; Petersen, N.; Lantz, A. E.; Gernaey, K. V. Matrix notation for efficient development of first-principles models within PAT applications: Integrated modeling of antibiotic production with *Streptomyces coelicolor*. *Biotechnology and Bioengineering* **2008**, *101*, 153–171.

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

(14) Müller, T.; Dürr, R.; Isken, B.; Schulze-Horsel, J.; Reichl, U.; Kienle, A. Distributed modeling of human influenza a virus-host cell interactions during vaccine production. *Biotechnology and Bioengineering* **2013**, *110*, 2252–2266.

(15) Zambrano-Bigiarini, M.; Rojas, R. A model-independent Particle Swarm Optimisation software for model calibration. *Environmental Modelling & Software* **2013**, *43*, 5–25.

(16) Larsen, T. A., Udert, K. M., Lienert, J., Eds. *Source separation and decentralization for wastewater management*; IWA Publishing, 2013.

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

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

(19) Van De Steene, M.; Van Vooren, L.; Ottoy, J. P.; Vanrolleghem, P. A. Automatic buffer capacity model building for advanced interpretation of titration curves. *Environmental Science & Technology* **2002**, *36*, 715–723.

(20) Mašić, A.; Udert, K.; Villez, K. Global parameter optimization for biokinetic modeling of simple batch experiments. *Environmental Modelling and Software* **2016**, *85*, 356–373.

(21) Bhatt, N.; Amrhein, M.; Bonvin, D. Incremental Identification of Reaction and Mass-Transfer Kinetics Using the Concept of Extents. *Industrial & Engineering Chemistry Research* **2011**, *50*, 12960–12974.

(22) Srinivasan, S.; Billeter, J.; D., B. Extent-based incremental identification of reaction systems using concentration and calorimetric measurements. *Chemical Engineering Journal* **2012**, *207-208*, 785–793.

(23) Billeter, J.; Srinivasan, S.; D., B. Extent-based Kinetic Identification using Spectroscopic Measurements and Multivariate Calibration. *Analytica Chimica Acta* **2013**, *767*, 21–34.

(24) Srinivasan, S.; Billeter, J.; D., B. Sequential Model Identification of Reaction Systems - The Missing Path between the Incremental and Simultaneous Approaches. *AIChE Journal* **2017**, submitted.

(25) Fumasoli, A. Nitrification of Urine as Pretreatment for Nutrient Recovery. Ph.D. thesis, ETH Zürich, 2016.

(26) Srinivasan, S.; Billeter, J.; Bonvin, D. Identification of Multiphase Reaction Systems with Instantaneous Equilibria. *Industrial & Engineering Chemistry Research* **2016**, *29*, 8034–8045.

(27) Alonso, C.; Zhu, X.; Suidan, M. T.; Kim, B. R.; Kim, B. J. Parameter estimation in biofilter systems. *Environmental Science & Technology* **2000**, *34*, 2318–2323.

(28) Zhou, Y. A. N.; Pijuan, M.; Zeng, R. J.; Yuan, Z. Free nitrous acid inhibition on nitrous oxide reduction by a denitrifying-enhanced biological phosphorus removal sludge. *Environmental Science & Technology* **2008**, *42*, 8260–8265.

(29) Ni, B. J.; Zeng, R. J.; Fang, F.; Xu, J.; Sheng, G. P.; Yu, H. Q. A novel approach to evaluate the production kinetics of extracellular polymeric substances (EPS) by activated sludge using weighted nonlinear least-squares analysis. *Environmental Science & Technology* **2009**, *43*, 3743–3750.

(30) Sathyamoorthy, S.; Chandran, K.; Ramsburg, C. A. Biodegradation and cometabolic modeling of selected beta blockers during ammonia oxidation. *Environmental Science & Technology* **2013**, *47*, 12835–12843.

(31) Nelder, J. A.; Mead, R. A simplex method for function minimization. *The Computer Journal* **1965**, *7*, 308–313.

(32) Chandran, K.; Smets, B. F. Estimating biomass yield coefficients for autotrophic am-

monia and nitrite oxidation from batch respirograms. *Water Research* **2001**, *35*, 3153–3156.

(33) Buendía, I. M.; Fernández, F. J.; Villaseñor, J.; Rodríguez, L. Feasibility of anaerobic co-digestion as a treatment option of meat industry wastes. *Bioresource Technology* **2009**, *100*, 1903–1909.

(34) Gernaey, K.; Bogaert, H.; Massone, A.; Vanrolleghem, P.; Verstraete, W. On-line nitrification monitoring in activated sludge with a titrimetric sensor. *Environmental Science & Technology* **1997**, *31*, 2350–2355.

(35) Jensen, P. D.; Ge, H.; Batstone, D. J. Assessing the role of biochemical methane potential tests in determining anaerobic degradability rate and extent. *Water Science & Technology* **2011**, *64*, 880–886.

(36) Santa Cruz, J. A.; Mussati, S. F.; Scenna, N. J.; Gernaey, K. V.; Mussati, M. C. Reaction invariant-based reduction of the activated sludge model ASM1 for batch applications. *Journal of Environmental Chemical Engineering* **2016**, *4*, 3654–3664.

(37) Rodrigues, D.; Srinivasan, S.; Billeter, J.; D., B. Variant and Invariant States for Chemical Reaction Systems. *Computers & Chemical Engineering* **2015**, *73*, 23–33.

(38) Bonvin, D.; Rippin, D. W. T. Target factor analysis for the identification of stoichiometric models. *Chemical Engineering Science* **1990**, *45*, 3417–3426.

(39) Mašić, A.; ; Billeter, J.; Bonvin, D.; Villez, K. Extent computation under rank-deficient conditions. *20th World Congress of the International Federation of Automatic Control (IFAC2017), 9-14 July 2017, Toulouse, France* **2017**, *Accepted for oral presentation*.

(40) Rieger, L.; Alex, J.; Winkler, S.; Boehler, M.; Thomann, M.; Siegrist, H. Progress in sensor technology – progress in process control? Part I: Sensor property investigation and classification. *Water Science & Technology* **2003**, *47*, 103–112.

(41) Rosén, C.; Rieger, L.; Jeppsson, U.; Vanrolleghem, P. A. Adding realism to simulated sensors and actuators. *Water Science & Technology* **2008**, *57*, 337–344.

(42) Schielke-Jenni, S.; Villez, K.; Morgenroth, E.; Udert, K. M. Observability of anammox activity in single-stage nitritation/anammox reactors using mass balances. *Environmental Science: Water Research & Technology* **2015**, *1*, 523–534.

(43) Mašić, A.; Srinivasan, S.; Billeter, J.; Bonvin, D.; Villez, K. Shape Constrained Splines as Transparent Black-Box Models for Bioprocess Modeling. *Computers and Chemical Engineering* **2017**, *99*, 96–105.

(44) The MathWorks Inc., *MATLAB Release 2014b*; Natick, Massachusetts, 2014.

(45) Villez, K.; Rengaswamy, R.; Venkatasubramanian, V. Generalized Shape Constrained Spline Fitting for Qualitative Analysis of Trends. *Computers & Chemical Engineering* **2013**, *58*, 116–134.

Summary

The Supplementary Information consists of:

- This text which consists of 34 pages and includes 27 figures.
- The latest version of the EMI software package which enables reproduction of our results in the Matlab environment.

Software

All software necessary to reproduce the results presented in this work is available as part of the self-sufficient Efficient Model Identification (EMI) package for Matlab or Octave. It is published under the GPL v3 open-source license and constitutes the Supporting Information together with this text.

Graphical overview of modeling via extents

The modeling procedure is illustrated in Fig. S.1 for the exemplary case studied in this work. The three main steps, i.e. *(i)* extent computation, *(ii)* extent modeling, and *(iii)* model fine-tuning, are shown from top to bottom. The experimental extents are split into two individual time series corresponding to the two reactions. After this, the parameters for four candidate rate laws are estimated for each reaction. The best-fit rate laws are combined into a joint model. The associated parameter values are used as an initial guess for the fine-tuning step.

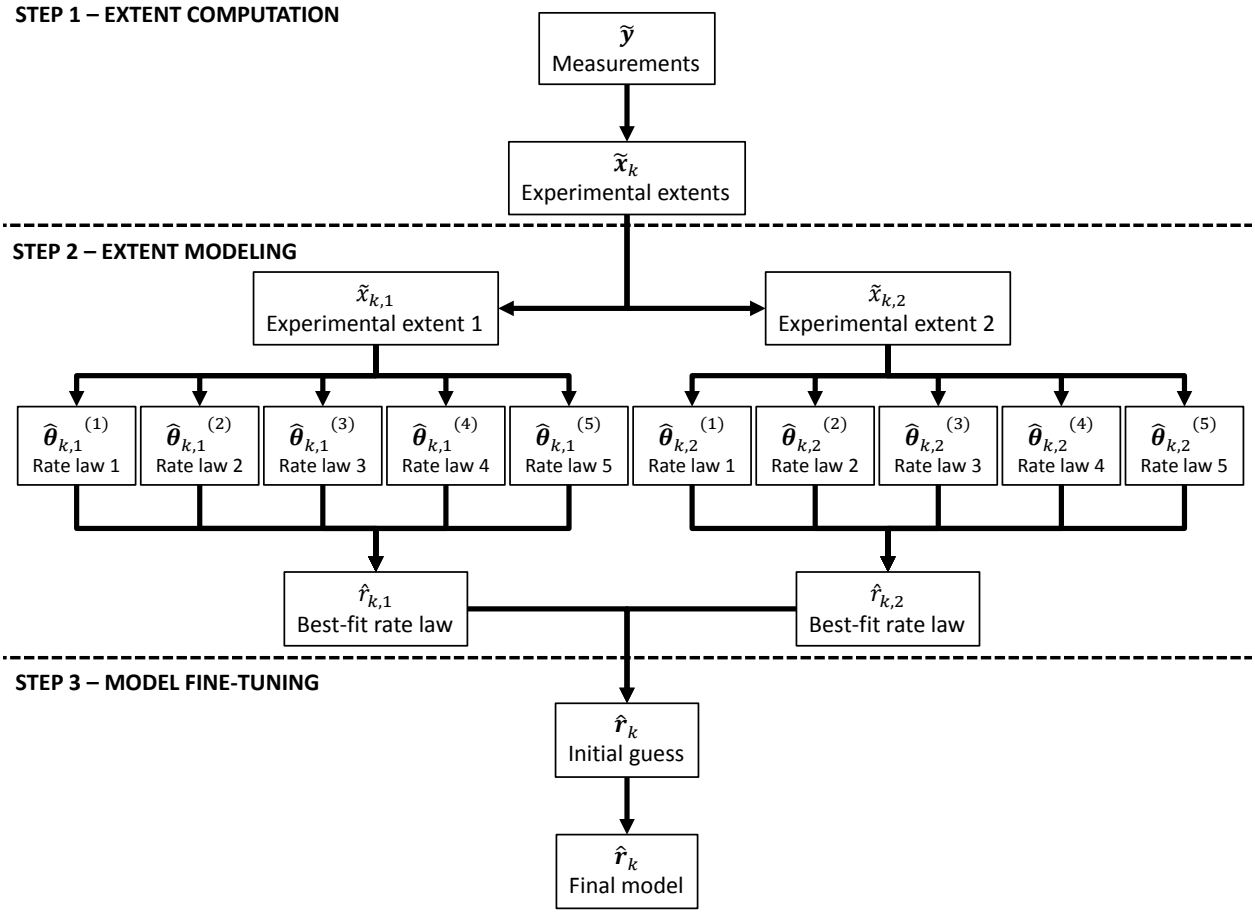

Figure S.1: **Extent-based modeling procedure.** Through the computation of experimental extents, kinetic modeling can be divided in smaller problems, each one focusing on the identification of the rate law and the corresponding parameters for a single reaction. A fine-tuning step is used at the end to obtain the final parameter estimates for the identified rate laws.

Bounding procedures

The estimation of the kinetic parameters in step 2 of the incremental model identification procedure is based on the branch-and-bound algorithm. Its use and application for biokinetic model parameter estimation has been demonstrated before[20]. The following bounding procedures constitute the only differences with this prior work. In what follows, we consider the estimation of a single parameter vector, $\boldsymbol{\theta}^{(j)}$, for a single candidate reaction rate law, $r_{k,i}^{(j)}$. For the sake of conciseness, these are given as $\boldsymbol{\theta}$ and r in what follows.

Definition of considered parameter set and parameter subsets

During the branch-and-bound algorithm, several hyper-rectangular parameter subsets are considered. These subsets are denoted here as Ω_a with a an integer indicating the chronology of the evaluated parameter subsets. Ω_0 corresponds to the root set, i.e. the set containing all feasible parameter values. Each parameter subset can be described as follows:

$$\boldsymbol{\theta} \in \Omega_a \Leftrightarrow \boldsymbol{\theta}_a^{\mathrm{L}} \leq \boldsymbol{\theta} \leq \boldsymbol{\theta}_a^{\mathrm{U}} \tag{38}$$

with $\boldsymbol{\theta}_a^{\mathrm{L}}$ and $\boldsymbol{\theta}_a^{\mathrm{U}}$ containing the lower and upper bounds for each element of $\boldsymbol{\theta}$. Inequalities between vectors are defined in an element-wise manner.

