Computers and Chemical Engineering 99 (2017) 96–105

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

Shape constrained splines as transparent black-box models for bioprocess modeling

Alma Mašić [a], Sriniketh Srinivasan [b], Julien Billeter [b], Dominique Bonvin [b], Kris Villez [a,*]

[a] Eawag: Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, CH-8600 Dübendorf, Switzerland
[b] Laboratoire d'Automatique, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

ARTICLE INFO

Article history:
Available online 3 January 2017

Keywords:
Mathematical models
Microbial growth-rate kinetics
Monod equation
Shape-constrained spline function
Wastewater treatment

ABSTRACT

Empirical model identification for biological systems is a challenging task due to the combined effects of complex interactions, nonlinear effects, and lack of specific measurements. In this context, several researchers have provided tools for experimental design, model structure selection, and optimal parameter estimation, often packaged together in iterative model identification schemes. Still, one often has to rely on a limited number of candidate rate laws such as Contois, Haldane, Monod, Moser, and Tessier. In this work, we propose to use shape-constrained spline functions as a way to reduce the number of candidate rate laws to be considered in a model identification study, while retaining or even expanding the explanatory power in comparison to conventional sets of candidate rate laws. The shape-constrained rate laws exhibit the flexibility of typical black-box models, while offering a transparent interpretation akin to conventionally applied rate laws such as Monod and Haldane. In addition, the shape-constrained spline models lead to limited extrapolation errors despite the large number of parameters.

1. Introduction

Despite major advances in computational tools, the task of building reliable models for process design, monitoring, operation, and automation remains difficult (e.g., Mašić and Eberl, 2014). Quite often, modeling is challenged by the complexity and nonlinearity of the process at hand. In the case of biological systems, especially mixed cultures, a large number of key variables cannot be measured. This typically includes the concentrations of active organisms and their internal metabolites.

The lack of completeness of experimental data has led to the formulation of the activated sludge model (ASM) family in the case of biological wastewater treatment systems with suspended biomass. These models represent mixed-culture biological systems in a simplified way by identifying the most important groups of bacteria and a macroscopic description of the growth and decay processes associated with them. In these models, one makes use of *switching functions* to describe the most important effects of substrates, products, and inhibiting compounds on the growth and decay processes (Henze et al., 2008). The Monod function (Monod, 1949) is most popular to describe substrate affinity. However, the Monod model is not considered a universal representation of all bacterial behaviors (Moser, 1985). Less popular alternatives include models by Moser (1958), Tessier (1942), and Contois (1959). Importantly, this approach is necessarily empirical. In other words, these switching functions describe empirically established relationships rather than laws derived from first principles. As a result, extrapolation errors can easily be observed when a model is used to optimize process controls (e.g., Sin et al., 2006).

Avoiding extrapolation errors can in part be solved by designing experiments carefully (e.g., Donckels et al., 2009). In addition, frequent model updating might help account for stochastic changes in the process. However, modifying both the model structure and its parameters on a frequent basis leads to large computational efforts for experimental design, model structure selection, and parameter estimation. With the methods proposed and applied in this work, we aim to reduce such efforts and thereby facilitate faster model identification procedures.

Our method relies on the observation that many switching functions have the same shape despite being different functions. This is the case for the affinity switching functions discussed above. Indeed, the Contois, Monod, Moser, and Tessier switching functions exhibit the same increasing and concave shape with respect to the substrate concentration. The Monod function is often used by default, mainly to avoid large computational efforts related to the selection among the list of candidates. However, this can lead

* Corresponding author.
E-mail addresses: alma.masic@eawag.ch (A. Mašić), sriniketh.srinivasan@epfl.ch (S. Srinivasan), julien.billeter@epfl.ch (J. Billeter), dominique.bonvin@epfl.ch (D. Bonvin), kris.villez@eawag.ch (K. Villez).

http://dx.doi.org/10.1016/j.compchemeng.2016.12.017
0098-1354/

to severe extrapolation errors during process design, as is also demonstrated in Neumann and Gujer (2008). Alternatively, one can consider several candidates in a library of rate laws and select via an iterative process of experimental design, parameter estimation, and model structure selection (e.g., Sin et al., 2005). The power of such an approach increases with the number of candidate rate laws, which however results in a larger computational effort. Even if the computational requirements can be satisfied, such an approach can still fail as a library cannot be guaranteed to be universal, that is, to encompass all feasible behaviors (Refsgaard et al., 2006).

