Pergamon

0043-1354(95)00106-9

Wat. Res. Vol. 29, No. 11, pp. 2571–2578, 1995

0043-1354/95 $9.50 + 0.00

STRUCTURAL IDENTIFIABILITY OF BIOKINETIC MODELS OF ACTIVATED SLUDGE RESPIRATION

D. DOCHAIN,[1] PETER A. VANROLLEGHEM[2]Ⓜ and MARC VAN DAELE[2]

[1]Chercheur qualifié FNRS, Cesame, Université Catholique de Louvain, Bât. Euler, av. G. Lemaître 4-6, B-1348 Louvain-La-Neuve and
[2]BIOMATH, University of Gent, Coupure Links 653, B-9000 Gent, Belgium

(First received August 1994; accepted in revised form March 1995)

Abstract—This paper deals with the identifiability of parameters of kinetic models describing the activated sludge process. The main concern of the paper is to present important aspects of the structural identifiability properties. The identifiability analysis is based on the availability of only on-line oxygen uptake rate data (given by a novel respirographic biosensor). Four model candidates (exponential, Monod, double Monod and modified IAWQ No. 1) are considered. Two different methods (Taylor series expansion, and transformation of the nonlinear model into a model linear-in-the-parameters) are considered, their advantages and drawbacks are illustrated with the four kinetic models. For each model it is found that only a smaller set of the original parameters are structurally identifiable on the basis of oxygen uptake rate data only.

Key words—activated sludge wastewater treatment, kinetic models, identifiability, respirometry

NOMENCLATURE

a, b, c = parameters
k_r, k_s = kinetic constants (min^{-1})
K_m = saturation or affinity constant (mg/l)
OUR = oxygen uptake rate ($mg\ O_2 \cdot l^{-1} \cdot min^{-1}$)
S = substrate concentration (mg/l)
t = time (min)
X = biomass concentration (mg/l)
X_r, X_s = hydrolyzable substrates (mg/l)
Y = yield coefficient
y, x_1, x_2 = model variables
z_1, z_2, z_3 = auxiliary variables

Greek letters

α = parameter combination
μ = specific growth rate (min^{-1})
μ_{max} = maximum specific growth rate (min^{-1})
θ = parameter combination

Index

ex = exogenous
$i(=1, 2)$ = index of pollutant substrate and related dynamics terms

1. INTRODUCTION

The identification of the dynamical models describing activated sludge processes is characterized by two important features:

1. The models are most often highly complex, they are usually high-order non linear systems incorporating a large number of state variables and parameters. For instance, the IAWQ activated sludge model No. 1 (Henze *et al.*, 1986) contains 13 state variables and 19 parameters.
2. There is, generally speaking, a lack of cheap and reliable sensors for on-line measurement of the key state variables, in particular those involved in the model. Despite considerable efforts, on-line sensor technology is still considered to be the weakest part in the real-time process control chain (Harremoës *et al.*, 1993; Vanrolleghem and Verstraete, 1993).

Both problems are common to all biotechnological processes, although particularly crucial in activated sludge processes, because of the inherent particularly complex nature of these processes, involving for instance many different microbial populations, and which, furthermore, are often difficult to reliably measure with the available instrumentation.

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. The central question of the identifiability analysis is the following:

> Assume that a certain number of the state variables are available for measurement; on the basis of the model structure (structural identifiability) or on the type and quality of available data (practical identifiability), can we expect to give via parameter estimation a unique value to the model parameters?

Simply speaking, one would wonder what is the use of trying to calibrate the parameters of a model which is, structurally or practically, unidentifiable. This above formulation is quite crude, but the answer to the identifiability analysis is often more subtle: it is not just a "yes or no" answer, but when it results in some conclusions (what is not a priori obvious with nonlinear models), these may tell that some subset or combinations of the model parameters are *a priori* identifiable.

The goal of our work is to study both the structural and practical identifiability of a class of models employing Monod type limitation kinetics used to describe activated sludge processes. In contrast to the studies found in the literature, the analysis will not start from the assumption that measurements of biomass and substrates are available but from the assumption that only oxygen uptake rate data are available. This choice is motivated by the context of the study: its main objective is the implementation of an on-line respirographic sensor that is capable of getting the best information possible on the basis of the available data and well-known biokinetic model structures for process monitoring and control via "in-sensor experiments" (Vanrolleghem, 1994). However our objective is to deal with the identifiability in a sufficiently general way so as to allow the extension of the proposed study to other practical situations.