Upper bound

An upper bound to the objective function value is easily obtained by evaluating the objective function in (32) at an arbitrary value for $\boldsymbol{\theta}$ within the considered set Ω_a:

$$Q^{\mathrm{U}} = q(\boldsymbol{\theta}) = \sum_{h=1}^{H} \frac{\left(\tilde{x}_{k,i}(t_h) - x_{k,i}(t_h)\right)^2}{\lambda_{i,h}}, \tag{39}$$

following simulation of the following DAE system:

$$\mathbf{g}\left(\mathbf{n}(t)/V\right) = \mathbf{0} \tag{40}$$

$$\bar{\mathbf{E}}\,\mathbf{n}(t) = \bar{\mathbf{n}}_{k,0} + \bar{\mathbf{N}}^{\mathrm{T}}\,\mathbf{x}_k(t) \tag{41}$$

$$\forall r = 1,\ldots,R_k:\; x_{k,r}(t) = \begin{cases} V\int_0^t r_{k,i}\left(\mathbf{n}(\tau)/V,\boldsymbol{\theta}\right)\,d\tau, \quad x_{k,i}(0)=0 & \text{if } r = i \\ \mathcal{I}(\mathbf{t},\tilde{\mathbf{x}}_{k,r},t), & \text{if } r \neq i \end{cases} \tag{42}$$

with definitions as in the main text. It is fairly trivial to see that Q^{U} is a valid upper bound. Indeed, at least one set of parameter values within Ω_a results in an objective value that is lower or equal to Q^{U}. This is true since the evaluated $\boldsymbol{\theta}$ is in the set and gives $q(\boldsymbol{\theta}) = Q^{\mathrm{U}}$.

Lower bound

As usual, obtaining a provable lower bound is more challenging. In this study, we follow the previously developed procedure.[20] The main difference is that there is no need *(i)* to apply a linearizing model reformulation or *(ii)* to bound the value of rate measurements. This is because *(i)* the initial conditions and the stoichiometric matrix are considered known at the stage of kinetic parameter estimation and *(ii)* extents are integral states. As a result, the bounding procedure remains fairly simple.

To start, consider that the reaction rate can be bounded as follows for each of the candidate rate laws (see main text, Table 2):

$$\boldsymbol{\theta} \in \Omega_a: \quad 0 \leq r\left(\mathbf{n}(\tau)/V,\boldsymbol{\theta}_a^{\mathrm{U}}\right) \leq r\left(\mathbf{n}(\tau)/V,\boldsymbol{\theta}\right) \leq r\left(\mathbf{n}(\tau)/V,\boldsymbol{\theta}_a^{\mathrm{L}}\right) \tag{43}$$

Indeed, thanks to the particular parameterization in Table 2, one can easily see that the highest (lowest) reaction rates are obtained for the lowest (highest) parameter values within

Ω_a, i.e. $\boldsymbol{\theta}_a^{\mathrm{L}}$ ($\boldsymbol{\theta}_a^{\mathrm{U}}$). In addition, the reaction rate is strictly non-negative at all times (i.e. irreversible reaction). Combining this positivity of the reaction rate with the bounds for the reaction rates means that one can write the following inequality for ith modeled extent of reaction:

$$\boldsymbol{\theta} \in \Omega_a : \quad x_{k,i}^{\mathrm{L}}(t) \leq x_{k,i}(t) \leq x_{k,i}^{\mathrm{U}}(t) \tag{44}$$

with

$$x_{k,i}^{\mathrm{L}}(t) = V \int_0^t r\left(\mathbf{n}(\tau)/V, \boldsymbol{\theta}_a^{\mathrm{U}}\right)\, d\tau, \quad x_{k,i}^{\mathrm{L}}(0) = 0 \tag{45}$$

$$x_{k,i}^{\mathrm{U}}(t) = V \int_0^t r\left(\mathbf{n}(\tau)/V, \boldsymbol{\theta}_a^{\mathrm{L}}\right)\, d\tau, \quad x_{k,i}^{\mathrm{U}}(0) = 0 \tag{46}$$

subject to (40–41) and all evaluations of the $r \neq i$ case in (42).

In words, the considered extent of reaction at time t is the highest (lowest) for the highest (lowest) reaction rates and thus the lowest (highest) parameter values. This statement follows from the fact that the extent is a monotonic function of time (positivity of the reaction rate) with the derivative defined by the reaction rate. This derivative takes its lowest (highest) attainable value for the highest (lowest) parameter values at any time t and for any possible state that has been reached at time t. It follows that two simulations delivering $x_{k,i}^{\mathrm{U}}(t)$ and $x_{k,i}^{\mathrm{L}}(t)$ deliver effective bounds to the extent profiles obtained with any feasible value for $\boldsymbol{\theta}$ within Ω_a.

Based on interval arithmetic, the squared residuals $s_{k,i}(t_h) = \left(\tilde{x}_{k,i}(t_h) - x_{k,i}(t_h)\right)^2$ can now be lower bounded as follows:

$$\boldsymbol{\theta} \in \Omega_a : \quad s^{\mathrm{L}}_{k,i}(t_h) \leq s_{k,i}(t_h)|_{\boldsymbol{\theta}} \tag{47}$$

with

$$d^{\mathrm{L}}_{k,i}(t_h) = \tilde{x}_{k,i}(t_h) - x^{\mathrm{U}}_{k,i}(t_h) \tag{48}$$

$$d^{\mathrm{U}}_{k,i}(t_h) = \tilde{x}_{k,i}(t_h) - x^{\mathrm{L}}_{k,i}(t_h) \tag{49}$$

$$s^{\mathrm{L}}_{k,i}(t_h) = \begin{cases} 0 & \text{if } d^{\mathrm{L}}_{k,i}(t_h) \leq 0 \leq d^{\mathrm{U}}_{k,i}(t_h) \\ \min\left(d^{\mathrm{L}}_{k,i}(t_h)^2, d^{\mathrm{U}}_{k,i}(t_h)^2\right) & \text{otherwise} \end{cases} \tag{50}$$

From this, it follows that:

$$\boldsymbol{\theta} \in \Omega_a : \quad Q^{\mathrm{L}} \leq q(\boldsymbol{\theta}) \tag{51}$$

with

$$Q^{\mathrm{L}} = \sum_{h=1}^{H} \frac{s^{\mathrm{L}}_{k,i}(t_h)}{\lambda_{i,h}}, \tag{52}$$

which proves that Q^{L} is a valid lower bound.

Implementation of the bounding procedures

The above procedures suggest simulation of the considered extent of reaction for three parameter vectors. The first is executed for an arbitrary feasible choice for $\boldsymbol{\theta}$ within Ω_a. The second and third simulation is executed for $\boldsymbol{\theta}^{\mathrm{L}}$ and $\boldsymbol{\theta}^{\mathrm{U}}$. These simulations are the computa-

tionally most expensive steps of the bounding procedures. For this reason, the upper bound procedure is evaluated for $\boldsymbol{\theta}^{\mathrm{L}}$ and $\boldsymbol{\theta}^{\mathrm{U}}$, since the corresponding extent simulations are required anyway for the lower bound. This means only two simulations are executed to compute both the lower and upper bound. In the process, one obtains two distinct upper bound values. The minimum of these two upper bounds is then reported as the best-known upper bound. A graphical scheme of the bounding procedures is given in Fig. S.2. Note that this scheme is fairly simple compared to the original bounding procedures.[20]

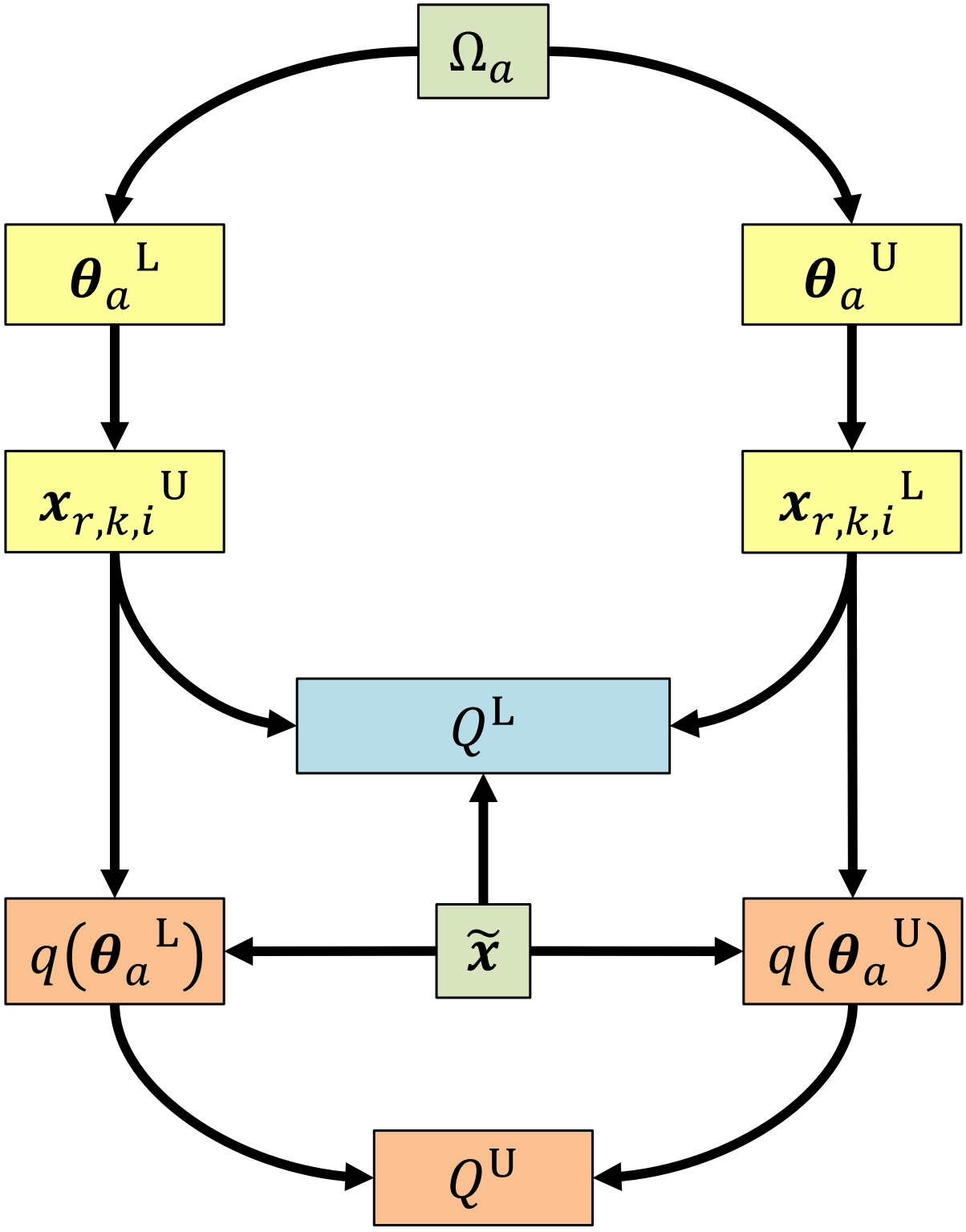

Figure S.2: **Illustration of the bounding procedures:** Two simulations are executed, one for both extremal parameter vectors ($\boldsymbol{\theta}_a^{\mathrm{L}}$, $\boldsymbol{\theta}_a^{\mathrm{U}}$) that bound the considered set (Ω_a). These deliver the bounding profiles for the extent of reaction ($\mathbf{x}_{k,i}^{\mathrm{L}}$ and $\mathbf{x}_{k,i}^{\mathrm{U}}$). By combining these two extremal profiles with the experimental extent series ($\tilde{\mathbf{x}}$), one can compute both the upper bound (Q^{U}) and the lower bound (Q^{L}). In this scheme, should $Q(\boldsymbol{\theta}_a^{\mathrm{L}})$ and $Q(\boldsymbol{\theta}_a^{\mathrm{U}})$ not be $q(\boldsymbol{\theta}_a^{\mathrm{L}})$ and $Q(\boldsymbol{\theta}_a^{\mathrm{U}})$? Also, the exponents L and U seem to be bold although they should not be...

Additional results

The next figures (Fig. S.3-S.27) show the simulation results obtained which each model obtained with Method 1 after parameter estimation.

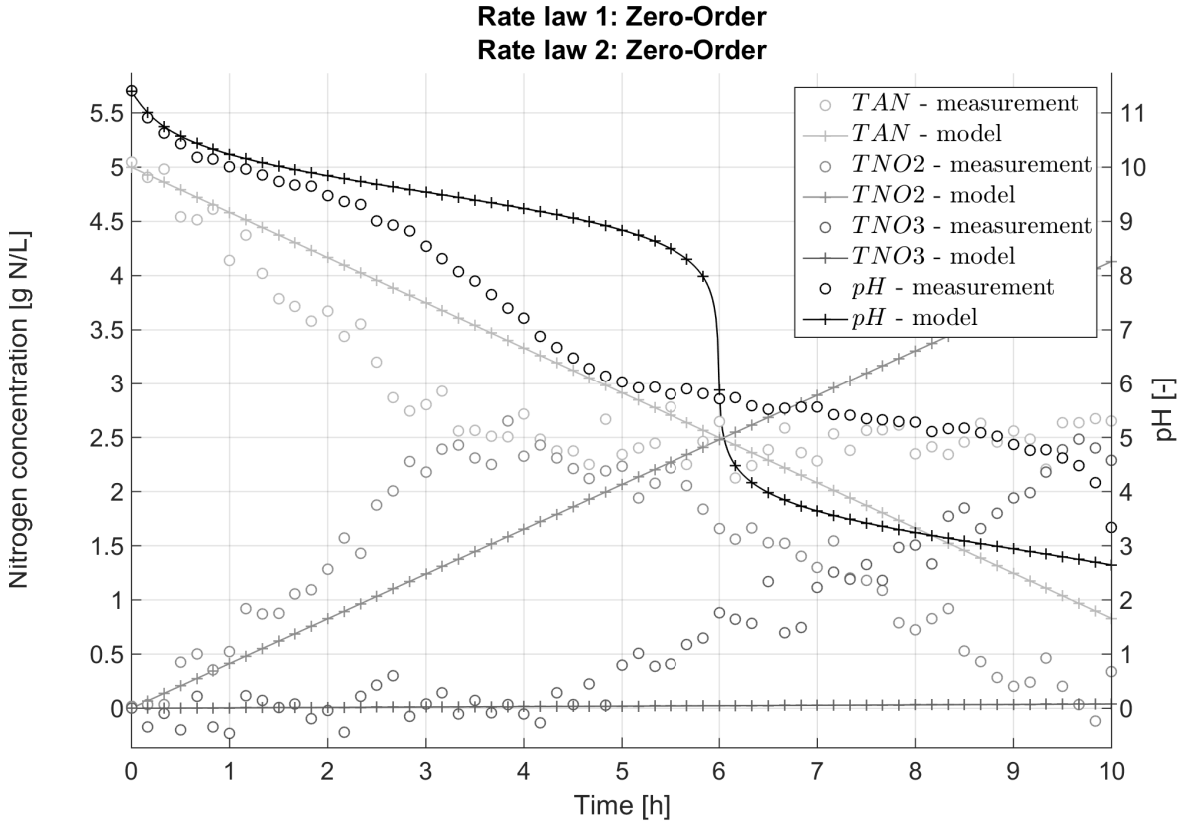

Figure S.3: **Method 1 - Simultaneous model identification - Model 1.** Measurements and simulation of the measured variables with model 1 after parameter estimation with the Nelder-Mead simplex method.

