# Biokinetic model identification via extents of reaction

Alma Mašić[1], Sriniketh Srinivasan[2], Julien Billeter[2], Dominique Bonvin[2] and Kris Villez[1]

[1]Eawag: Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, P.O. Box 611, CH-8600 Dübendorf, Switzerland
[2]Laboratoire d'Automatique, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland
[*]Corresponding author (Email: *alma.masic@eawag.ch*)

**Abstract**
Model structure selection and parameter identification for biokinetic modeling of biological wastewater treatment processes is broadly accepted to be a complicated task. Contributing factors include (i) nonlinear behavior, (ii) lack of knowledge, (iii) lack of (accurate) measurements, and (iv) a large number of model parameters to estimate. Several strategies have been proposed in the wastewater engineering literature to deal with the complexity of the modeling task. These include (i) experimental design, (ii) determination of identifiable parameters, and (iii) stochastic nonlinear optimization. Despite these developments, model identification remains challenging. Extent-based modeling simplifies this task by identifying each reaction kinetics separately. The available method fits in a strategy where the reaction network (graph) and its stoichiometry (matrix) are first identified. Then, the extents of reaction are computed and the identification of the individual rate functions is made in terms of extents. In this work, the original extent-based method is modified to take nonlinear constraints and measurements into account. A simulated batch process is used to demonstrate the method.

**Keywords**
Biokinetic models; Extents of reaction; Good modeling practice; Model structure identification

*Methods.* For simplicity, we assume that the (bio)chemical process that includes *R* reactions and *S* species occurs in a batch reactor without gas exchange. The stoichiometry is given by the (*R* x *S*)-dimensional stoichiometric matrix **N**. The numbers of moles of the *S* chemical species, $\mathbf{n}(t)$, can be expressed as functions of the extents of reaction, $\mathbf{x}_r(t)$:

$$\mathbf{n}(t) = \mathbf{n}_0 + \mathbf{N}^{\mathrm{T}}\mathbf{x}_r(t). \tag{1}$$

The extent of reaction $\mathbf{x}_{r,i}(t)$ expresses the number of moles produced or consumed by the *i*th reaction in the period $[0, t]$, as described in [1]. Concentrations are given by

$$\mathbf{c}(t) = \frac{\mathbf{n}(t)}{V}, \tag{2}$$

where V is the volume inside of the reactor. Assuming equilibrium for acid-base reactions, one can write a number of algebraic constraints between concentrations:

$$\mathbf{g}\big(\mathbf{c}(t)\big) = \mathbf{0}. \tag{3}$$

Finally, the measurements $\mathbf{y}(t)$ can be written as functions of concentrations,

$$\mathbf{y}(t) = \mathbf{f}\big(\mathbf{c}(t)\big) + \mathbf{e}(t), \tag{4}$$

with $\mathbf{e}(t)$ the measurement errors. This last equation allows considering nonlinear measurements, such as pH, as well as measurements of lumped states, such as total ammonia nitrogen (TAN). The extents of reaction can be estimated in the weighted least-squares sense by solving the problem

$$\hat{\mathbf{x}}_r(t) = \arg\min_{\mathbf{x}_r(t)} \Big(\mathbf{f}\big(\mathbf{c}(t)\big) - \mathbf{y}(t)\Big)^{\mathrm{T}} \mathbf{W}(t) \Big(\mathbf{f}\big(\mathbf{c}(t)\big) - \mathbf{y}(t)\Big) \tag{5}$$

$$\text{s.t.} \quad \mathbf{c}(t) = (\mathbf{n}_0 + \mathbf{N}^{\mathrm{T}}\mathbf{x}_r(t))/V \tag{6}$$

$$\mathbf{g}\big(\mathbf{c}(t)\big) = \mathbf{0}, \tag{7}$$

with $\mathbf{W}(t)$ denoting the weighting matrix, usually the inverse of the measurement variance-covariance matrix. The estimation procedure corresponds to maximum likelihood estimation if the measurement errors are normally distributed with zero mean and $\mathbf{W}(t)$ is the variance-covariance matrix.