The paper is organized as follows. The theoretical framework of the identification study will be addressed in Section 2, i.e. some important definitions are reviewed and basic concepts for the structural identifiability tests are introduced. In Section 3, the class of models studied in the paper and the considered assumptions are given. The structural identifiability of the models is studied in Section 4. Two different techniques (Taylor series expansion, and transformation of the nonlinear models) have been considered for the analysis of each candidate model, yet for limited space reasons, only the results of one method per model is presented. One objective is here to illustrate the possibilities and limitations of both identifiability tests.

In this paper, we shall mainly concentrate on the structural identifiability. A second paper (Vanrolleghem *et al.*, 1995) is intended to deal with optimal experimental design for parameter identification, to which the practical identifiability is more closely related: that's why our intention is to postpone the discussion about the practical identifiability in that paper. However, in order to clarify both concepts of structural and practical identifiability as best as possible, the latter will be illustrated in Section 2 with the Monod model.

2. THEORETICAL FRAMEWORK

The notion of structural identifiability is related to the possibility to give a unique value to each parameter of a mathematical model. In simple words, the structural identifiability of a model can be formulated as follows (a rigorous definition can be found e.g. in Godfrey and DiStefano III (1987)): given a model structure and perfect (i.e. that fit perfectly to the model) data of model variables, are all the parameters of the model identifiable? From the structural identifiability analysis one may conclude that only combinations of the model parameters are identifiable. If the number of resulting combinations is lower than the original model parameters, or if there is not a one-to-one relationship between both parameter sets, then *a priori* knowledge about some parameters may be required to achieve identifiability. A simple example may illustrate this: in the model $y = ax_1 + bx_2 + c(x_1 + x_2)$, only the parameters $a + c$ and $b + c$ are structurally identifiable (and not the three parameters a, b, c; two parameters (e.g. a and b) will be identifiable if the value of a third one (here, c) is known *a priori*.

For linear systems, the structural identifiability is rather well understood, and besides classical identifiable models (like dynamical models in canonical form), there exists a number of tests for the identifiability (e.g. Laplace transform method, Taylor series expansion of the observations, Markov parameter matrix approach, modal matrix approach, ..., see e.g. Godfrey and DiStefano III (1987)). However, for models that are nonlinear in the parameters (like the models used in this work), the problem is much more complex. There exist also several structural identifiability tests, but they are usually very complex [they typically require the (very helpful) use of symbolic software packages (Raksanyi *et al.*, 1985), as will be illustrated below]. In the following several similar approaches are used wherein the models are transformed into linear ones, after which the analysis is based on the linear model.

Practical identifiability on the other hand is related to the quality of the data and their "informative" content: are the available data informative enough for identifying the model parameters and for giving accurate values? In the model $y = ax_1 + bx_2$ the parameters are structurally identifiable but they will not be practically identifiable if the experimental conditions are such that x_1 and x_2 are proportional ($x_1 = \alpha x_2$) (then only the combination $a\alpha + b$ is identifiable). A typical example is the Monod model in simple microbial growth processes, which has been shown to be structurally identifiable from (perfect) data of substrate and biomass (Aborhey and Williamson, 1978), but which was found to be often not practically identifiable because of the usually poor quality of the data and the limited number of datapoints (e.g. Holmberg, 1982).

While the structural identifiability is studied under the assumption of perfect, i.e. noiseless, 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

predicted by the theoretical analysis, may no longer be guaranteed, because a change in one parameter can be compensated almost completely by a proportional shift in another, 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; Marsili-Libelli, 1992). Consequently, the estimated parameters may vary over a broad range and little physical interpretation can be given to the parameter values obtained.

The Monod-model [μ_{max} is the maximum specific growth rate (min^{-1}), K_m is the saturation constant (mg/l)],

$$\mu(S) = \frac{\mu_{max} S}{K_m + S} \tag{1}$$

is probably the best-known example in biological systems of a model in which parameter estimates may be highly correlated (Boyle and Berthouex, 1974; Holmberg, 1982; Munack, 1989). In many cases the experiments provide only sufficient information to estimate the ratio between both parameters in this model, μ_{max}/K_m. A simple example may illustrate this (Fig. 1): if only growth rates are measured for substrate concentrations ranging between 0 and 0.1 mg/l, no distinction can be made between different parameter sets, i.e. the Monod model is unidentifiable. In order to overcome this problem, it has been proposed to use additional a priori information (e.g. a known maximum growth rate) to impose parameter bounds (Holmberg, 1982), to sample more frequently in defined periods of the experiment in order to increase the informative content of the collected data (Vialas *et al.*, 1986), or to design experiments by which more informative data can be collected (Munack, 1989). Optimal experimental design for parameter estimation of the Monod model based on oxygen uptake rate data will also be the object of our following paper (Vanrolleghem *et al.*, 1995).