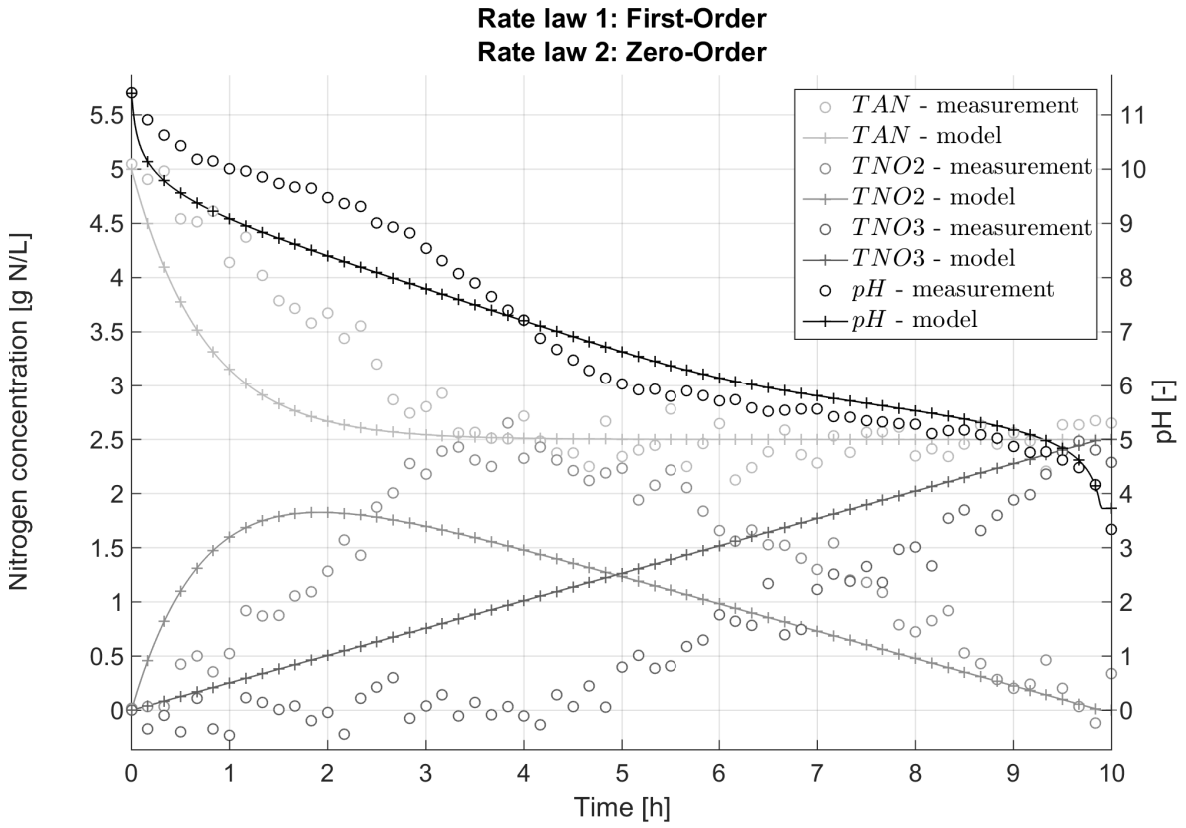

Figure S.4: **Method 1 - Simultaneous model identification - Model 2.** Measurements and simulation of the measured variables with model 2 after parameter estimation with the Nelder-Mead simplex method.

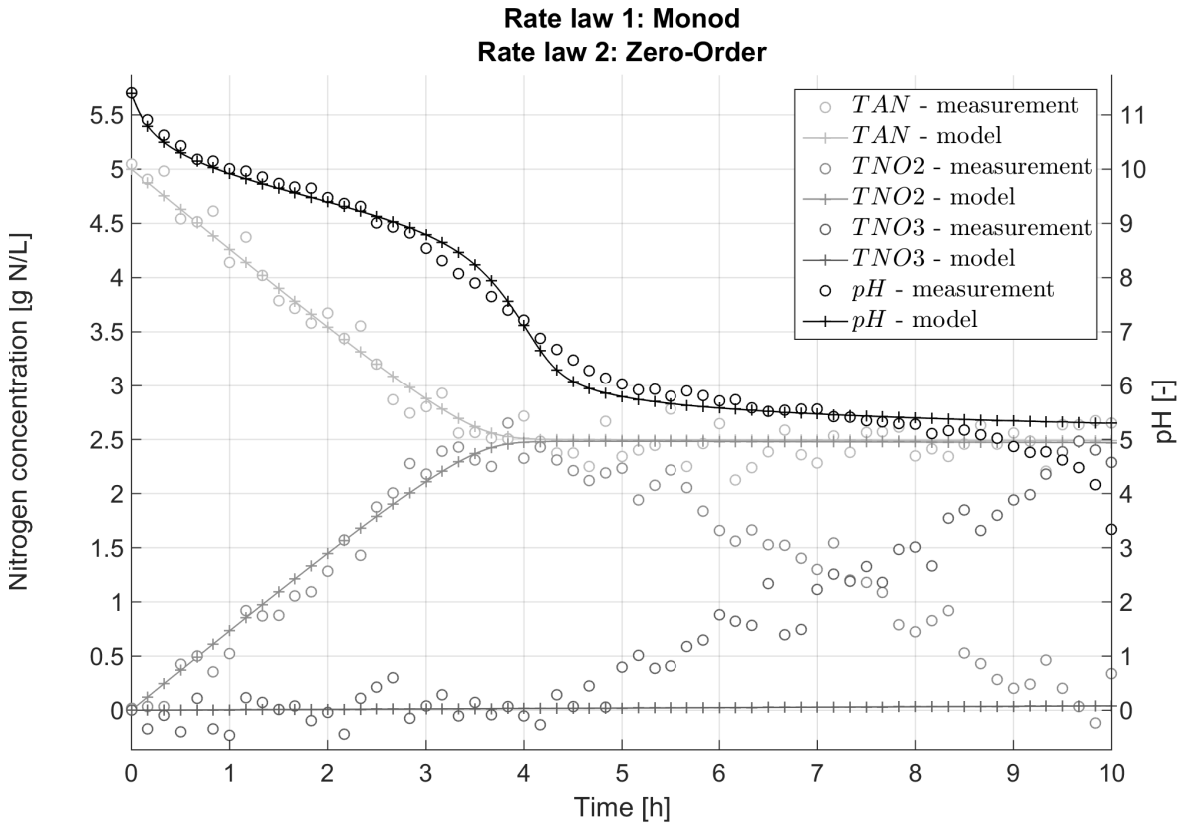

Figure S.5: **Method 1 - Simultaneous model identification - Model 3.** Measurements and simulation of the measured variables with model 3 after parameter estimation with the Nelder-Mead simplex method.

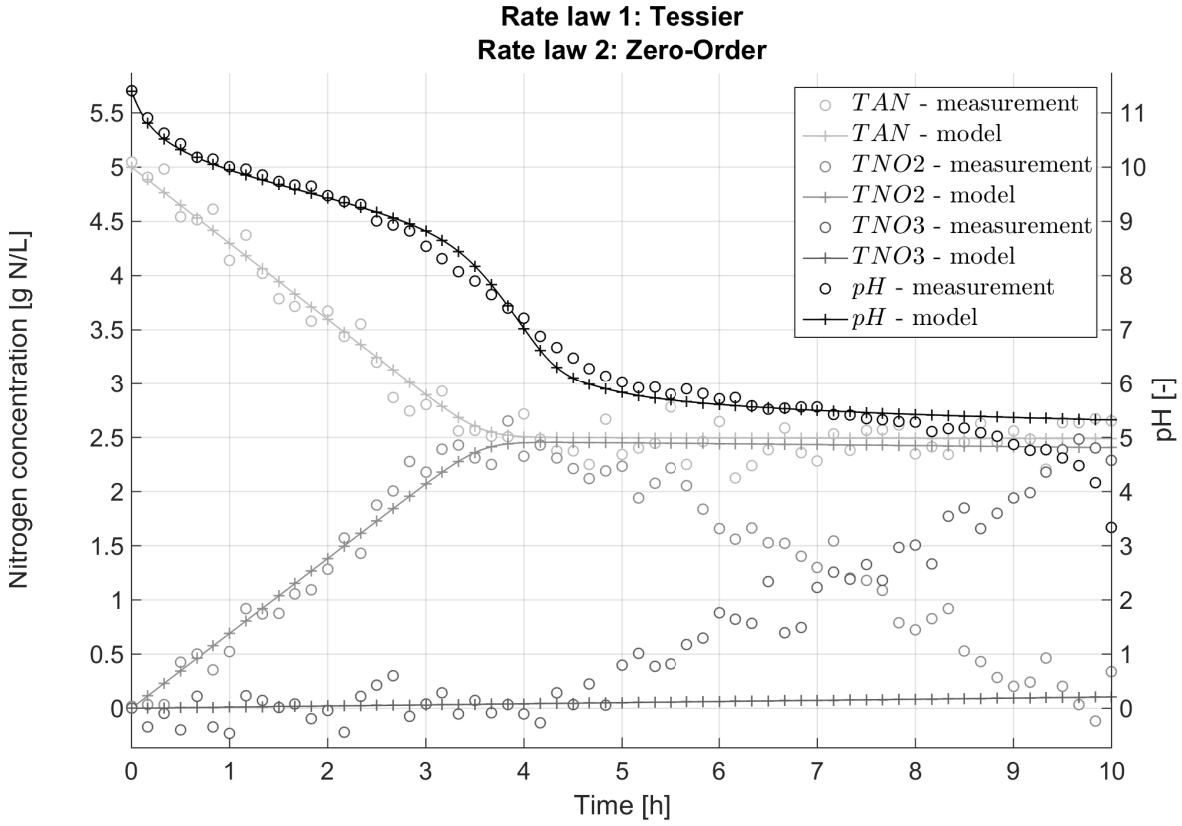

Figure S.6: **Method 1 - Simultaneous model identification - Model 4.** Measurements and simulation of the measured variables with model 4 after parameter estimation with the Nelder-Mead simplex method.

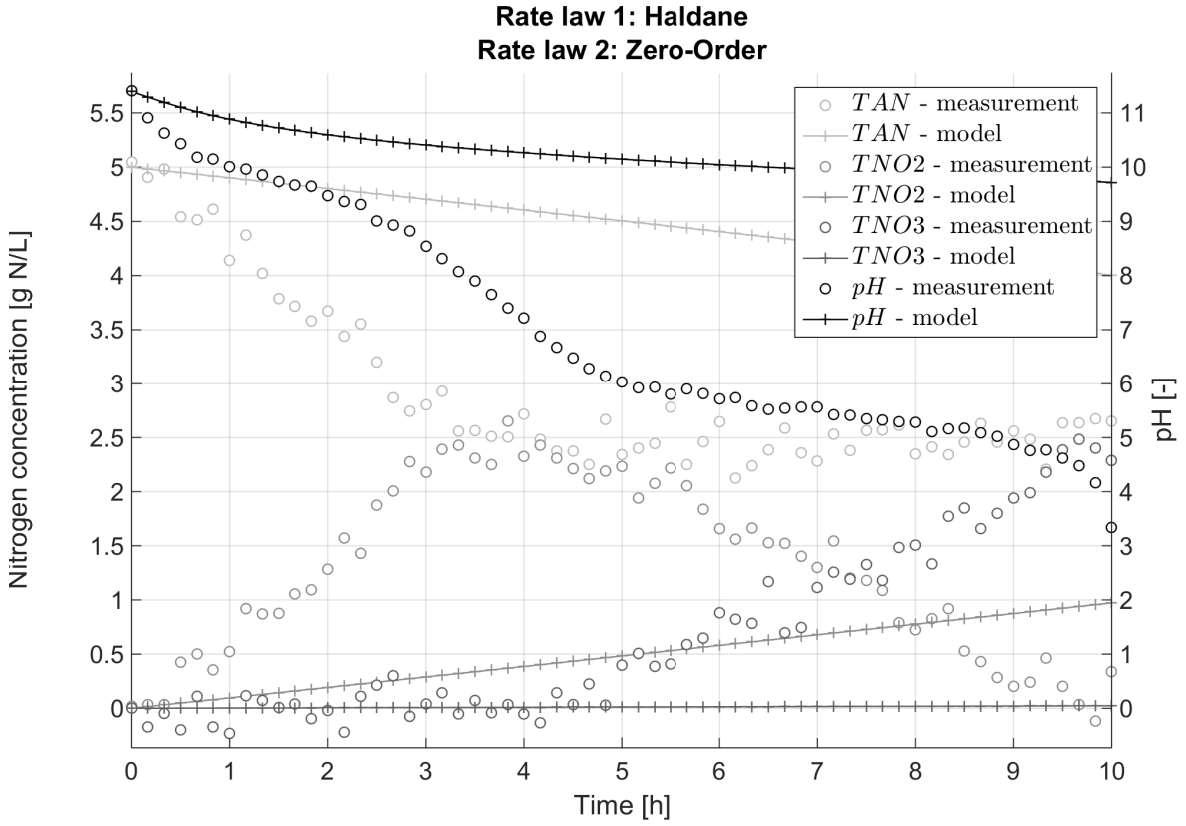

Figure S.7: **Method 1 - Simultaneous model identification - Model 5.** Measurements and simulation of the measured variables with model 5 after parameter estimation with the Nelder-Mead simplex method.

