Preprints of the 6th Int. Conference on
Computer Applications in Biotechnology
Garmisch-Partenkirchen, Germany
May 14 - 17, 1995

IDENTIFIABILITY OF BIOKINETIC MODELS IN AN ACTIVATED SLUDGE PROCESS

P. Vanrolleghem*, M. Van Daele* and D. Dochain**

* *Laboratory of Microbial Ecology, University of Gent, Coupure Links 653, B-9000 Gent, Belgium, Fax : 32-9-2646248, e-mail : Peter.Vanrolleghem@rug.ac.be*

***Chercheur qualifié FNRS, Cesame, Université Catholique de Louvain, Bât. Euler, av. G. Lemaître 4-6, 1348 Louvain-La-Neuve, Belgium, Fax : 32-10-472180, e-mail : dochain@auto.ucl.ac.be*

Abstract. This paper deals with the identifiability of parameters of kinetic models used in activated sludge processes. The identifiability analysis is based on the on-line oxygen uptake rate data given by a novel respirometric biosensor. Four model candidates (exponential, Monod, double Monod and IAWQ $n^\circ 1$) are considered. For each model, only a smaller set of the original parameters are structurally identifiable. It is also illustrated how to design experiments so as to obtain on-line improved parameter estimates.

Key Words. Activated Sludge Process, Identifiability, Optimal Experimental Design

1. INTRODUCTION

Because of the model complexity and the scarcity of on-line sensors, the identifiability study of the dynamical models, prior to any identification, is certainly a key question in activated sludge processes. While the theoretical identifiability is studied under the assumption of perfect data, the problem with highly correlated parameters arises when a limited set of experimental, noise-corrupted data is used for parameter estimation. Under such conditions the uniqueness of parameter estimates, though predicted by the theoretical analysis, is no longer guaranteed, because a change in one parameter can be compensated almost completely by a proportional shift in another one, still producing a satisfying fit between experimental data and model predictions. In addition, the numerical algorithms that perform the nonlinear parameter estimation show poor convergence when faced with this type of ill-conditioned optimization problems, the estimates being very sensitive to the initial parameter values given to the algorithm (Holmberg, 1982). Consequently, the estimated parameters may vary over a broad range, resulting in little physical interpretation of the parameter values. Few studies were directed at the design of experiments by which more informative data can be collected. In (Munack, 1989), it is shown how to modify batch experiments in order to improve parameter confidences by optimal experimental design techniques.

The goal of this paper is to study both the theoretical and practical identifiability of a class of Monod-based biological models used to describe activated sludge processes. In contrast to the studies found in the literature, the analysis will start from the assumption that only oxygen uptake rate data are available. The experiments that are evaluated with respect to their information content are short-term batch experiments (typically 30 minutes), with an extension towards simple fed-batch systems. Optimal experimental designs are proposed and validated with real-life experiments. A novel aspect of this work is that the procedures are implemented for on-line use in a respirometric biosensor installed at the treatment plant, allowing to update the parameters in the process models on-line. The so-called "In-Sensor-Experiments" performed in this respirometer are preferred over data obtained from the full-scale treatment plant since identification-in-the-loop may be subject to serious experimental constraints, leading to important practical identifiability problems. The paper is organized as follows. In section 2, the models considered in the paper are presented. In section 3, theoretical identifiability results are given. In section 4, some problems of practical identifiability of the models are illustrated. In section 5, the optimal experimental designs for parameter estimation are validated with real-life data.

2. MATHEMATICAL MODELS

The models considered in this study express the dependence of the exogenous oxygen uptake rate OUR_{ex} on the biodegradation of k substrates S_i

present in the mixed liquor:

$$OUR_{ex} = -\sum_{i=1}^{k}(1 - Y_i)\frac{dS_i}{dt} \tag{1}$$

In the above expression, Y_i (the yield coefficient) is the fraction of pollutant S_i which is not oxidized but converted into new biocatalyst X. As usual, all concentrations are expressed in chemical oxygen demand (COD) units. The different model complexities taken into account in this work express the different number k of pollutants to be considered and the degradation mechanism. The four types of wastewater/sludge interaction included are:

Type 1 (Exponential) : One pollutant, first order kinetics (k=1)

$$\frac{dS_1}{dt} = -\frac{\mu_{max1}X}{Y_1}S_1 \tag{2}$$

Type 2 (Single Monod) : One pollutant, Monod kinetics (k=1)

$$\frac{dS_1}{dt} = -\frac{\mu_{max1}X}{Y_1}\frac{S_1}{K_{m1} + S_1} \tag{3}$$

Type 3 (Double Monod) : Two pollutants, double Monod kinetics (k=2)

$$\frac{dS_i}{dt} = -\frac{\mu_{maxi}X}{Y_i}\frac{S_i}{K_{mi} + S_i}, \; i = 1, 2$$