To accommodate for the lack of universality discussed above, we propose shape-constrained spline functions (SCS, Villez et al., 2013) as an alternative way to formulate rate laws. Instead of evaluating multiple candidate rate laws with approximately the same shape, we propose to use a single shape-constrained spline function for each considered shape. In other words, we replace all candidate rate laws with a given shape with a single generic rate law. Initial results obtained with this approach were presented at the DYCOPS-CAB2016 conference (Mašić et al., 2016a). The present work expands and completes this study. In Mašić et al. (2016a), simplified biological processes were simulated by assuming that the net growth is zero at all time. This led to the analysis of univariate processes. In this work, this assumption is removed, thus leading to a more general multivariate approach. In addition, while Mašić et al. (2016a) only dealt with the increasing-concave case described above, we consider here rate laws that include inhibition effects as well. Furthermore, the simulation study in this work includes *(i)* more realism, *(ii)* a single improved parameter estimation method for parameter estimation in practical conditions, *(iii)* a validation test demonstrating that extrapolation errors are limited, and *(iv)* a more detailed interpretation and discussion of the results.

The considered spline functions are flexible thanks to the use of a large number of parameters. As a result, they can describe a wide range of kinetic behaviors, akin to black-box modeling approaches (Guay et al., 2004). Note that the application of shape constraints ensures the identifiability and straightforward interpretation of the resulting models, as will be shown below. Shape restrictions are commonly applied for fitting hazard models (Meyer, 2008). More recently, SCS functions were adopted for fault detection and fault diagnosis in a qualitative trend analysis framework (Villez et al., 2013; Villez and Habermacher, 2016). The main difference with these previous studies is that the SCS functions now appear inside a set of nonlinear differential equations.

2. Differences with prior work

The differences with the previous DYCOPS-CAB study (Mašić et al., 2016a) are:

- The substrate and biomass concentrations are considered as state variables, as opposed to the DYCOPS-CAB case, where the biomass concentration was assumed constant. As a consequence, the estimated parameters are associated with the stoichiometry, the growth rate, and the decay rate, whereas the DYCOPS-CAB study only considered the growth rate.
- The estimation of additional parameters in the multivariate case called for the development of a new parameter estimation method (see Section 3.4.2 below).
- The simulated experiments have been modified to appear more realistic. In particular, the sampling frequency used in the DYCOPS-CAB paper has been reduced significantly.
- In this work, only one parameter estimation procedure is used for indirect model fitting of every model (Section 3.4.2). In contrast, the DYCOPS-CAB study used different parameter estimation procedures for the conventional rate laws and the SCS-based rate laws.
- This work includes a validation test, in which the identified models are tested for their extrapolative capability. Such a test was not part of the DYCOPS-CAB study.
- All figures in this paper are new. Although Figs. 4–6b bear similarity with figures in the DYCOPS-CAB study, the data and their interpretation have been modified according to the changes made in the simulations. Furthermore, Figs. 1–3b and 7a–8, which describe new ideas and results, were not in the DYCOPS-CAB study.
- The discussion and conclusion sections were modified and expanded significantly.

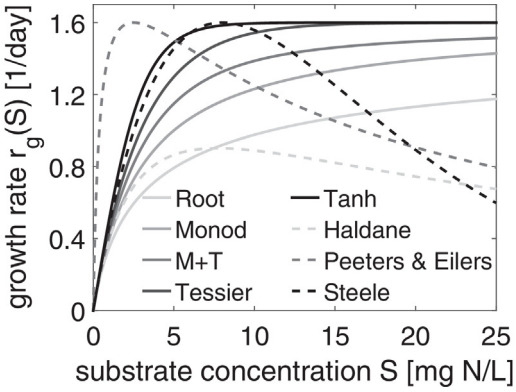

Fig. 1. Growth-rate laws considered in this work, as functions of the substrate concentration. The rate laws are defined in Table 1.

3. Mathematical model & methods

3.1. Model description

In this study, simple models describing bacterial growth and decay are used. The models are similar in structure to the activated sludge models discussed in Henze et al. (2008). Let $S(t)$ and $X(t)$ denote the substrate and biomass concentrations at time t. The change in these concentrations with respect to time can be expressed as

$$\frac{dS}{dt} = -\frac{r_g(S)}{Y}\, X, \quad S(0) = S^0 \tag{1}$$

$$\frac{dX}{dt} = r_g(S)\, X - r_d(X), \quad X(0) = X^0 \tag{2}$$

where $r_g(S)$ and $r_d(X)$ are rate laws expressing the bacterial growth and decay as a function of S and X, respectively. The metabolic product concentration $P(t)$ can be computed as

$$P(t) = S^0 - S(t). \tag{3}$$

The initial concentrations are S^0 and X^0. This model describes growth and decay as distinct processes in contrast to Mašić et al. (2016a) which implicitly assumed the two process rates are the same at all times.

The expression $r_g(S)$ for the specific growth rate can be varied to express the effects of substrates, products, and other chemical species. In this study, we consider a classical set of rate laws describing both uninhibited and inhibited bacterial growth processes. This reflects a situation where no a priori knowledge is available about the structure of the kinetic growth-rate law. The considered growth-rate laws are described in the next section. For