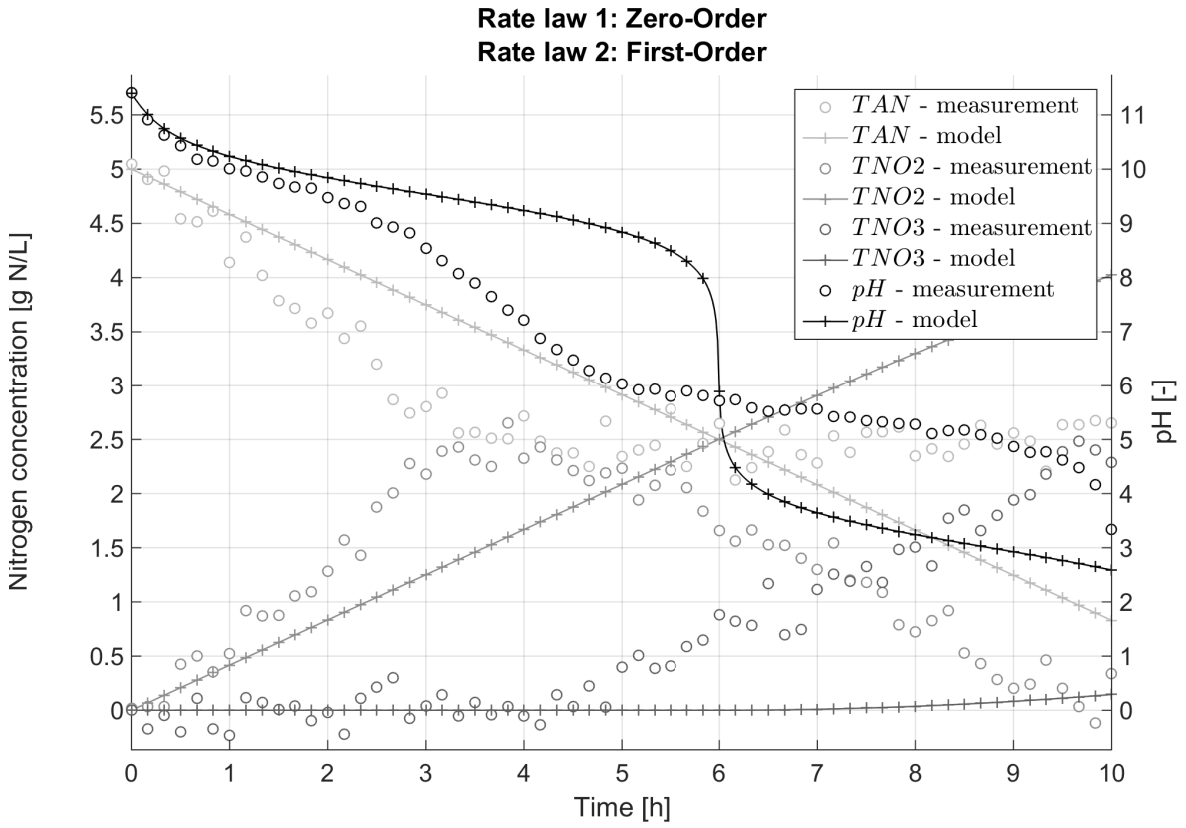

Figure S.8: **Method 1 - Simultaneous model identification - Model 6.** Measurements and simulation of the measured variables with model 6 after parameter estimation with the Nelder-Mead simplex method.

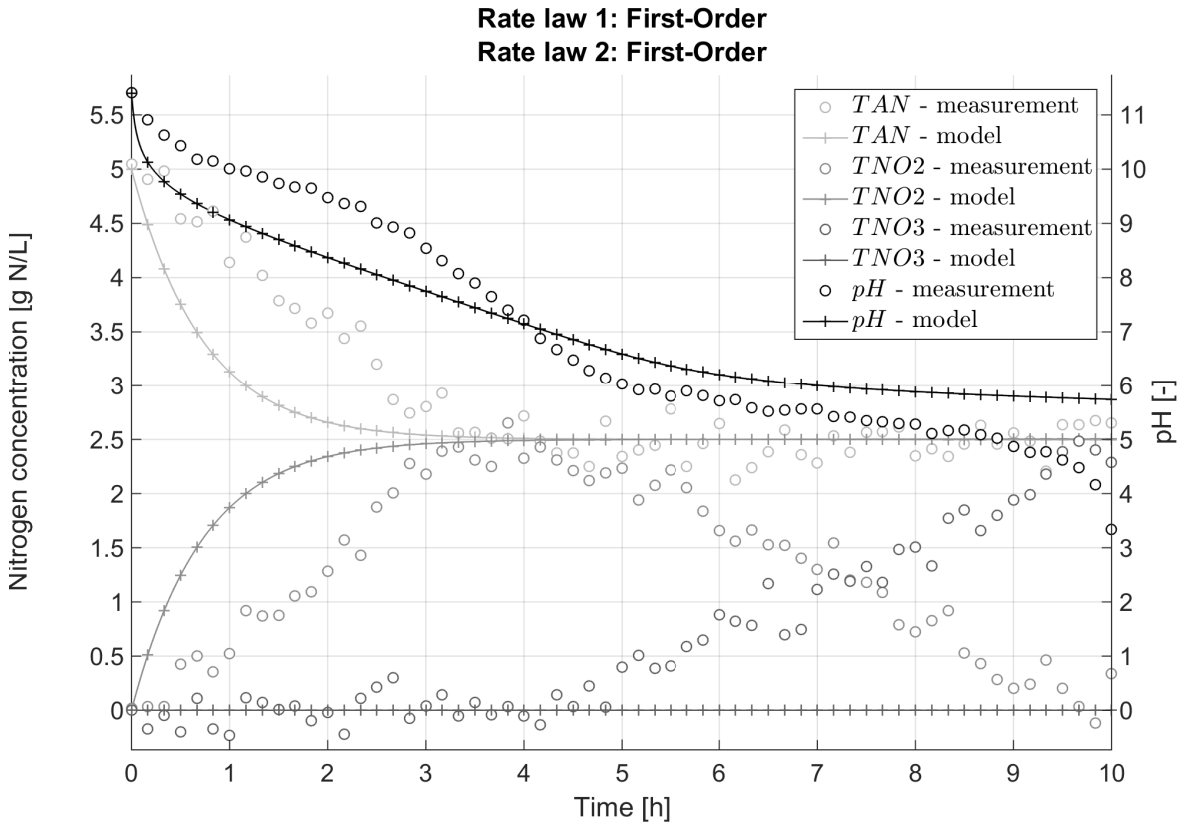

Figure S.9: **Method 1 - Simultaneous model identification - Model 7.** Measurements and simulation of the measured variables with model 7 after parameter estimation with the Nelder-Mead simplex method.

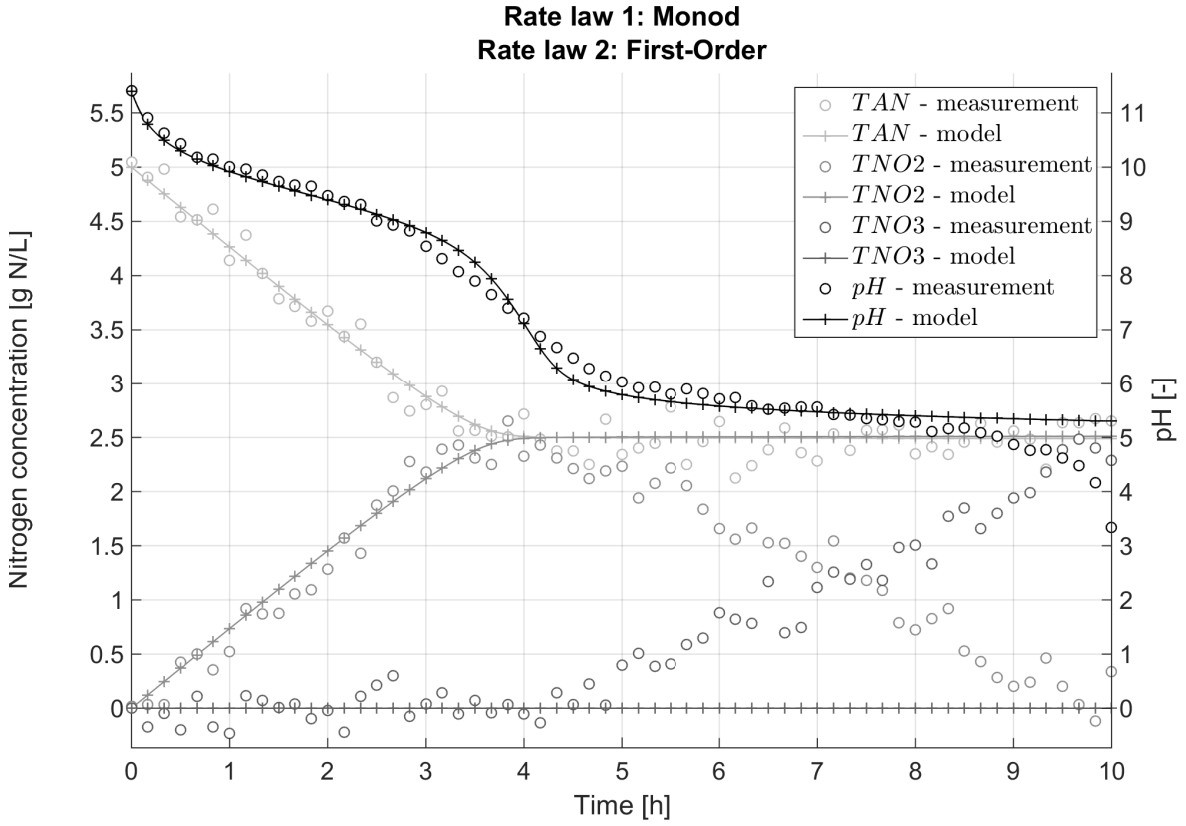

Figure S.10: **Method 1 - Simultaneous model identification - Model 8.** Measurements and simulation of the measured variables with model 8 after parameter estimation with the Nelder-Mead simplex method.

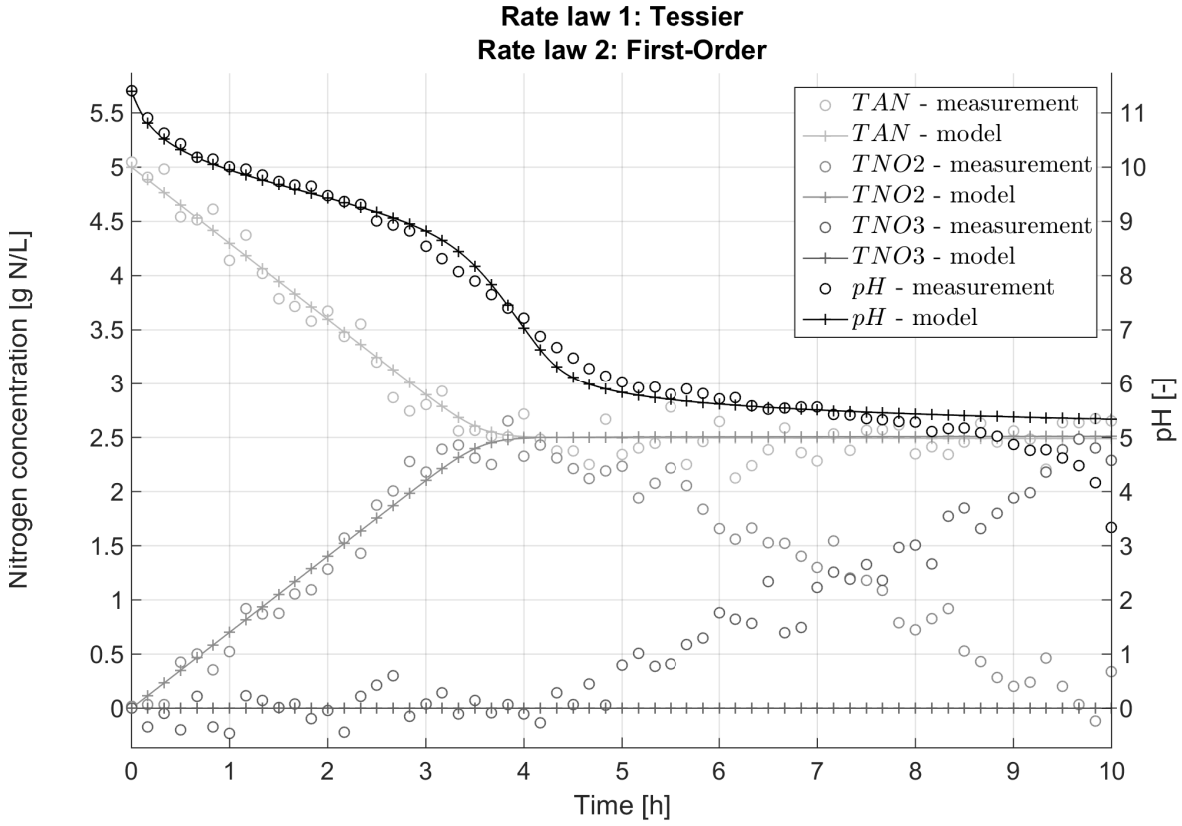

Figure S.11: **Method 1 - Simultaneous model identification - Model 9.** Measurements and simulation of the measured variables with model 9 after parameter estimation with the Nelder-Mead simplex method.

