

Biokinetic process model diagnosis with shape-constrained spline functions

A. Mašić*, S. Srinivasan**, J. Billeter**, D. Bonvin**, K. Villez*

* Swiss Federal Institute of Aquatic Science and Technology (EAWAG), Überlandstrasse 133, 8600 Dübendorf, Switzerland

** Laboratoire d'Automatique, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

Abstract: Model-structure identification is important for the optimization and design of biokinetic processes. Standard Monod and Tessier functions are often used by default to describe bacterial growth with respect to a substrate, leading to significant optimization errors in case of inappropriate representation. This paper introduces shape-constrained spline (SCS) functions, which share the qualitative behavior of a number of conventional growth-rate functions expressing substrate affinity effects. A simulated case study demonstrates the capabilities in terms of model identification of SCS functions, which offer a high parametric flexibility and could replace incomplete libraries of functions by a single biokinetic model structure. Moreover, the diagnostic ability of the spline functions is illustrated for the case of Haldane kinetics, which exhibits a distinctively different shape. The major benefit of these spline functions lies in their model discrimination capabilities by indicating in a quick and conclusive way the presence of other effects than substrate affinity.

Keywords: bacterial growth-rate kinetics; shape-constrained spline functions; wastewater treatment

Introduction

Biological wastewater treatment makes use of bacterial growth to remove undesired compounds, such as nitrogen and phosphorus, from wastewater (Henze et al., 2008). The biological treatment step involves many complex biological and biochemical processes. These processes can be described by mathematical models, which are helpful tools in the analysis and understanding of processes. Moreover, mathematical modeling can be used for design, interpretation, and optimization of wastewater treatment. The possibility to use a model for the above purposes often depends on the model structure. Biokinetic process models are well established for biological wastewater treatment, including the activated sludge models (Henze et al., 2008) and the biofilm models (Wanner et al., 2006). Most of these models rely on Monod or Tessier kinetics (Bastin&Dochain, 1990) to express bacterial growth rates with respect to a given substrate – an assumption that is rarely verified experimentally. Bacteria are unlikely to follow such idealized kinetics, which can lead to design errors and incorrect optimization when a wrong model structure is used (Neumann&Gujer, 2008). Even when the model is initially verified, it is possible that the bacterial behavior can change over time, resulting in a wrong model for the process at hand.

The model structure can be identified by fitting experimental data to a library of well-known bacterial kinetic growth-rate functions. This approach requires performing multiple parameter identifications, one for each function in the library. Optimization time can be reduced by constructing a smaller library, but this immediately excludes certain functions and decreases the chance of finding the correct or most appropriate function. Regardless of its size, such a conventional library can never be completely exhaustive (Refsgaard et al., 2006). A new approach is to use shape-constrained spline (SCS) functions (Mašić et al., submitted) to determine the kinetic growth-rate functions.

Substrate affinity can be described as a monotonically increasing function with a concave shape. This qualitative description can be used in the construction of SCS functions (Villez et

al., 2013), which are flexible enough to reproduce the shape and serve as a single biokinetic growth model that can be interpreted as any conventional model.

In previous work (Mašić et al., submitted), we investigated the use of SCS functions to describe simulated bacterial growth rates, and we compared their performance to the standard approach of fitting a library of functions. The library consisted of the Monod and Tessier functions, while data were simulated with other growth-rate functions such as hyperbolic tangent, square root and combined Monod-Tessier functions. It was shown that it is possible to fit an SCS function directly to growth-rate data and that this SCS function is a suitable alternative to standard growth-rate functions for all the studied cases.

The results above were based on the assumption that good measurements of growth rates were available for direct fitting. However, this is not always the case. Hence, we now study the fitting of the growth rates described by simple ordinary differential equations and simulated measurements of concentrations of compounds commonly available in bioreactors. Moreover, we focus on the diagnostic feature of the SCS functions and investigate how the knowledge and the exclusion of certain shapes can help assess the model structure. To this end, we introduce the Haldane growth-rate function (Haldane, 1930), which has a different shape reflecting substrate-inhibition effects.