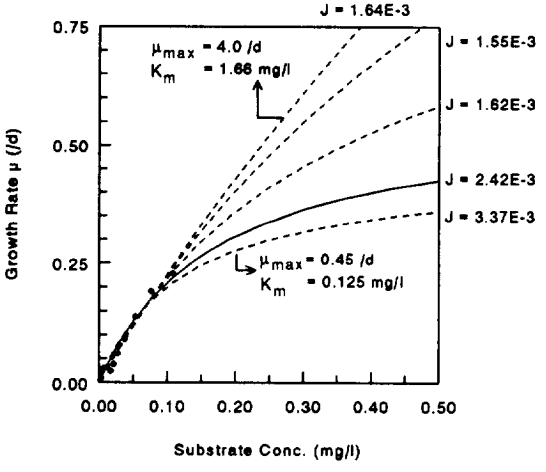

Fig. 1. Practical identifiability of the Monod model parameters.

3. 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

$$OUR_{ex} = -\sum_{i=1}^{k} (1 - Y_i) \frac{dS_i}{dt}. \tag{2}$$

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 ($k = 1$)

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

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

$$\frac{dS_1}{dt} = -\frac{\mu_{max1} X}{Y_1} \frac{S_1}{K_{m1} + S_1}. \tag{4}$$

Type 3 (Double Monod): Two pollutants simultaneously degraded without mutual interaction, double Monod kinetics ($k = 2$)

$$\frac{dS_1}{dt} = -\frac{\mu_{max1} X}{Y_1} \frac{S_1}{K_{m1} + S_1}. \tag{5}$$

$$\frac{dS_2}{dt} = -\frac{\mu_{max2} X}{Y_2} \frac{S_2}{K_{m2} + S_2}. \tag{6}$$

Type 4 (Modified IAWQ model No. 1) (Sollfrank and Gujer, 1991): 3 pollutants, 2 hydrolysed into the first substrate which is used for growth according to the Monod kinetics ($k = 1$)

$$\frac{dS_1}{dt} = -\frac{\mu_{max1} X}{Y_1} \frac{S_1}{K_{m1} + S_1} + k_r X_r + k_s X_s \tag{7}$$

$$\frac{dX_r}{dt} = -k_r X_r \tag{8}$$

$$\frac{dX_s}{dt} = -k_s X_s. \tag{9}$$

Nitrification and its associated oxygen consumption have not been included in the model. Because of the hydrolysis terms in (8, 9), OUR_{ex} (2) should be rewritten as follows

$$OUR_{ex} = -(1 - Y_1)\left(\frac{dS_1}{dt} - k_r X_r - k_s X_s\right). \tag{10}$$

The μ_{maxi} ($i = 1, 2$) and k_j ($j = r, s$) are rate constants, and the K_{mi} ($i = 1, 2$) are the so-called affinity constants expressing the dependency of the degradation rate on the concentration of pollutant S_i ($i = 1, 2$). Note that the models as given above can be used to describe batch experiments only.

As far as known by the authors, no studies of the structural 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 $S(t=0) \ll X(t=0)$];
- the oxygen uptake rate 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;
- the oxygen concentration is always maintained above 3 mg/l so that oxygen is not limiting.

4. STRUCTURAL IDENTIFIABILITY

In this section attention will focus on the structural identifiability of the 4 models introduced above.

The structural identifiability of the first model is rather straightforward. First note that the relation between OUR_{ex} and S_1 [$OUR_{ex} = -(1-Y_1)(dS_1/dt)$] can be also written after multiplication of both sides by $d\tau$ and integration between time $\tau = 0$ and time $\tau = t$

$$\int_0^t OUR_{ex}(\tau)\,d\tau = -(1-Y_1)[S_1(t) - S_1(0)] \quad (11)$$

which can be used to give an expression of $S_1(t)$

$$S_1(t) = S_1(0) - \frac{1}{(1-Y_1)}\int_0^t OUR_{ex}(\tau)\,d\tau. \quad (12)$$

Now let us define the variable $y(t)$

$$y(t) = \int_0^t OUR_{ex}(\tau)\,d\tau \quad (13)$$

which also means that $OUR_{ex} = dy/dt$. Then by combining the model equation (3) with the above expressions (12) and (13), one readily obtains the following dynamical linear equation for $y(t)$

$$\frac{dy}{dt} = \theta_1 y(t) + \theta_2 \quad (14)$$

with

$$\theta_1 = -\frac{\mu_{max1} X}{Y_1}, \quad \theta_2 = \frac{\mu_{max1} X}{Y_1}(1-Y_1)S_1(0). \quad (15)$$

Because $dy(t)/dt$ is measured and $y(t)$ is readily calculated from these data, the parameters θ_1 and θ_2 of the linear dynamical model (14) are (structurally) identifiable, i.e. it is theoretically possible to generate signals for $y(t)$ and $dy(t)/dt$ (e.g. via an impulse response) such that the parameters θ_1 and θ_2 can be uniquely identified. With respect to the original model parameters, and by considering the above definitions (15), this means that the model parameter combinations $\mu_{max1} X/Y_1$ and $(1-Y_1)S_1(0)$ are structurally identifiable.