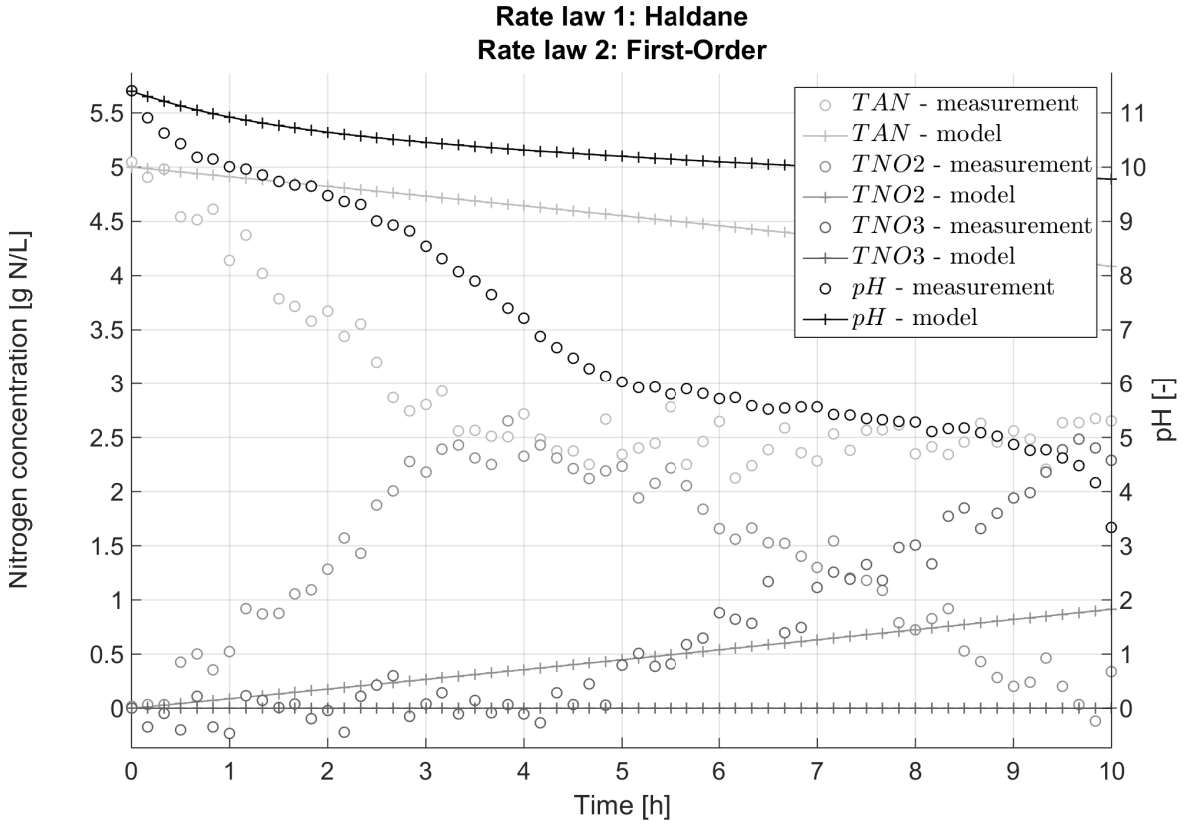

Figure S.12: **Method 1 - Simultaneous model identification - Model 10.** Measurements and simulation of the measured variables with model 10 after parameter estimation with the Nelder-Mead simplex method.

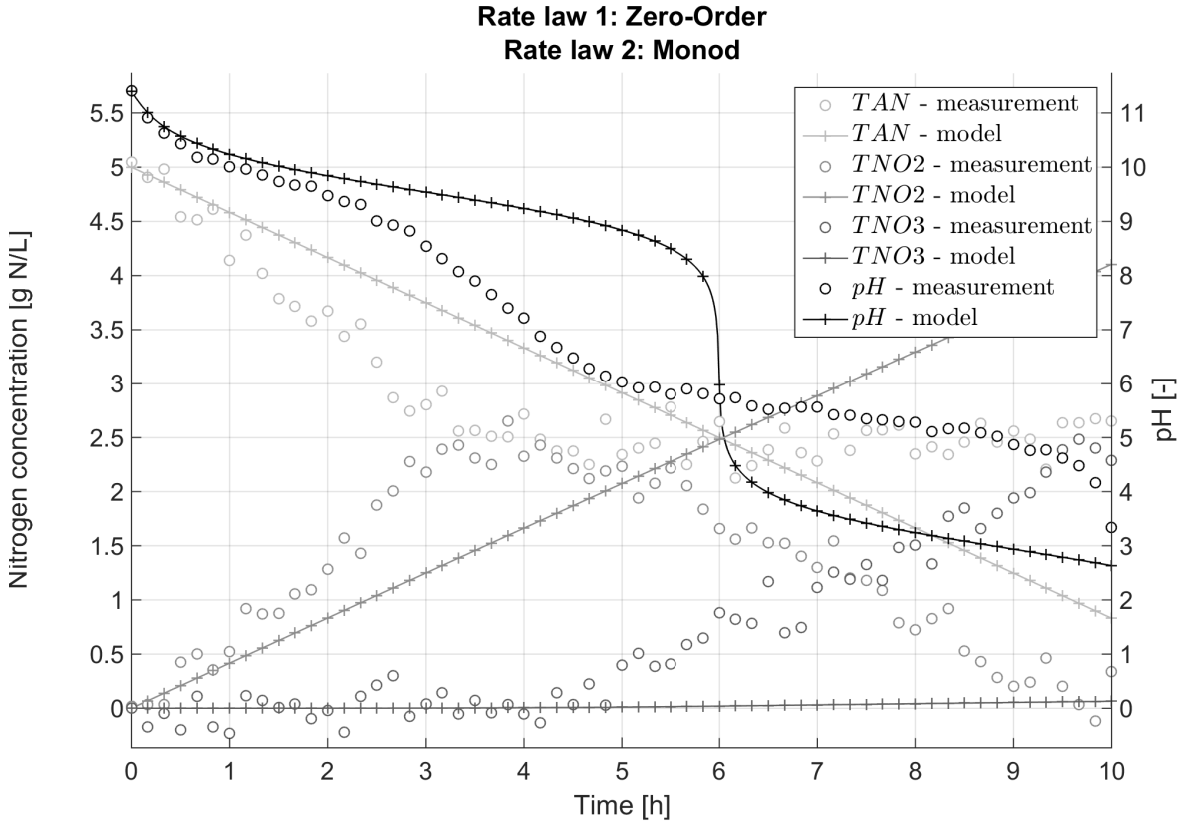

Figure S.13: **Method 1 - Simultaneous model identification - Model 11.** Measurements and simulation of the measured variables with model 11 after parameter estimation with the Nelder-Mead simplex method.

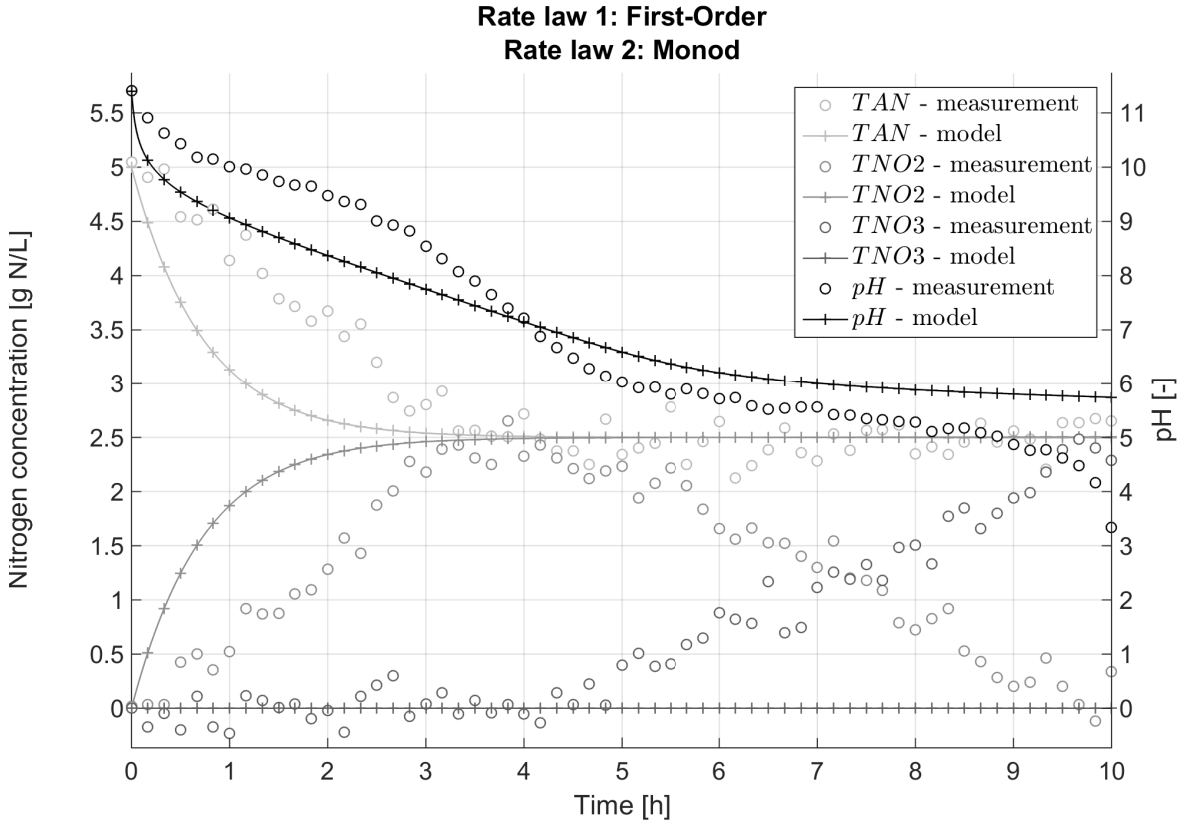

Figure S.14: **Method 1 - Simultaneous model identification - Model 12.** Measurements and simulation of the measured variables with model 12 after parameter estimation with the Nelder-Mead simplex method.

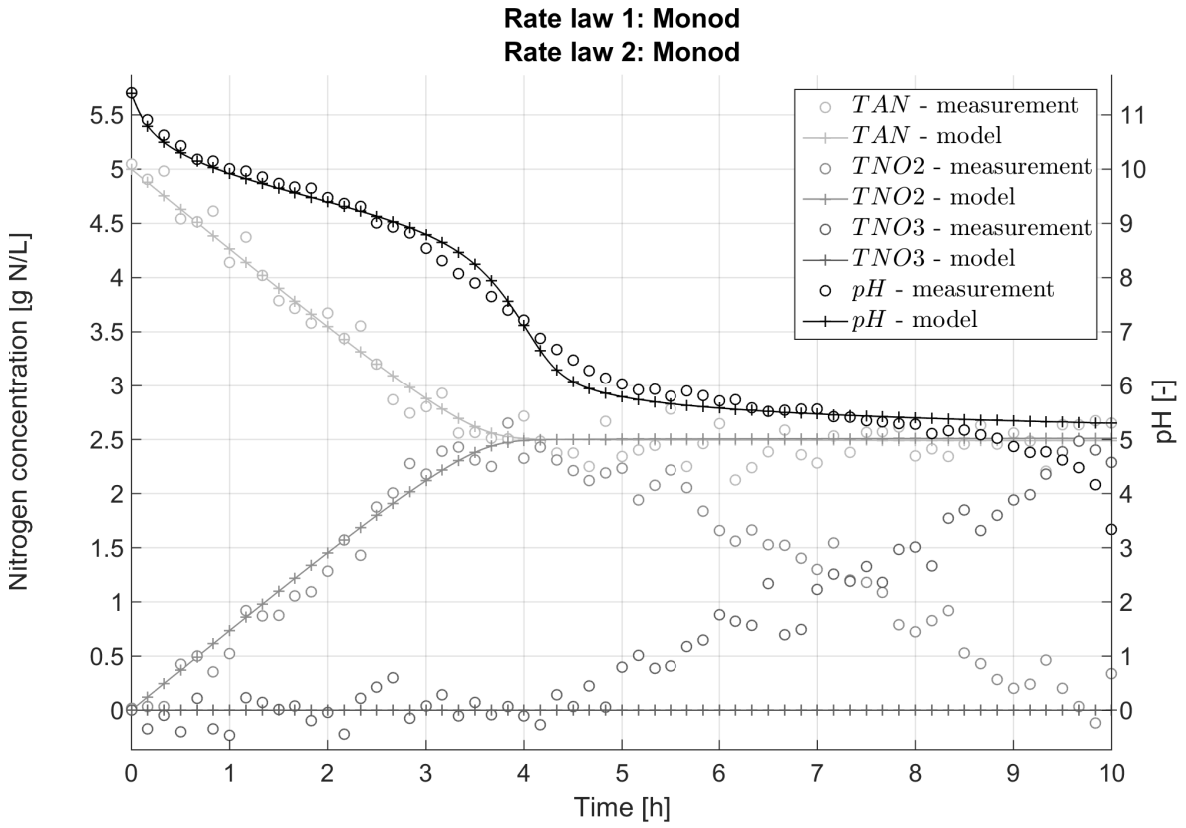

Figure S.15: **Method 1 - Simultaneous model identification - Model 13.** Measurements and simulation of the measured variables with model 13 after parameter estimation with the Nelder-Mead simplex method.