Material and Methods

Let us assume a simple model of bacterial substrate consumption, interpreted as the conversion of ammonia to nitrite by ammonia oxidizing bacteria, as $S \rightarrow P$, described by

$$\frac{d}{dt}S(t) = -\mu(S(t)), \quad S(0) = S^0 \quad (1)$$

where S denotes the concentration of the substrate (total ammonia nitrogen, TAN) and $\mu(\cdot)$ is the growth rate. The concentration of the product P (total nitrite nitrogen, TNO₂) is computed as $P(t) = S^0 - S(t)$. The bacterial activity μ can be described by a variety of functions. The library of candidate rate functions consists of two conventional functions, the Monod μ_M and Tessier μ_T functions. Three more exotic functions (hyperbolic tangent μ_{HT} , square root μ_R , combined Monod and Tessier μ_{MT}) as well as an inhibition function (Haldane μ_H) are considered but not included in the library:

$$\mu_M(S) = a^{max} \frac{S}{K_S + S} \quad (2) \quad \mu_T(S) = a^{max} (1 - e^{-SK_S}) \quad (5)$$

$$\mu_{HT}(S) = a^{max} \tanh \frac{S}{K_S} \quad (3) \quad \mu_R(S) = a^{max} \frac{\sqrt{4+5S/\sqrt{1.5}-2}}{\sqrt{4+5S/\sqrt{1.5}-2+K_S}} \quad (6)$$

$$\mu_{MT}(S) = \frac{a^{max}}{2} \left((1 - e^{-SK_S}) + \frac{S}{K_S + S} \right) \quad (4) \quad \mu_H(S) = a^{max} \frac{S}{K_S + S + S^2/K_I} \quad (7)$$

where K_S is the half saturation coefficient, a^{max} is the maximum activity of the biomass, and K_I is the substrate inhibition coefficient. Rate equations (2) – (7) rely on the assumption of negligible biomass net growth.

The SCS function is a piecewise-cubic polynomial function (using a natural cubic spline basis) in the substrate concentration. It can be formulated as a weighted sum of spline basis functions $\mu_{SCS} = \mathbf{b}_0(S)^T \boldsymbol{\theta}$, where $\mathbf{b}_0(S)$ is the $(n_k + 1)$ -dimensional vector of spline basis functions evaluated at the substrate concentration S , and $\boldsymbol{\theta}$ the $(n_k + 1)$ -dimensional vector of parameters (Villez et al., 2013). Equidistantly placed $(n_k + 1)$ knots determine the piecewise behavior of the function. In addition, shape constraints are imposed on the parameters $\boldsymbol{\theta}$.

Estimation of the parameters K_S , a^{max} , K_I , and θ was performed numerically in Matlab with built-in optimization functions `lsqnonlin`, `fminunc`, `fmincon`, and with the SCS toolbox (Villez et al., 2013). The values $S^0 = 25$ mg N/L, $K_S = 2$ mg N/L, $a^{max} = 11$ mg N/(L·day) and $K_I = 25$ mg N/L were used for simulation. The SCS function was constructed with 27 knots. The simulated measurements were obtained by solving Eq. (1) with a selected rate function for $t = [0, 4]$ h at a sampling time of 2.5 min, resulting in 97 measurement points. A zero-mean Gaussian noise with standard deviation $\sigma = 0.1$ was added to each point.