Let us illustrate the structural identifiability concept via (real-life) data [Fig. 2(a)] with an initial substrate concentration $S_1(0)$. Figure 2(b) shows the data pairs [OUR_{ex}, $\int_0^t OUR_{ex}(\tau)\,d\tau$] corresponding to the [$OUR_{ex}(t)$, t] data presented in Fig. 2(a): θ_2 is given by the initial value of OUR_{ex}, and θ_1 is the slope.

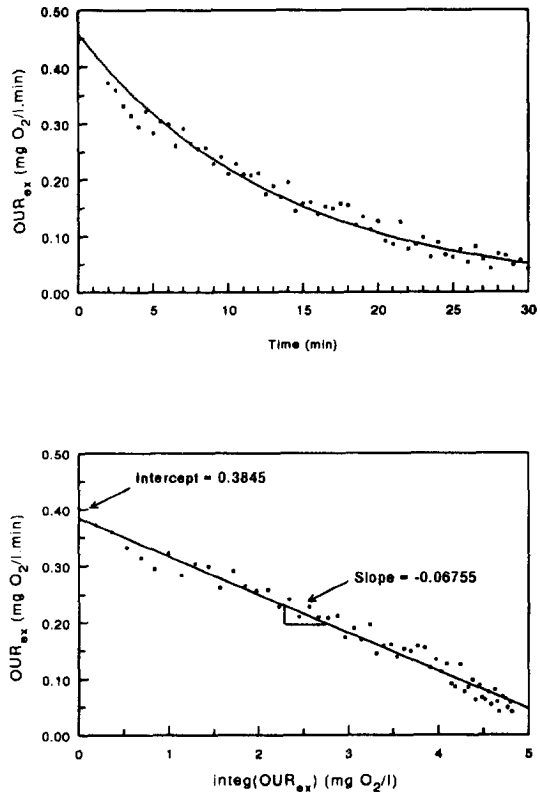

Fig. 2. Transformation of OUR data corresponding to the exponential model (a) into a linear regression form (b).

For the other three models, two approaches are presented:

1. The Taylor series expansion (Pohjanpalo, 1978),
2. The transformation of the models into models linear in the parameters (which is indeed the method used in the above exponential model).

4.1. *Taylor series expansion*

This method has been applied successfully for the first two models by a symbolic algebra software package (Mathematica, Wolfram Research Ltd). The complexity of the computer calculations for the identifiability analysis of the third model was such that the computer was not able to end the computation procedure and to give any conclusive results.

The method is based on Taylor series expansion of the observations [here, $OUR_{ex}(t)$] around time $t = 0$

$$OUR_{ex}(t) = OUR_{ex}(0) + t\frac{dOUR_{ex}}{dt}(0) + \frac{t^2}{2!}\frac{d^2OUR_{ex}}{dt^2}(0) + \cdots \quad (16)$$

and consists of looking at the successive derivatives to check if they contain information about the parameters to be identified. The approach is illustrated

with the second (Single Monod) model. Let us compute the successive derivatives: this will formally lead to an infinite set of terms. Let us consider the first terms

$$\mathrm{OUR}_{ex}(0) = \frac{\mu_{max1} X(1 - Y_1)}{Y_1} \frac{S_1(0)}{K_{m1} + S_1(0)} \quad (17)$$

$$\frac{d\mathrm{OUR}_{ex}}{dt}(0) = -\frac{\mu_{max1}^2 X^2(1 - Y_1)}{Y_1^2} \frac{K_{m1} S_1(0)}{[K_{m1} + S_1(0)]^3} \quad (18)$$

$$\frac{d^2\mathrm{OUR}_{ex}}{dt^2}(0) = \frac{\mu_{max1}^3 X^3(1 - Y_1)}{Y_1^3} \times \frac{K_{m1} S_1(0)[K_{m1} - 2S_1(0)]}{[K_{m1} + S_1(0)]^5} \quad (19)$$

$$\frac{d^3\mathrm{OUR}_{ex}}{dt^3}(0) = -\frac{\mu_{max1}^4 X^4(1 - Y_1)}{Y_1^4} \frac{K_{m1} S_1(0)}{[K_{m1} + S_1(0)]^7} \times [K_{m1}^2 - 8K_{m1} S_1(0) + 6S_1(0)^2] \quad (20)$$

$$\frac{d^4\mathrm{OUR}_{ex}}{dt^4}(0) = \frac{\mu_{max1}^5 X^5(1 - Y_1)}{Y_1^5} \frac{K_{m1} S_1(0)}{[K_{m1} + S_1(0)]^9} \times [K_{m1}^3 - 22K_{m1}^2 S_1(0) + 58K_{m1} S_1(0)^2 - 24S_1(0)^3]. \quad (21)$$

There are five parameters to be identified: Y_1, μ_{max1}, X, K_{m1} and $S_1(0)$. The key question is then the following: are they all structurally identifiable, or only combinations of them?