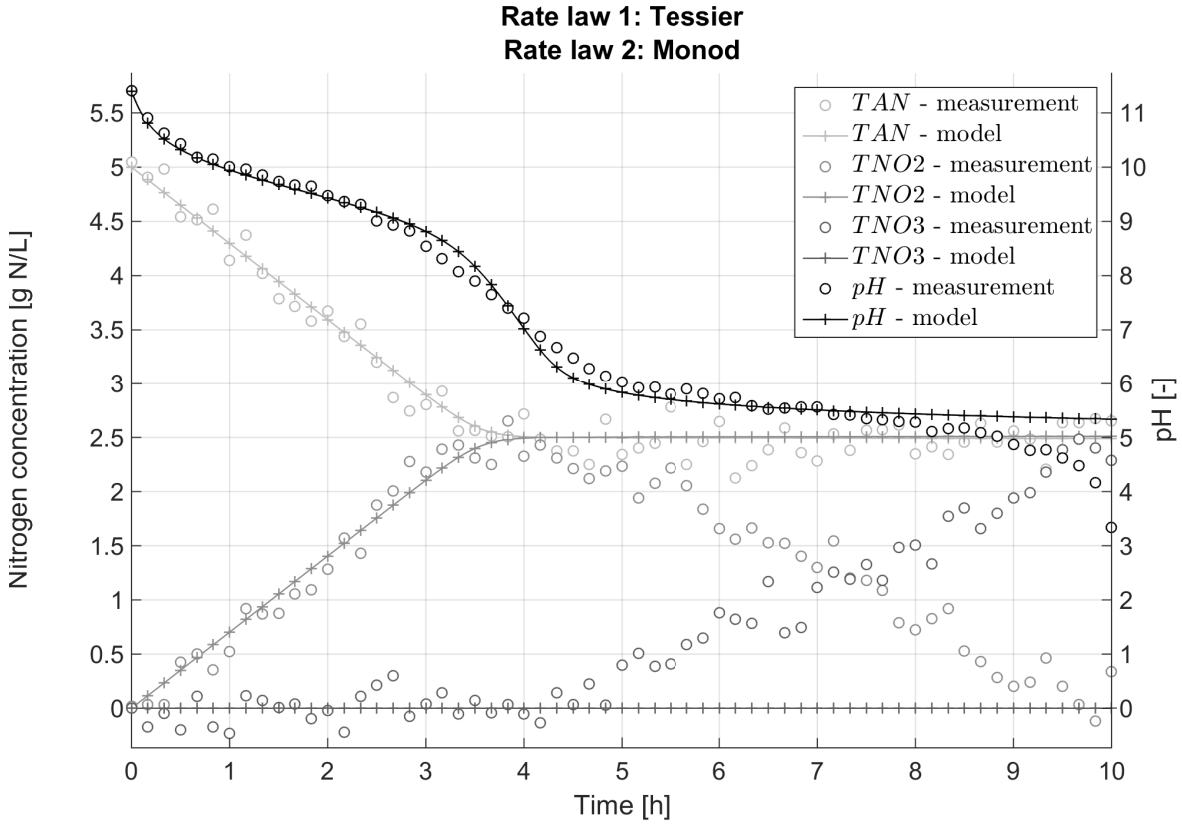

Figure S.16: **Method 1 - Simultaneous model identification - Model 14.** Measurements and simulation of the measured variables with model 14 after parameter estimation with the Nelder-Mead simplex method.

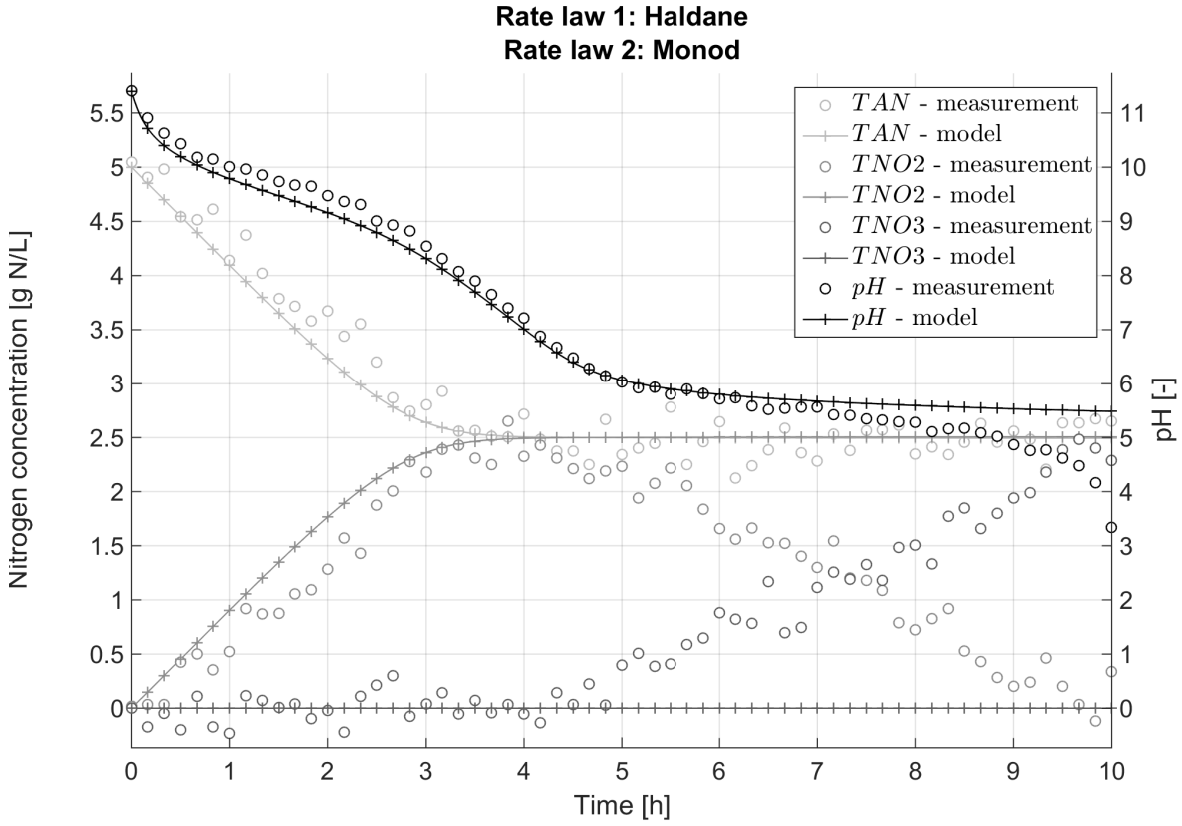

Figure S.17: **Method 1 - Simultaneous model identification - Model 15.** Measurements and simulation of the measured variables with model 15 after parameter estimation with the Nelder-Mead simplex method.

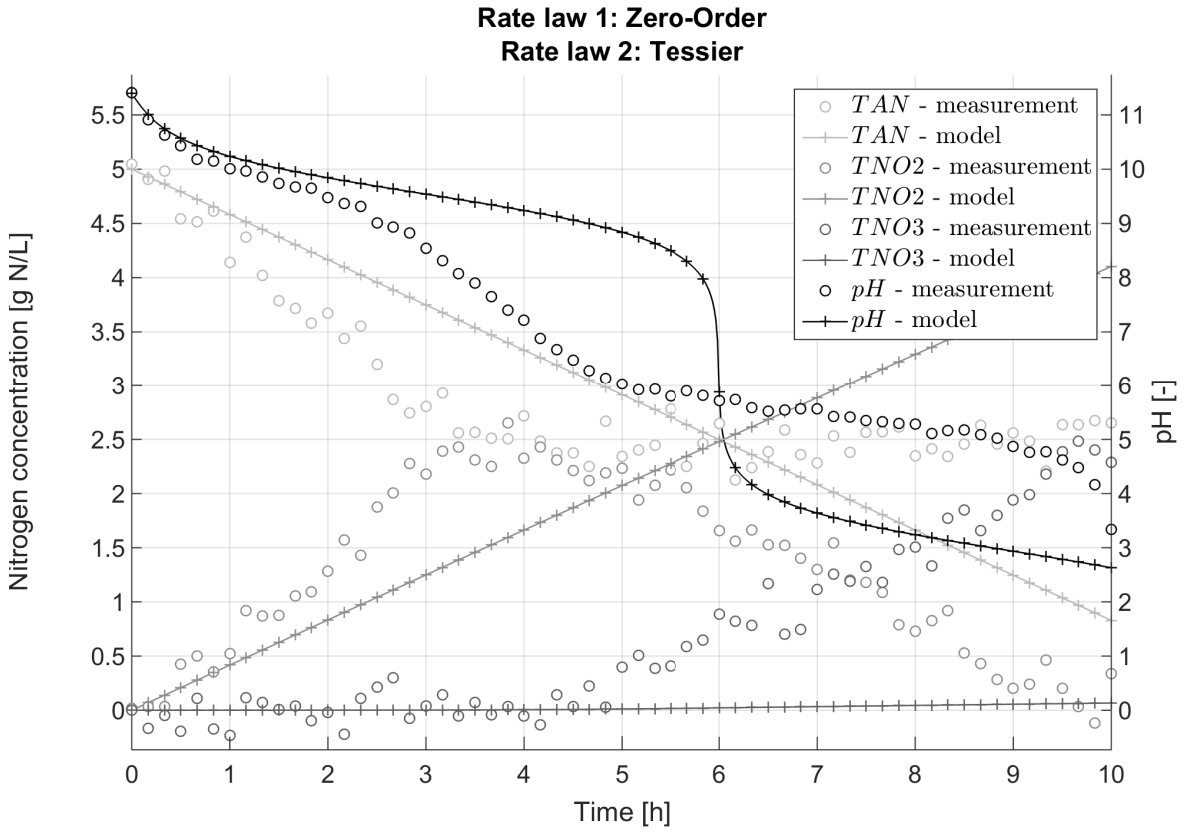

Figure S.18: **Method 1 - Simultaneous model identification - Model 16.** Measurements and simulation of the measured variables with model 16 after parameter estimation with the Nelder-Mead simplex method.

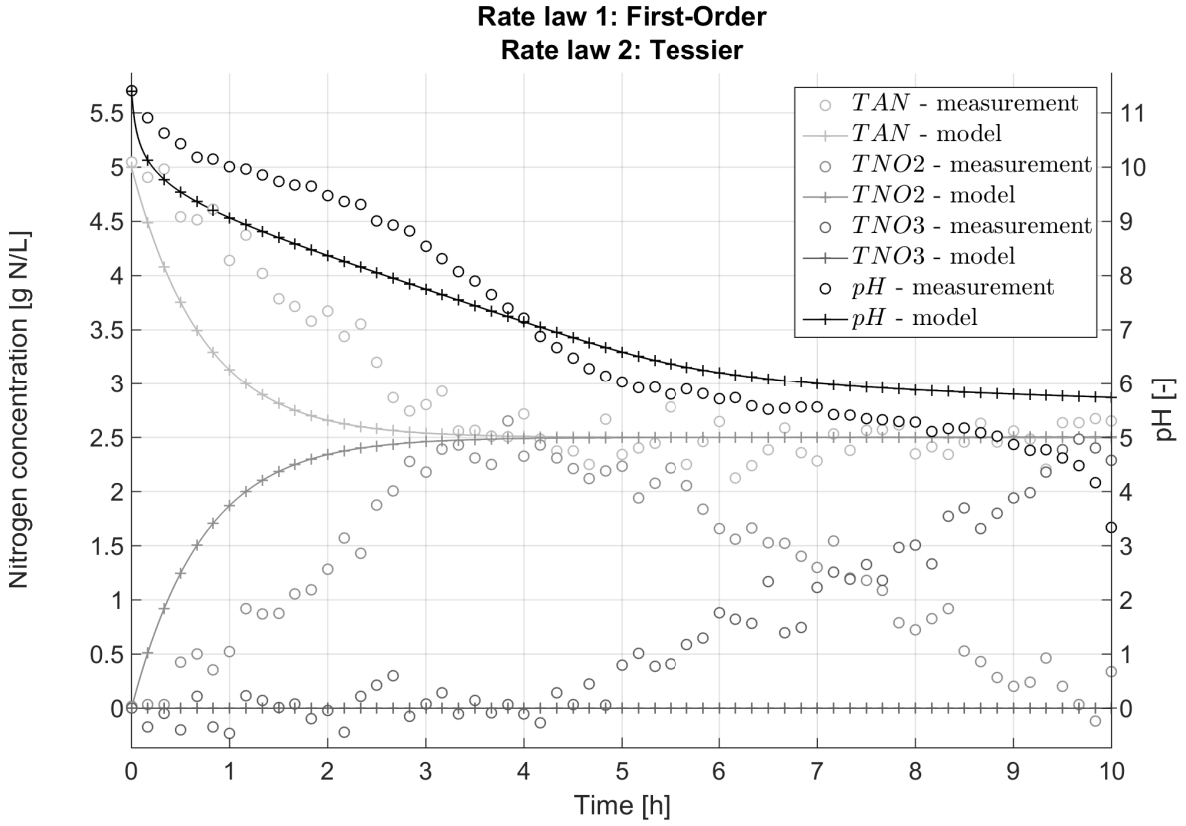

Figure S.19: **Method 1 - Simultaneous model identification - Model 17.** Measurements and simulation of the measured variables with model 17 after parameter estimation with the Nelder-Mead simplex method.