Type 4 (Modified IAWQ model n^o1) : 3 pollutants, 2 hydrolysed into the first substrate which is used for growth according to the Monod kinetics (k=1)

$$\begin{aligned}\frac{dS_1}{dt} &= -\frac{\mu_{max1}X}{Y_1}\frac{S_1}{K_{m1} + S_1} + k_rX_r + k_sX_s \\ \frac{dX_r}{dt} &= -k_rX_r \\ \frac{dX_s}{dt} &= -k_sX_s\end{aligned}$$

In this model, OUR_{ex} should be rewritten as follows :

$$OUR_{ex} = (1 - Y_1)(\frac{dS_1}{dt} - k_rX_r - k_sX_s)$$

The μ_{maxi} and k_i are rate constants, and the K_{mi} are the so-called affinity constants expressing the dependency of the degradation rate on the concentration of pollutant S_i. Note that the models as given above can be used to describe batch experiments. An additional term in the mass balance is required to describe the fed-batch experiments.

As far as known by the authors, no studies of the theoretical identifiability of the parameters in Monod-type models have been reported based on the assumption that only oxygen uptake rate data are available to the experimenter. The experiments on which the model identification is to be based, are performed in such a way that :

- the change in biomass concentration can be assumed negligible (which is a fair assumption provided S(t=0) $<<$ X(t=0));
- the OUR data are only due to exogenous (= substrate induced) respiration (OUR_{ex}), i.e. endogenous respiration is either assumed negligible or is eliminated from the data.

3. STRUCTURAL IDENTIFIABILITY

In this section attention will focus on the structural identifiability of the 4 models introduced above. Let us first concentrate on the structural identifiability of the first model : if the equation of the model is combined with the one which relates OUR to S_1 (1), and one defines $y(t) = \int_0^t OUR(\tau)d\tau$, one readily obtains :

$$\begin{aligned}\frac{dy}{dt} &= \theta_1 y(t) + \theta_2 \\ \theta_1 &= -\frac{\mu_{max1}X}{Y_1}, \; \theta_2 = \frac{\mu_{max1}X}{Y_1}(1 - Y_1)S_1(0)\end{aligned}$$

Because $dy(t)/dt$ is measured and $y(t)$ is readily calculated, it is straightforward to see that the parameters θ_1 and θ_2 (i.e. the combinations of the model parameters $\mu_{max1}X/Y_1$ and $(1 - Y_1)S_1(0)$) are theoretically identifiable. We can proceed similarly for the other models : the results are summarized in Table 1 (for further details, see (Dochain et al., 1994)).

Exponential	Single Monod
$(1 - Y_1)S_1(0)$ $\mu_{max1}X(1 - Y_1)/Y_1$	$(1 - Y_1)S_1(0)$ $\mu_{max1}X(1 - Y_1)/Y_1$ $(1 - Y_1)K_{m1}$
Double Monod	**Modified IAWQ n^o1**
$(1 - Y_1)S_1(0)$ $\mu_{max1}X(1 - Y_1)/Y_1$ $(1 - Y_1)K_{m1}$ $(1 - Y_2)S_2(0)$ $\mu_{max2}X(1 - Y_2)/Y_2$ $(1 - Y_2)K_{m2}$	$(1 - Y_1)S_1(0)$ $\mu_{max1}X(1 - Y_1)/Y_1$ $(1 - Y_1)K_{m1}$ $(1 - Y_1)X_r(0)$ k_r, k_s $(1 - Y_1)X_s(0)$

Table 1. Identifiable parameter combinations

4. OPTIMAL EXPERIMENTAL DESIGN

An essential reason for optimal experimental design for parameter estimation (OED/PE) is that the measurement data must allow unique determination of the (combinations of) parameters shown to be theoretically identifiable, i.e. produce "informative" experiments (Goodwin, 1987). A second goal must be to increase the preciseness of (some of) the parameter estimates. Different strategies have been developed to reach these goals. The Fisher Information Matrix F or, equivalently, the covariance matrix are the corner-stone of these optimal experimental design procedures because these matrices summarize the information content of an experiment or the preciseness of the parameter estimates. Depending on the requirements imposed by the application different scalar measures of these matrices are optimized. A number of such deduced variables that have been proposed are (Munack, 1991) : A-optimal design criterion (min $\mathrm{tr}(F^{-1})$), modified A-optimal design criterion (max $\mathrm{tr}(F)$), D-optimal design criterion (max $\det(F)$), E-optimal design criterion (max $\lambda_{min}(F)$), modified E-optimal design criterion (min $\frac{\lambda_{max}(F)}{\lambda_{min}(F)}$), in which $\lambda_{min}(F)$ and $\lambda_{max}(F)$ are the smallest and largest eigenvalue of the Fisher information matrix.