Let us first notice that the following parameter combinations

$$\theta_1 = \frac{\mu_{max1} X(1 - Y_1)}{Y_1}, \quad \theta_2 = (1 - Y_1)S_1(0), \quad \theta_3 = (1 - Y_1)K_{m1} \quad (22)$$

are combined in the first three derivatives. Indeed by noting

$$z_i = \frac{d^i\mathrm{OUR}_{ex}}{dt^i}(0), \quad i = 0, 1, 2, \ldots \quad (23)$$

equations (17), (18) and (19) can then be rewritten under the following (equivalent) form

$$z_0 = \frac{\theta_1 \theta_2}{\theta_2 + \theta_3} \quad (24)$$

$$z_1 = -\frac{\theta_1^2 \theta_2 \theta_3}{(\theta_2 + \theta_3)^3} \quad (25)$$

$$z_2 = \frac{\theta_1^3 \theta_2 \theta_3 (\theta_3 - 2\theta_2)}{(\theta_2 + \theta_3)^5}. \quad (26)$$

Therefore the "parameters" θ_1, θ_2 and θ_3 can be formally calculated from the values of z_i (which can be theoretically calculated from the [OUR(t), t] dataset) by inverting the above expressions, i.e.

$$\theta_1 = \frac{z_0(z_0 z_2 - 3z_1^2)}{z_0 z_2 - z_1^2} \quad (27)$$

$$\theta_2 = -\frac{2z_0^2 z_1}{z_0 z_2 - 3z_1^2} \quad (28)$$

$$\theta_3 = -\frac{4z_0^2 z_1^3}{(z_0 z_2 - 3z_1^2)(z_0 z_2 - z_1^2)}. \quad (29)$$

The question is then the following: can we expect to increase the number of identifiable parameters by considering higher order derivatives?

If we look at the additional derivatives for $i \geq 3$, the above parameter combinations are still combined basically in the same way as for the lower derivative terms, without any possibility to put in evidence other parameter combinations which could lead to a larger set of identifiable parameters. The conclusion is the same if we consider even higher order derivatives: only the above parameter combinations θ_1, θ_2 and θ_3 are structurally identifiable.

Besides the above parameter combination is basically the only one that fits in the above identifiability analysis (other combinations can also be considered (e.g. θ_1 and θ_2 as in (22), and $\theta_3 = (1 - Y_1)[K_{m1} + S_1(0)]$, but they are combinations of the above parameter combinations (22), and therefore basically the same as the one proposed). Note also that the next method (transformation of the nonlinear model) has also been applied to the Single Monod model and leads to the same conclusions.

Finally it is worth noting that the symbolic software has been used to compute the successive derivatives and, once a parameter combination has been chosen, to perform the subsequent computations [e.g. (24), (25), (26) and (27), (28), (29) here above].

4.2. *Transformation of the nonlinear models*

Another way to analyze the structural identifiability is to transform the nonlinear model into a model linear in the parameters, and then look at the identifiability of the linear model.

The line of reasoning is similar to the one considered for the exponential model hereabove [development (11)–(15)]. Let us see how it applies to the Double Monod model (5, 6). Recall that here the oxygen uptake rate OUR_{ex} is the sum of the contribution of two substrates S_1 and S_2

$$\mathrm{OUR}_{ex} = -(1 - Y_1)\frac{dS_1}{dt} - (1 - Y_2)\frac{dS_2}{dt}. \quad (30)$$

A typical OUR_{ex} profile is shown in Fig. 3. In the following, we assume (as it is suggested in Fig. 3) that one substrate (S_1) is completely eliminated from the mixed liquor after the first part of the experiment (note that there is only one pathological case when this assumption does not hold: when S_1 and S_2 are eliminated at exactly the same time). With this assumption the oxygen uptake rate can be subdivided in two parts corresponding with the degradation of each substrate. Hence, the identifiability analysis reduces to the analysis of the Single Monod model performed in two steps: for $0 \leq t < t_1$ for the first Monod model, and for $t_1 \leq t \leq t_2$ for the second one.