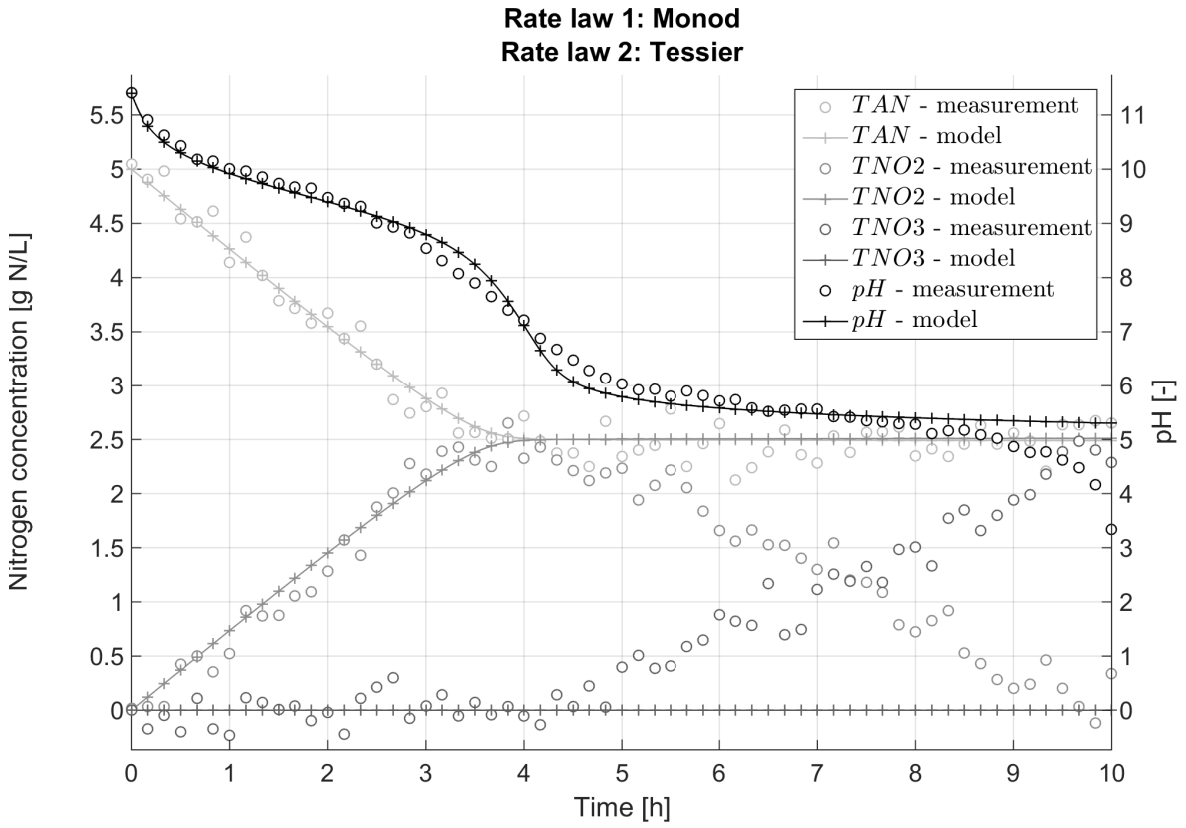

Figure S.20: **Method 1 - Simultaneous model identification - Model 18.** Measurements and simulation of the measured variables with model 18 after parameter estimation with the Nelder-Mead simplex method.

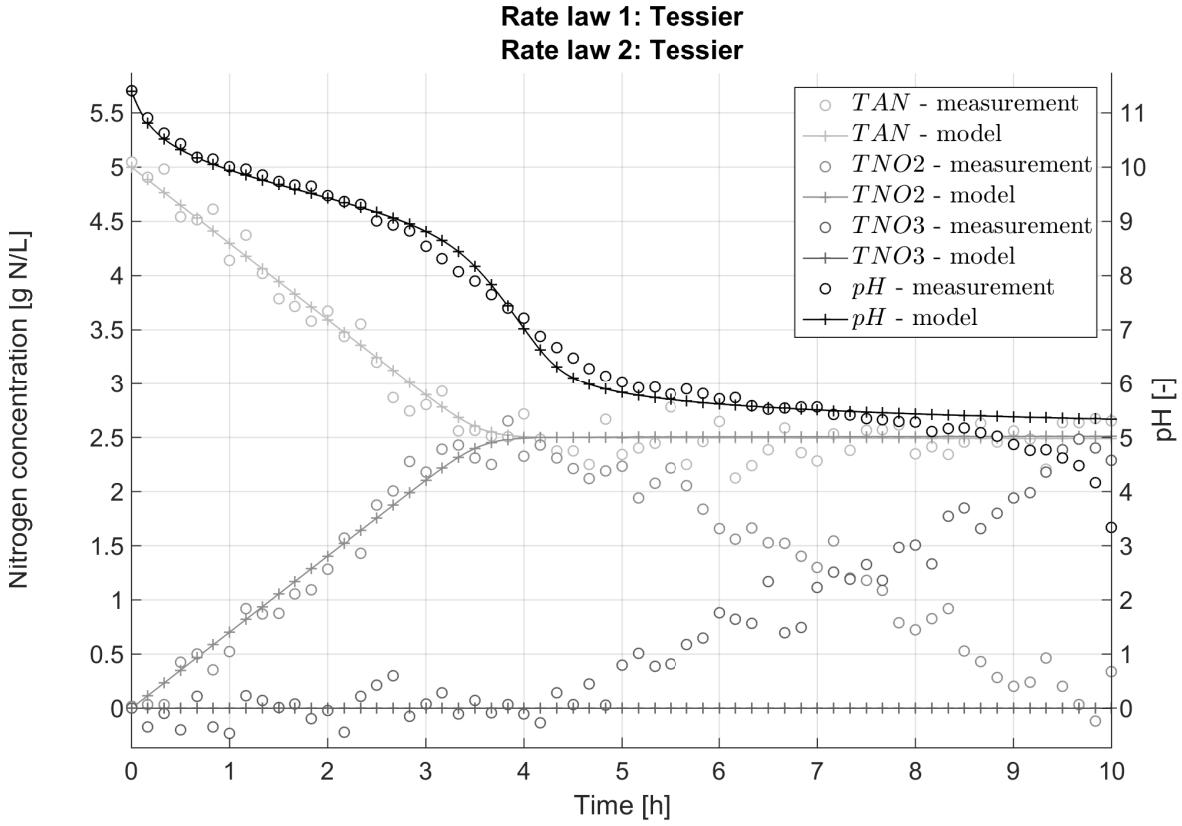

Figure S.21: **Method 1 - Simultaneous model identification - Model 19.** Measurements and simulation of the measured variables with model 19 after parameter estimation with the Nelder-Mead simplex method.

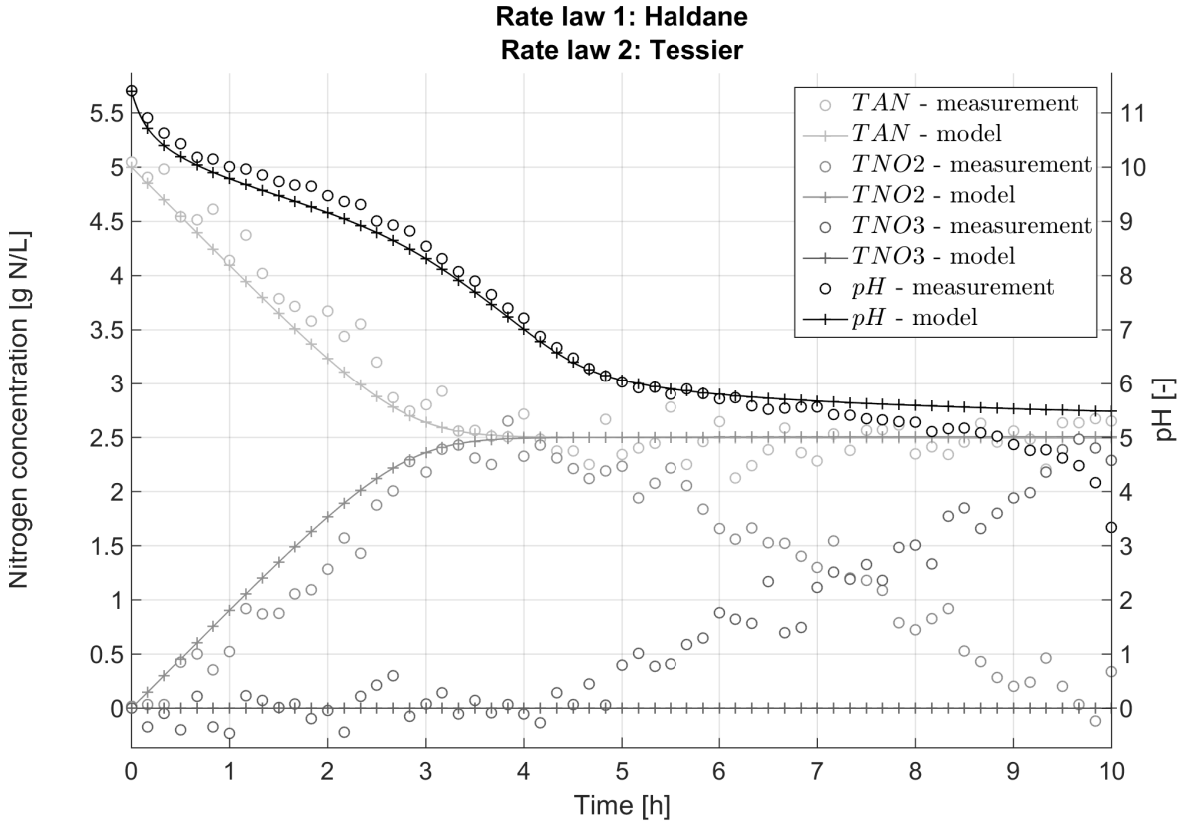

Figure S.22: **Method 1 - Simultaneous model identification - Model 20.** Measurements and simulation of the measured variables with model 20 after parameter estimation with the Nelder-Mead simplex method.

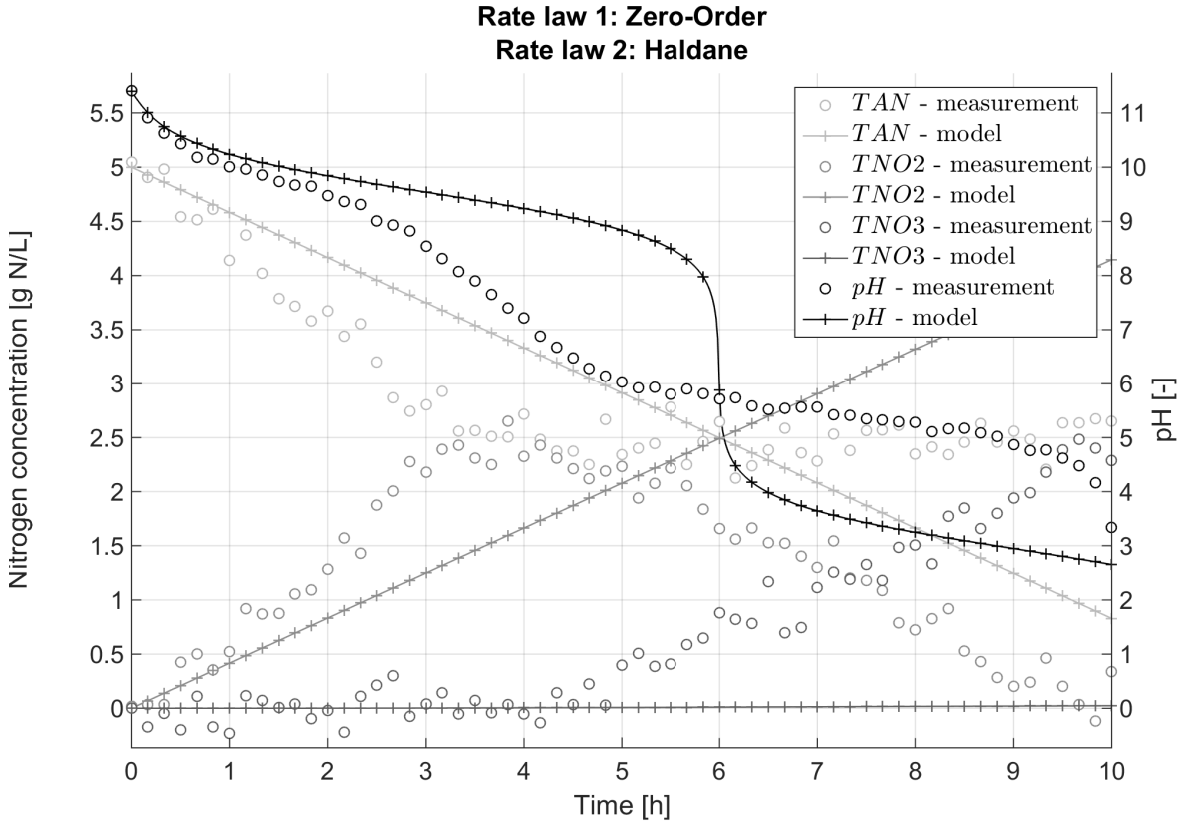

Figure S.23: **Method 1 - Simultaneous model identification - Model 21.** Measurements and simulation of the measured variables with model 21 after parameter estimation with the Nelder-Mead simplex method.