Results and Conclusions

Parametric flexibility of the SCS approach: The considered growth rates are shown in Figure 1a. Five out of the six functions exhibit the same qualitative shape, namely, a steep increase at low substrate concentrations and a plateau at high concentrations. The Haldane function has a distinctively different shape, with a decreasing rate at high substrate concentrations, which can be explained by substrate inhibition. Figure 1b shows the weighted root mean square residuals (WRMSR) for the fitting of the six growth rates using the library functions (Monod and Tessier growth rates) and the SCS function, based on simulated noisy measurements of TAN and TNO₂. The SCS function performs just as well as or better than the library functions in all cases. The Monod function does not perform well when it is fitted against the hyperbolic tangent or Tessier functions. Hence, the use of the Monod kinetics as default rate function is not always appropriate. Extrapolation of results and conclusions based on Monod kinetics to other cases might generally be unreliable. On the other hand, the approach based on the SCS function provides a single biokinetic model, encompassing all six models, which can easily be adjusted to each situation.

Diagnostic ability of the SCS approach: The Haldane function – a very different function compared to the other five – was used to simulate noisy measurements of TAN and TNO₂, as shown in Figure 1c. It is clear that neither the best library function – in this case the Monod function – nor the SCS function is able to fit the data well. By looking at Figure 1d one can see that the residuals are large and autocorrelated, indicating that the applied models do not fit the data well. This is also noticeable for the Haldane function in Figure 1b, where the WRMSR values are much larger than for any of the other simulated functions. If one was only using the Monod function to model this case, one could not exclude that another unidentified function with the same shape exists. However, by using the flexible SCS approach, it is now possible to conclusively reject the hypothesis of an isotonic-concave shape for this process behavior. Beside its fitting flexibility, this diagnostic ability is a major benefit of the SCS model.

Towards a single biokinetic model structure: Further work will include several different shapes for the SCS function, mimicking the library of functions, but with a significantly larger potential to determine the rate-function structure. The elements in the library will be ordered according to the probability of finding a particular shape among bacterial growth rates. Although the SCS functions are described by more parameters than conventional growth-rate functions, the computation time is comparable to or lower than that of a standard library approach. The benefit of the SCS functions lies in the fact that they are exhaustive representations of each shape. This new methodology is expected to be beneficial in various applications, including in the biological nitrification of collected source-separated urine for resource recovery, the success of which relies on a good model of the nitrification process.