Table 1. Identifiable parameter combinations of the 4 models

Exponential	Single Monod	Double Monod	Modified IAWQ
$(1-Y_1)S_1(0)$	$(1-Y_1)S_1(0)$	$(1-Y_1)S_1(0)$	$(1-Y_1)S_1(0)$
$\frac{\mu_{max1}X(1-Y_1)}{Y_1}$	$\frac{\mu_{max1}X(1-Y_1)}{Y_1}$	$\frac{\mu_{max1}X(1-Y_1)}{Y_1}$	$\frac{\mu_{max1}X(1-Y_1)}{Y_1}$
	$(1-Y_1)K_{m1}$	$(1-Y_1)K_{m1}$	$(1-Y_1)K_{m1}$
		$(1-Y_2)S_2(0)$	$(1-Y_1)X_1(0)$
		$\frac{\mu_{max2}X(1-Y_2)}{Y_2}$	k_r
		$(1-Y_2)K_{m2}$	$(1-Y_1)X_3(0)$
			k_s

Let us first proceed for the first step and denote the first term of the right hand side of (30) by OUR_1

$$OUR_1 = -(1-Y_1)\frac{dS_1}{dt}. \tag{31}$$

The integration of the above equation gives

$$S_1(t) = S_1(0) - \frac{1}{1-Y_1}\int_0^t OUR_1(\tau)\,d\tau. \tag{32}$$

By introducing equation (5), the oxygen uptake rate equation (31) can be rewritten as follows

$$OUR_1 = \frac{1-Y_1}{Y_1}\frac{\mu_{max1}S_1}{K_{m1}+S_1}. \tag{33}$$

By multiplying both sides of the above equation by $(K_{m1}+S_1)$, and by considering (in order to have more compact notations) the variable $y_1(t)$ defined as follows

$$y_1(t) = \int_0^t OUR_1(\tau)\,d\tau \tag{34}$$

equation (33) after much manipulation becomes

$$y_1\frac{dy_1}{dt} = \alpha_1 - \alpha_2 y_1 + \alpha_3\frac{dy_1}{dt} \tag{35}$$

where the parameters α_i $(i=1,2,3)$ are indeed combinations of the parameters θ_i $(i=1,2,3)$ defined in (22)

$$\alpha_1 = \theta_1\theta_2, \quad \alpha_2 = \theta_2, \quad \alpha_3 = \theta_2+\theta_3. \tag{36}$$

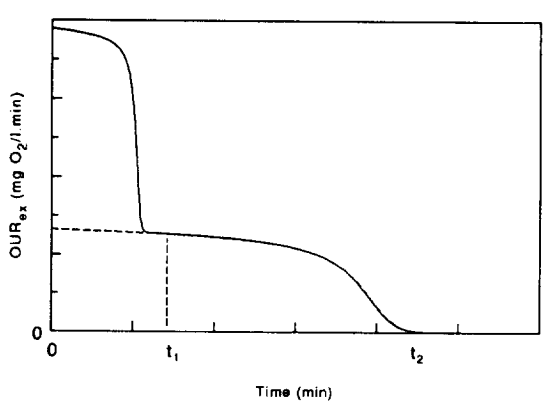

Fig. 3. Typical OUR profile with double Monod kinetics.

There is clearly a one-to-one relation between these two sets of parameters

$$\theta_1 = \frac{\alpha_1}{\alpha_2}, \quad \theta_2 = \alpha_2, \quad \theta_3 = \alpha_3 - \alpha_2. \tag{37}$$

Then we can conclude that with independent data of $y_1(dy_1/dt)$, y_1 and dy_1/dt (generated via an appropriate experiment design), the parameters $\alpha_1, \alpha_2, \alpha_3$, and therefore the parameters $\theta_1, \theta_2, \theta_3$, are identifiable. This result corresponds to the one obtained in the preceding section for the Single Monod model (for which we had used the Taylor series expansion approach).

We can proceed similarly for the second step $t_1 \leqslant t \leqslant t_2$, and using similar definitions for OUR_2 and $y_2(t)$, it is straightforward that the parameters

$$\theta_4 = \frac{\mu_{max2}X(1-Y_2)}{Y_2}, \quad \theta_5 = (1-Y_2)S_2(0),$$
$$\theta_6 = (1-Y_2)K_{m2} \tag{38}$$

are identifiable. This means that only six combinations of the nine original parameters [Y_1, $S_1(0)$, μ_{max1}, K_{m1}, Y_2, $S_2(0)$, μ_{max2}, K_{m2}, and X] are structurally identifiable.

A similar approach has been used for the IAWQ No. 1 model (which is summarized in the Appendix). The identifiability results for the four models are summarized in Table 1.