A nice feature of the extents $\hat{\mathrm{x}}_r(t)$ is that each one contains information about a single reaction. This helps evaluate and optimize the quality of a model by considering each rate function individually. Practically, kinetic model identification proceeds by identifying the structure of the rate function and the value of the rate parameters separately for each reaction. The rate of change of each extent can be written as a function of the species concentrations,

$$d\hat{x}_i/dt = r_i(\hat{\mathbf{c}}(t), \boldsymbol{\theta_i}), \quad i = 1, \dots, R, \tag{8}$$

where $\boldsymbol{\theta_i}$ denotes the rate parameters of the *i*th reaction. However, the species concentrations are expressed in terms of all extents $\hat{\mathbf{x}}_r(t)$ as

$$\hat{\mathbf{c}}(t) = (\mathbf{n}_0 + \mathbf{N}^{\mathrm{T}}\hat{\mathbf{x}}_r(t))/V. \tag{9}$$

In the general case, the dynamics of the extents are described by means of a system of differential algebraic equations. In simple cases, the rate of change of a given extent is only a function of that extent and of its corresponding parameters. However, it is more common that the extents of the other reactions are required as inputs to some rate functions. In this study, this case is treated by means of linear interpolation of the extents, similarly to [2]. Given this analytical framework, one can then identify each rate function separately.

*Results.* A simplified two-step nitrification process is simulated to demonstrate the method. The biomass concentration is assumed to be constant. The stoichiometric matrix is given as follows:

$$\mathbf{N} = \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 & 1 \end{bmatrix}. \tag{10}$$

The columns of the stoichiometric matrix correspond to the $S = 7$ species $\{NH_3, NH_4^+, NO_2^-, HNO_2, NO_3^-, HNO_3, H^+\}$. The first $R = 2$ rows represent the biological oxidation reactions and the last three the acid-base reactions, which are modeled as instantaneous equilibria. Assuming no change in biomass concentration during the reaction, the biological reaction rate expressions can be formulated as

$$r_{NH_3}(t) = r_1(\mathbf{c}(t)) = a_{max,NH_3} \cdot \frac{c_{NH_3}(t)}{K_{A,NH_3} + c_{NH_3}(t)} \cdot \frac{K_{I,NH_3}}{K_{I,NH_3} + c_{NH_3}(t)} \quad \textit{(Haldane)} \tag{11}$$

$$r_{HNO_2}(t) = r_2(\mathbf{c}(t)) = a_{max,HNO_2} \cdot \left(1 - \exp\left(-\frac{c_{NH_3}(t)}{K_{A,HNO_2}}\right)\right) \quad \textit{(Tessier)} \tag{12}$$

The initial conditions and parameters are listed in Table 1. TAN, TNO2, TNO3, and pH are measured every 5 minutes and zero-mean normally distributed noise with standard deviations of 0.05 g N/l for TAN, TNO2 and TNO3, and 0.01 for the pH is added to the simulated data. Figure 1 shows the simulated noisy profiles.

Figures 2a and 2c show the estimated extents with their point-wise confidence intervals. The expected variances are obtained by error propagation after linearization of Eqs. 5-7 at the optimum. One can see that *(i)* the first extent is more precisely estimated than the second one, and *(ii)* the precision of the first extent changes over time. In contrast, the second extent exhibits a fairly constant variance. This can be explained by the fact that the change in pH affects the first reaction but not the second one.

The estimated extent can be used for model identification. For each reaction, the same three candidate rate functions are considered, namely the Haldane (Eq. 11) and Tessier (Eq. 12) functions as well as the Monod function with the subscript *s* denoting the species involved in the modeled reaction:

$$r_s(t) = r(\mathbf{c}(t)) = a_{max,s} \cdot \frac{c_s(t)}{K_{A,s} + c_s(t)}. \quad \textit{(Monod)} \tag{13}$$

To model the first extent, the second extent is interpolated linearly and provided as a known input. The parameters of each of the three candidate functions are adjusted by means of a nonlinear search. Figure 2a shows the resulting best-fit simulations. As expected, the Haldane function fits the first extent best. On the other hand, the best fit for the second extent is obtained with the Tessier model (see Figures 2b and 2d).