Designing identification experiments requires several choices, e.g. what outputs should be measured at what time instants and at what frequency, and what inputs to manipulate and in what way. An illustrative example of the effect of various output combinations is given in (Munack, 1991) while the problem of optimal timing of sampling is addressed in (Holmberg, 1982) and (Vialas et al., 1986). In this work, the output and sampling frequency are no longer available to the experimenter since they are fixed by the hardware used in the study. The only degree of freedom left is the design of the input. Optimal experimental design therefore reduces to finding the input functions $u(t)$ that lead to the most informative experiments. Finally, it must be stressed that optimal experiment designs for nonlinear model are influenced by the parameter values since the design criteria are based on the Fisher information matrix which is then parameter dependent. Hence, as the bioprocess considered here is highly time-varying, it must be emphasized that the optimal design is to be adjusted on-line.

Four OED/PE examples have been studied (optimal initial substrate, optimal additional pulse with fixed initial substrate, optimal additional pulse and initial substrate,multiple additional pulses) (Vanrolleghem et al., 1994a). Let us present the results of the second case (optimal additional pulse with fixed initial substrate) for a single Monod model.

If one considers that the pulse characteristics are fixed by the hardware used, i.e. pulse volume of the sample pump and mixing intensity in the reactor, the only degree of freedom to be evaluated here is the time of pulse addition, t_{puls}. As a matter of illustration, let us consider an additional pulse of 8 mg/l given at the optimal time in a fed-batch experiment, according to the modified E criterion (t_{puls} = 18.2 min). The covariance matrices V of the parameter vector $[\mu_{max1}\ K_{m1}]^T$ for the reference (no pulse) and optimal experiment for $S_1(t=0)$ = 23 mg/l shows that all variances and covariances have improved:

$$V_{reference} = \begin{pmatrix} 1.175\ 10^{-8} & 3.742\ 10^{-4} \\ 3.742\ 10^{-4} & 15.8 \end{pmatrix}$$

$$V_{OED/PE} = \begin{pmatrix} 9.623\ 10^{-9} & 1.752\ 10^{-4} \\ 1.752\ 10^{-4} & 6.735 \end{pmatrix}$$

The OED with an additional pulse is especially attractive for a more accurate estimation of the affinity constant. Indeed, while the confidence interval for the μ_{max1} only decreased with 10 percent, the K_{m1} precision increased with more than 50 percent. In addition the results show that the covariance between both biokinetic parameters is reduced almost to the same extent (For the results with the other OED/PE criteria, see (Vanrolleghem et al., 1994a)).

5. EXPERIMENTAL VALIDATION OF OED/PE

The real-life experiments used here are obtained from an on-line respirometric biosensor that performs (fed-)batch experiments automatically. Typically 30 minute records of OUR_{ex} data are produced and subjected to an identification procedure. The device producing the data used to identify biokinetic parameters in activated sludge models was developed some ten years ago with the purpose of determining the load and toxicity of wastewaters entering an activated sludge plant. The RODTOX (Kelma bvba, Niel, Belgium), acronym for Rapid Oxygen Demand and TOXicity tester, has since then been implemented on both industrial and municipal wastewater treatment plants (Vanrolleghem et al., 1994b). All experimental results reported were obtained with sludge taken from the wastewater treatment plant at Maria Middelares, Gent.

The effect of an additional substrate pulse was validated with 3 experiments in which the substrate concentration in the bioreactor was increased with a pulse of 2 mg COD/l. Different injection times were tested so as to illustrate the effect of an optimal t_{puls}.

The calculations are based on a single Monod model with parameters deduced from a reference batch experiment, and result in criterion values versus injection time, which exhibit differences in optimal time for the different design criteria. The A-, D- and E-criteria propose to inject a pulse after 14.6 minutes (after complete degradation of the initial substrate). The modified A-criterion based experiment consists of a prolonged batch-phase. And the modified E criterion OED results in a respirogram in which the oxygen uptake reaccelerates before complete disappearance of the initial amount of substrate.

The three experiments that were performed had injection times of 13, 14.1 and 14.6 minutes respectively. The resulting OUR_{ex} profile for t_{puls} = 14.1 is given in Figure 1. The following conclusions can be drawn by focussing on the effect of these fed-batch experiments on the error functional shape and parameter variances.