4.3. *Discussion*

The examples presented in the above Sections 4.1 and 4.2 are illustrative of the advantages and drawbacks of the considered methods. The implementation of the Taylor series expansion method has the advantage of being systematic in the sense that it follows a clearly identified route. The above example is illustrative of the potential difficulties with the Taylor series expansion:

- How many derivatives of OUR_{ex} are needed to obtain conclusive results? For certain models the question may arise whether one can achieve better identifiability properties by considering more terms in the expansion. Here we found that the additional evaluated terms did not yield additional information. Generally speaking the approach may imply

more and more symbolic computations, and yet not lead to conclusive results (as experienced for the Double Monod model).

- How can we derive the right combinations of identifiable parameters? There is indeed no general systematic rule for selecting these combinations, and therefore the procedure may look a little tricky. However the structure of the different terms of the expansion is often a source for good initial guesses. For instance, the choice of $\theta_1 = -[\mu_{max1} X(1 - Y_1)]/Y_1$ looks quite obvious from equations (17), (18) and (19).

On the other hand, the nonlinear transformation may suffer from the difficulty to easily find out the transformation that will *a priori* suit to the problem (although in the proposed example, the choice of the transformation (multiplication by the denominator of the Monod model) is rather straightforward).

Let us also make some comments about the obtained identifiability results.

First note that the yield coefficient(s) Y_1 (Y_2) is present in all the parameter combinations (except in k_r and k_s in the IAWQ model). This is not surprising since, just on the basis of OUR_{ex} data, one can have no idea which quantity of substrate has been transformed into biomass. This explains why the term $(1 - Y_1)S_1(0)[(1 - Y_2)S_2(0)]$ i.e. the fraction of substrate which is oxidized, appears as a parameter combination. The same remark applies to $(1 - Y_1)K_{m1}[(1 - Y_2)K_{m2}]$ which can be viewed as a rescaling of the substrate affinity constant.

Secondly the parameter combination

$$\frac{\mu_{max1} X(1 - Y_1)}{Y_1}\left(\frac{\mu_{max2} X(1 - Y_2)}{Y_2}\right)$$

is an expression of the total activity of the sludge, and with that respect can be considered as giving an information different from that of the individual parameters.

Finally let us point out that *a priori* information about some individual parameters (e.g. the yield coefficient(s) Y_1 (Y_2) values obtained via separate experiments) can be incorporated in the parameter evaluation procedure. Then individual parameters [e.g. $S_1(0)$ or K_{m1}] can be estimated.

5. CONCLUSIONS

This paper has dealt with the identifiability of parameters of kinetic models describing the activated sludge process. The main concern of the paper was to present important aspects of the structural identifiability properties. The identifiability analysis is based on the availability of on-line oxygen uptake rate data only (given by a novel respirographic biosensor). Four model candidates (exponential, Monod, double Monod and IAWQ No. 1) are considered. For each model we found that only a smaller set of the combinations of the original parameters are structurally identifiable on the basis of oxygen uptake rate data only. The yield coefficient is present in almost all parameter combinations. Separate experiments can provide the necessary information to obtain the individual parameters.

Acknowledgements—This paper presents research results of the Belgian Programme on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science, Technology and Culture. The scientific responsibility rests with its authors. The authors would also like to express their gratitude to one of the anonymous reviewers for his very valuable comments.

REFERENCES

Aborhey S. and Williamson D. (1978) State and parameter estimation of microbial growth processes. *Automatica* **14,** 493–498.

Boyle W. and Berthouex P. (1974) Biological wastewater treatment model building: fits and misfits. *Biotechnol. Bioengng* **16,** 1139–1159.

Godfrey K. and Di Stefano III J. (1987) Identifiability of model parameters. In *Identifiability of Parametric Models* (Edited by Walter E.), pp. 1–19. Pergamon Press, Oxford.

Harremoës P., Capodaglio A., Hellström B., Henze M., Jensen K., Lynggaard-Jensen A., Otterpohl R. and Soeberg H. (1993) Wastewater treatment plants under transient loading—performance, modelling and control. *Wat. Sci. Technol.* **27,** (12), 71–115.

Henze M., Grady C. P. L. Jr, Gujer W., Marais G. and Matsuo T. (1986) Activated sludge model No. 1. IAWPRC, Scient. Tech. Rep. No. 1.

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

Marsili-Libelli S. (1992) Parameter estimation of ecological models. *Ecol. Modell.* **62,** 233–258.

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

Pohjanpalo H. (1978). System identifiability based on the power series expansion of the solution. *Math. Biosci.* **41,** 21–33.

Raksanyi A., Lecourtier Y., Walter E. and Venot A. (1985) Identifiability and distinguishability testing via computer algebra. *Math. Biosci.*, **77,** 245–266.