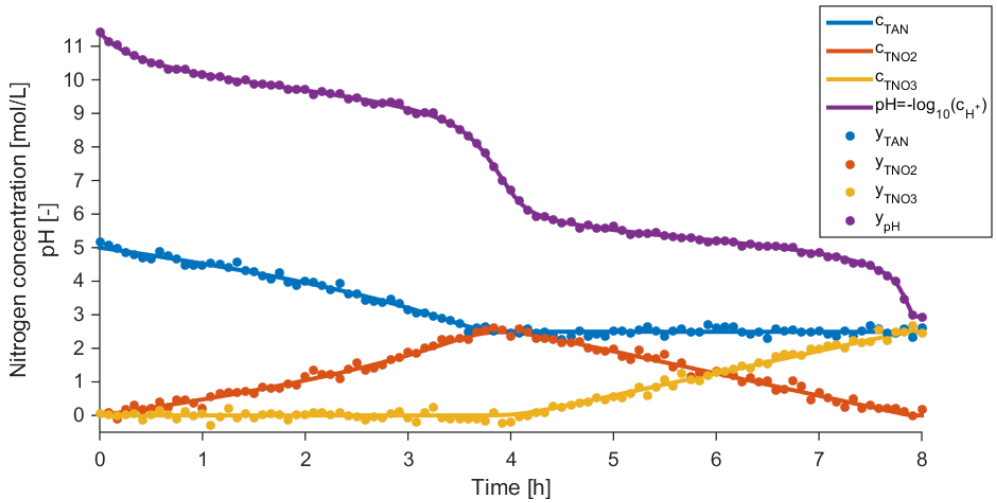


**Table 1**. Model parameters and initial conditions.

| Symbol | Value | Unit |
|---|---|---|
| $a_{max,NH_3}$ | 140 | g/L·d |
| $K_{A,NH_3}$ | 1 | g/L |
| $K_{I,NH_3}$ | 0.5 | g/L |
| $a_{max,HNO_2}$ | 16 | g/L·d |
| $K_{A,HNO_2}$ | 0.0025 | g/L |
| $c_{TAN}(0)$ | 5 | g/L |
| $c_{TNO2}(0)$ | 0 | g/L |
| $c_{TNO3}(0)$ | 0 | g/L |
| $pH(0)$ | 11.4 | - |

**Figure 1**. Simulated (lines) and measured (dots) values of TAN, TNO2, TNO3, and pH.

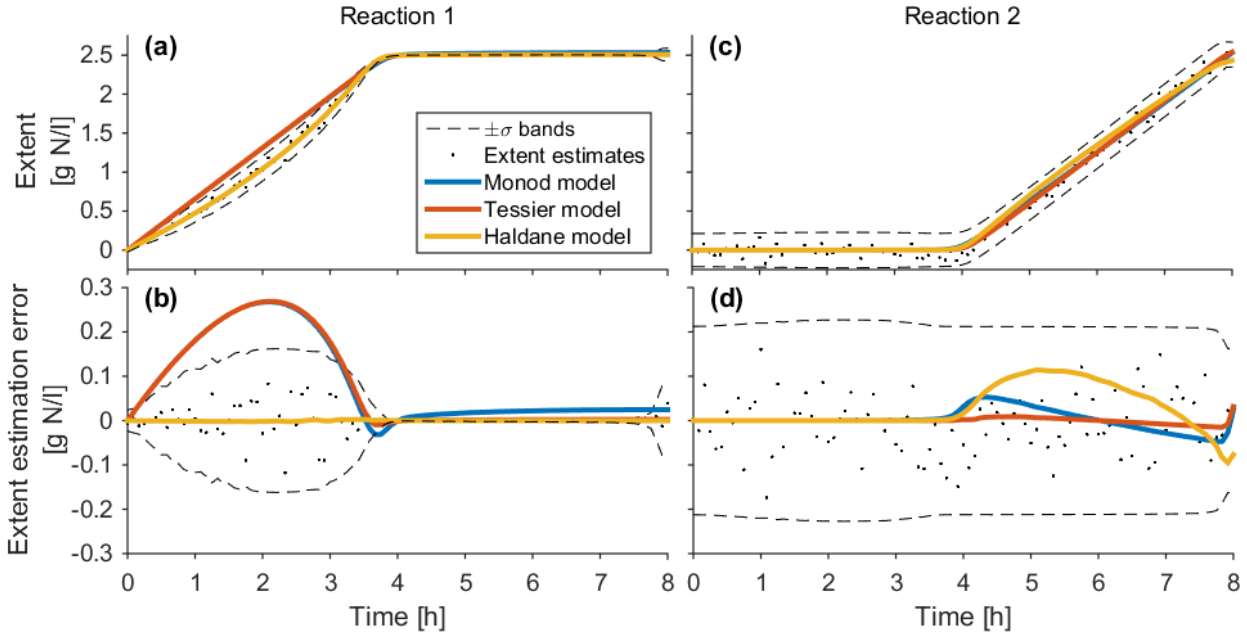


**Figure 2**. Measured (dots, 66% confidence intervals shown with dashed lines) and fitted extents (using Monod, Tessier and Haldane models, respectively the blue, red and yellow lines) for the first reaction (left, insets a and b) and the second reaction (right, insets c and d)

*Acknowledgments*. This study is financed by Eawag Discretionary Funds (PSP: 5221.00492.009.03).

*References.*

[1] Rodrigues, D., Srinivasan, S., Billeter, J. & Bonvin D. (2015). Variant and Invariant States for Chemical Reaction Systems. Comp. Chem. Eng., 73, 23-33.

[2] Bhatt, N., Amrhein, M., & Bonvin, D. (2011). Incremental Identification of Reaction and Mass–Transfer Kinetics Using the Concept of Extents. Ind. Eng. Chem. Res., 50(23), 12960-12974.