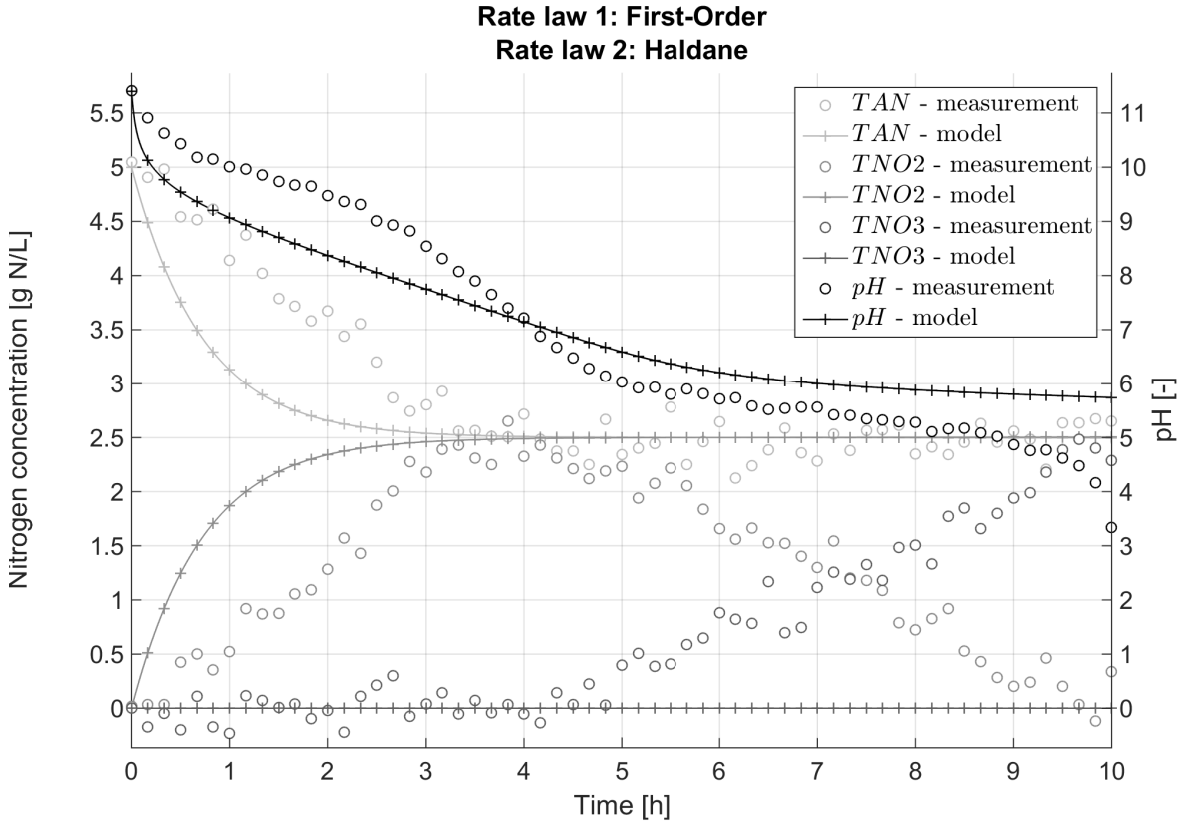

Figure S.24: **Method 1 - Simultaneous model identification - Model 22.** Measurements and simulation of the measured variables with model 22 after parameter estimation with the Nelder-Mead simplex method.

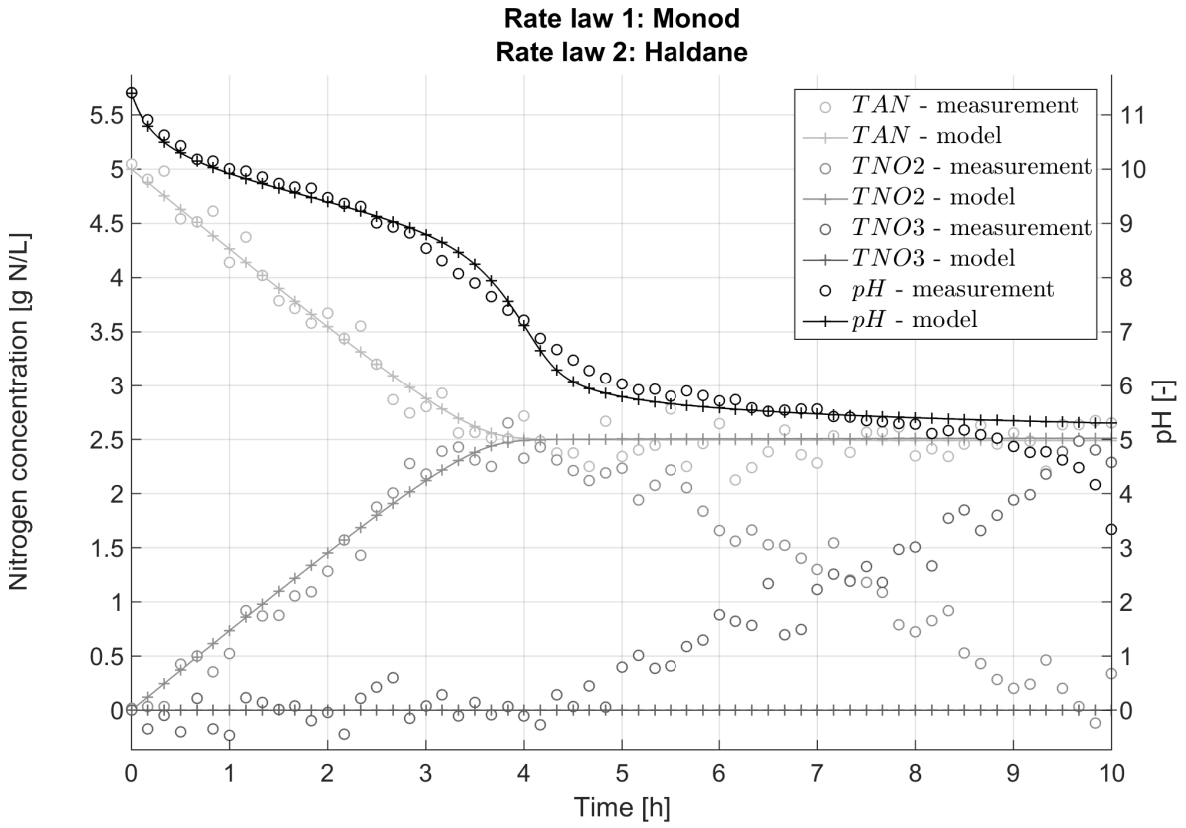

Figure S.25: **Method 1 - Simultaneous model identification - Model 23.** Measurements and simulation of the measured variables with model 23 after parameter estimation with the Nelder-Mead simplex method.

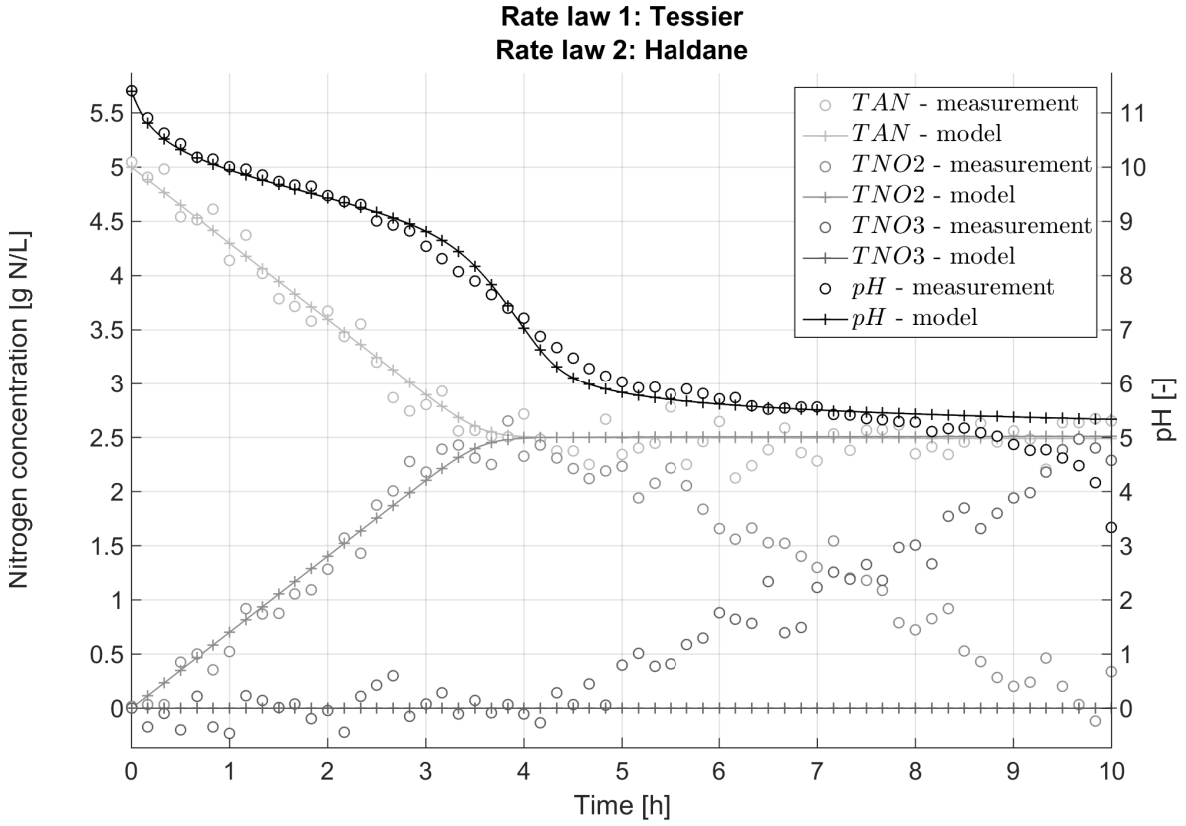

Figure S.26: **Method 1 - Simultaneous model identification - Model 24.** Measurements and simulation of the measured variables with model 24 after parameter estimation with the Nelder-Mead simplex method.

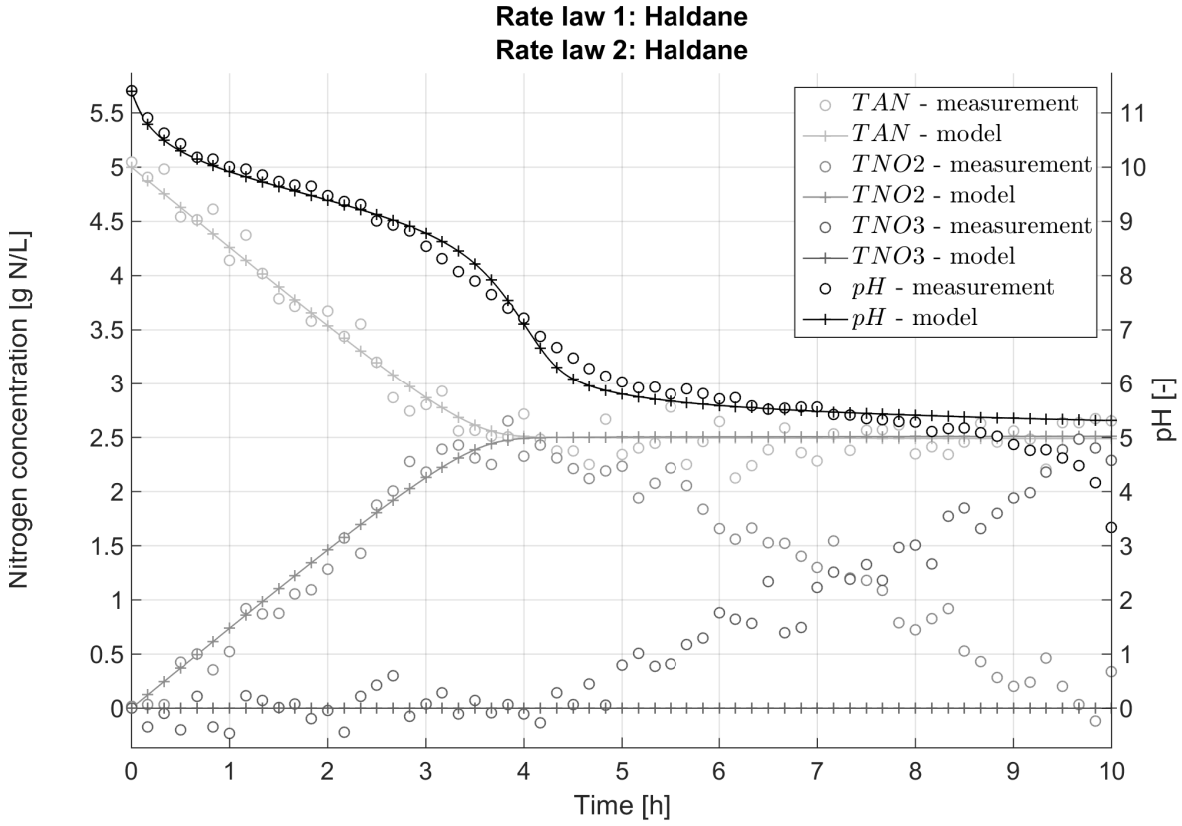

Figure S.27: **Method 1 - Simultaneous model identification - Model 25.** Measurements and simulation of the measured variables with model 25 after parameter estimation with the Nelder-Mead simplex method.

Graphical TOC Entry