Sollfrank U. and Gujer W. (1991) Characterisation of domestic wastewater for mathematical modelling of the activated sludge process. *Wat. Sci. Technol.* **23,** (4–6), 1057–1066.

Vanrolleghem P. (1994) On-line modelling of activated sludge processes: development of an adaptive sensor. Ph.D. thesis, Universiteit Gent, Belgium.

Vanrolleghem P., Van Daele M. and Dochain D. (1995) Practical identifiability of a biokinetic model of activated sludge respiration. *Wat. Res.* **29,** 2561–2570.

Vanrolleghem P. and Verstraete W. (1993) On-line monitoring equipment for wastewater treatment processes: State of the art. *Proc. TI-KVIV Studiedag Optimalisatie van Waterzuiveringsinstallaties door Proceskontrole en -sturing*, pp. 1–22.

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

APPENDIX

The analysis carried out for the modified IAWQ model is based on the model equations (7), (8), (9) and (10). As in Section 4.2, the first step in the analysis consists of considering that during a part of the experiment, the concentration of the rapidly hydrolysable substrate X_r should be approximately zero.

The effects of the two substrates cannot be decoupled, unlike in the Double Monod model where this was possible due to the saturation in the kinetics of S_2. Therefore we start the analysis for time $t_1 \leqslant t \leqslant t_2$ when $X_r = 0$: this means that in the first step only equations (7), (9) and (10) with $X_r = 0$ are considered. Note first that by introducing equation (7), the OUR_{ex} equation (10) can be rewritten as follows

$$OUR_{ex} = \frac{\mu_{max1} X (1 - Y_1)}{Y_1} \frac{S_1}{K_{m1} + S_1}. \tag{39}$$

Then the analysis is performed as follows:

1. Integration of equations (10) and (9). This gives the following relation for $S_1(t)$

$$S_1(t) = S_1(0) - \frac{1}{1 - Y_1} \int_0^t OUR_{ex}(\tau)\, d\tau + \int_0^t k_s X_s(\tau)\, d\tau \tag{40}$$

$$= S_1(0) - \frac{1}{1 - Y_1} \int_0^t OUR_{ex}(\tau)\, d\tau + k_s X_s(0)(1 - e^{-k_s t}). \tag{41}$$

2. Linearization of the exponential term $e^{-k_s t}$ around $t = 0$ (in order to carry out the analysis with a model linear in the parameters)

$$e^{-k_s t} \cong 1 - k_s t + \frac{k_s^2 t^2}{2} \tag{42}$$

(we stop the series expansion at the second order term since the additional terms do not add extra useful information for the analysis).

3. Introduction of these results in equation (39) and rewriting by multiplying both sides by

$$(1 - Y_1)K_{m1} + (1 - Y_1)S_1(0) - \int_0^t OUR_{ex}(\tau)\, d\tau + (1 - Y_1)k_s X_s(0)\left(k_s t - \frac{k_s^2 t^2}{2}\right)$$

and considering $y(t)$ defined in (13)

$$y \frac{dy}{dt} = \beta_1 + \beta_2 y + \beta_3 \frac{dy}{dt} + \beta_4 t \frac{dy}{dt} - \beta_5 \frac{t^2}{2} \frac{dy}{dt} - \beta_6 t + \beta_7 \frac{t^2}{2} \tag{43}$$

$$\beta_1 = -\theta_1 \theta_2, \quad \beta_2 = \theta_1, \quad \beta_3 = \theta_2 + \theta_3 \tag{44}$$

$$\beta_4 = \theta_7 \theta_8^2, \quad \beta_5 = \theta_7 \theta_8^3, \quad \beta_6 = \theta_1 \beta_4, \quad \beta_7 = \theta_1 \beta_5 \tag{45}$$

with $\theta_1, \theta_2, \theta_3$ as defined in (22), and θ_7 and θ_8 defined as follows

$$\theta_7 = (1 - Y_1)X_s(0), \quad \theta_8 = k_s. \tag{46}$$

Among the seven parameters β_i, only five are independent (β_6 and β_7 are related to β_4, β_5, and β_2). Therefore five parameter combinations are identifiable, i.e. $\beta_1, \beta_2, \beta_3, \beta_4$ and β_5, or equivalently $\theta_1, \theta_2, \theta_3, \theta_7$ and θ_8.

The second step for $0 \leqslant t \leqslant t_1$ considers the dynamics are given by equations (7), (8), (9) and (10) with $X_r \neq 0$, and that the values of the parameters β_i ($i = 1$–7) are given from the first step, i.e. data for times between t_1 and t_2. Then by following the same line of reasoning as in step 1, one obtains two more identifiable parameter combinations: $(1 - Y_1)X_r(0)$ and k_r.