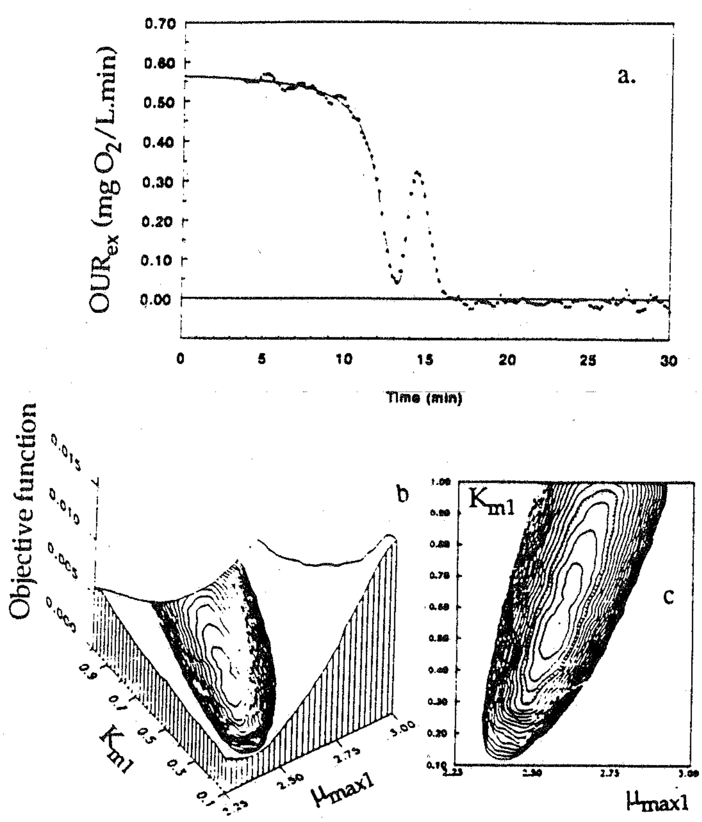

Figure 1. 3D- (b) and contour plot (c) of the objective function as a function of the Monod parameters with additional subtrate pulse at t = 14.1 min (a).

t_{puls}	Mod. E	var(μ_{maxl})	var(K_{ml})	covar.
-	1	1	1	1
13.0	0.676	0.411	0.422	0.381
14.1	0.624	0.535	0.465	0.468
14.6	0.619	0.480	0.409	0.417

Table 2. Modified E criterion and parameter variances' dependence versus pulse addition time

Although the expected values for the modified E criterion and the variances may change to a certain extent from the actually observed values due to changes in noise level, experimental error and also biological changes, the trends set by the theoretical analysis are confirmed with these results. Predicted modified E criterion values for instance were approx. 20% underestimated compared to the actual values. However, the data given in Table 2 clearly illustrate that still a significant improvement in shape of the error functional is obtained with fed-batch experiments.

A second conclusion concerns the variances. The experimental results confirm that significant improvements in parameter estimation accuracy can be obtained by this small extension of the experiment. The variances have decreased with more than 50 % (Table 2). One should note that a similar effect can be obtained by repeating the experiment twice, but this would double the experimentation time while the approach taken here increases the experiment duration with only 3 min., i.e. 10 % of normal operation.

References

Dochain, D., Vanrolleghem, P., and Van Daele, M. (1994). Structural identifiability of biokinetic models in activated sludge processes. *submitted for publication.*

Goodwin, G. (1987). Identification: Experiment design. In Singh, M., editor, *Systems and Control Encyclopedia*, volume 4, pages 2257–2264. Pergamon Press, Oxford.

Holmberg, A. (1982). On the practical identifiability of microbial growth models incorporating michaelis-menten type nonlinearities. *Math. Biosci.*, 62:23–43.

Munack, A. (1989). Optimal feeding strategy for identification of monod-type models by fed-batch experiments. *Proc. Comp. Appl. in Ferm. Techn. : Mod. and Control of Biotech. Proc.*, pages 195–204.

Munack, A. (1991). Optimization of sampling. In Schugerl, K., editor, *Biotechnology, a Multi-volume Comprehensive Treatise*, volume 4, pages 251–264. VCH, Weinheim.

Vanrolleghem, P., Van Daele, M., and Dochain, D. (1994a). Practical identifiability of a biokinetic model in activated sludge processes. *submitted for publication.*

Vanrolleghem, P., Kong, Z., Rombouts, G., and Verstraete, W. (1994b). An on-line respirographic biosensor for the characterization of load and toxicity of wastewaters. *J. Chem. Technol. Biotechnol.*, 59:321–333.

Vialas, C., Cheruy, A., and Gentil, S. (1986). An experimental approach to improve the monod model identification. In Johnson, A., editor, *Modelling and Control of Biotechnological Processes*, pages 155–160. Pergamon, Oxford.