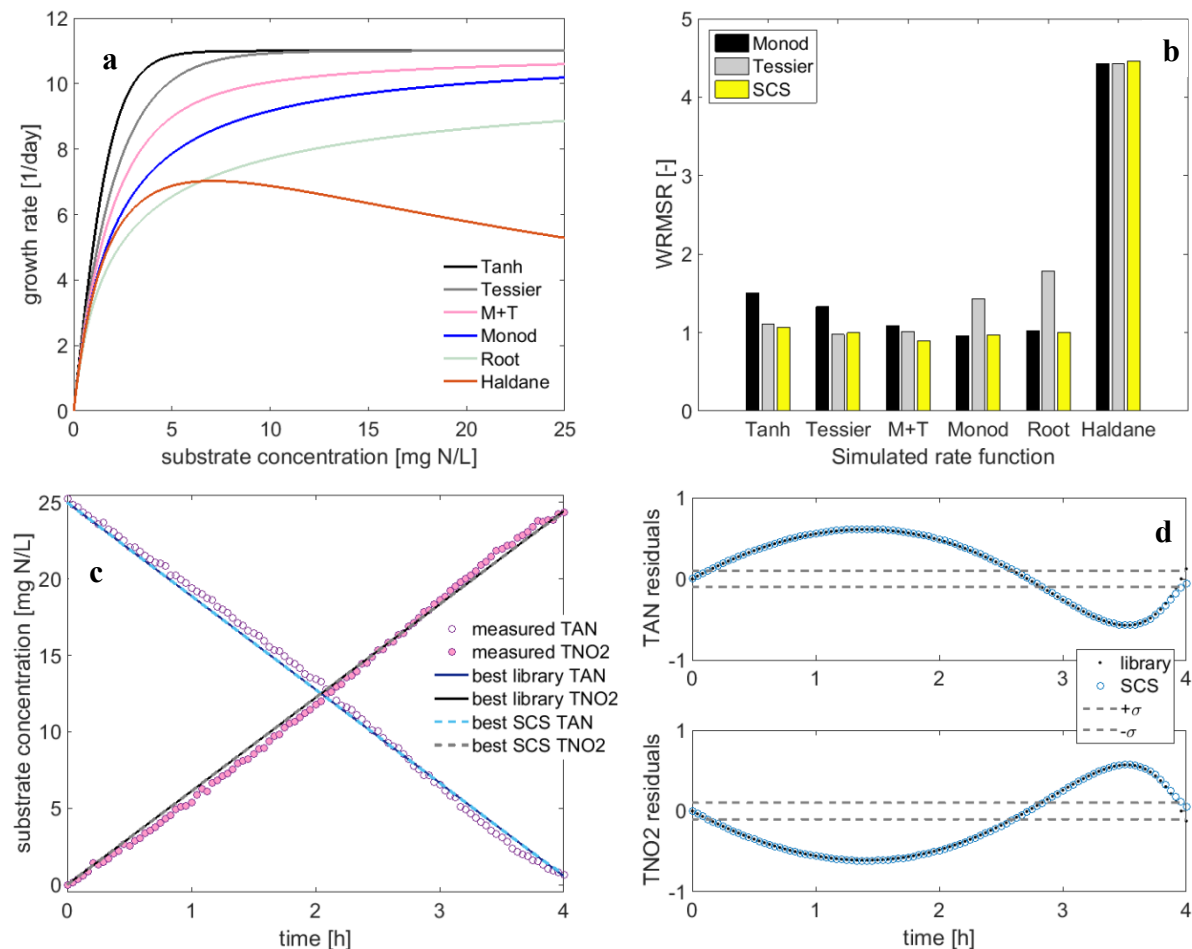


Figure 1.1 a) Bacterial growth rates as functions of substrate concentration, b) weighted root mean square residuals for the fitting of the library of functions (Monod and Tessier) and an SCS function to data simulated with six different growth-rate functions, c) simulated TAN and TNO2 concentrations obtained with the Haldane function, and best library function (Monod) and SCS function, d) TAN and TNO2 residuals from figure c) for the best library (Monod) and SCS functions with 68% confidence intervals ($\pm \sigma$).

References

- Bastin, G., Dochain, D. (1990). *On-line estimation and adaptive control of bioreactors*, Elsevier.
- Haldane, J.B.S. (1965). *Enzymes, 1930*. Reprint by MIT, Cambridge.
- Henze, M., van Loosdrecht, M., Ekama, G., Brdjanovic, D. (eds.) (2008). *Biological wastewater treatment: Principles, modelling and design*. IWA Publishing.
- Mašić, A., Srinivasan, S., Billeter, J., Bonvin, D., Villez, K. (submitted) Shape-constrained splines as transparent black-box models for biokinetic process modeling. Submitted to the *11th IFAC Symposium on Dynamics and Control of Process Systems, including Biosystems*, Trondheim, Norway, June 6-8, 2016.
- Neumann, M., Gujer, W. (2008). Underestimation of uncertainty in statistical regression of environmental models: influence of model structure uncertainty. *Environ. Sci. Technol.*, **42**(11), 4037-4043.
- Refsgaard, J., der Sluijs, J.V., Brown, J., der Keur, P.V. (2006). A framework for dealing with uncertainty due to model structure error. *Adv. Water Resour.*, **29**(11), 1586-1597.
- Villez, K., Venkatasubramanian, V., Rengaswamy, R. (2013). Generalized shape-constrained spline fitting for qualitative analysis of trends. *Comp. Chem. Eng.*, **58**, 116-134.
- Wanner, O., Eberl, H., Morgenroth, E., Noguera, D.R., Picioreanu, C., Rittmann, B., van Loosdrecht, M. (2006). *Mathematical modeling of biofilms*. In Scientific and Technical Report No.18. IWA Publishing.