Pergamon

*Wat. Sci. Tech.* Vol. 33, No. 2, pp. 91–105, 1996.

0273–1223/96 $15·00 + 0·00

PII: S0273-1223(96)00192-8

# IDENTIFICATION OF BIODEGRADATION MODELS UNDER MODEL AND DATA UNCERTAINTY

Peter A. Vanrolleghem* and Karel J. Keesman**

** Department of Applied Mathematics, Biometrics and Process Control, University of Gent, Coupure Links 653, B-9000 Gent, Belgium*
*** Wageningen Agricultural University, Department of Agricultural Engineering and Physics, Bomenweg 4, NL-6703 HD Wageningen, The Netherlands*

## ABSTRACT

In this paper a number of nonlinear parameter estimation methods are evaluated with respect to their ability to identify biodegradation models from "real-world" data. Important aspects are then the sensitivity to local minima, rate of convergence, required prior knowledge and direct or indirect availability of parameter estimates uncertainty. Furthermore, it is important whether a method is robust against invalid assumptions. In addition to the final parameter values, covariance and correlation matrices, confidence intervals and residual sequences are presented to obtain information about the validity of the models and noise assumptions. Finally, recommendations on the method's applicability range are provided. Copyright © 1996 IAWQ. Published by Elsevier Science Ltd

## KEYWORDS

Activated Sludge; identification; mathematical modelling; numerical methods.

## INTRODUCTION

Wastewater treatment process models have become a major tool to increase the understanding of the underlying biodegradation mechanisms, to communicate using it as a common ground, in the design of treatment plants, control and operating strategies, for training of operators and process engineers, and so forth. In these examples the mathematical model can only be applied successfully if it is a proper description, in terms of both model structure and model parameters, of the underlying process. Hence, theoretical modelling is most often not enough. There is also a need for identification of the model from experimental data.

However, in practice the identification exercises are often hampered by small data sets, questionable normality assumptions, possible heteroscedascity of the error variance, the occurrence of outliers and the effects of the nonlinearity of the models describing a biodegradation process. In many cases standard non-recursive parameter estimation tools are utilized to accomplish the identification of wastewater treatment models (Coté *et al.*, 1995; Reichert *et al.*, 1995; Schmidt & Isaacs, 1995). However, these methods rely on certain assumptions that are often violated by the specific data used. While methods have been proposed that impose less stringent requirements, application of these new methods in the area of wastewater treatment

modelling has been rare. The intention of this paper is to evaluate some robust identification methods and provide recommendations on their applicability range. Special attention will be focused on the application of these methods for fast tracking of transient events via on-line model identification.

Strikingly, uncertainty analysis has, in contrast to sewer system and water quality modelling, drawn little attention so far in the area of wastewater treatment systems (Beck and Chen, 1994). It must be appreciated, however, that input, model and parameter uncertainty are ubiquitous for these systems. Some of the issues of parameter uncertainty will be addressed.

The application used in the study is the identification of biodegradation models on the basis of respirometric data. Recently, a clear increase can be observed in the efforts directed at the use of (on-line) respirometers for monitoring, modelling and control of wastewater treatment plants (Spanjers & Keesman, 1994, Vanrolleghem *et al.*, 1995). Hence, the identification tools used will also partly determine the quality of the results obtained in such studies.

## IDENTIFICATION OF BIODEGRADATION MODELS

### Model structures

The general model structure for biodegradation as the result of the addition of k pollutants in an activated sludge process in a batch reactor and assuming Monod kinetics is given by:

$$\frac{dS_i(t)}{dt} = - r_{Si}(t) \tag{1}$$

$$r_{Si}(t) = \frac{\mu_{max,i} X}{Y_i} \frac{S_i(t)}{K_{m,i} + S_i(t)} \tag{2}$$

$$OUR_{ex}(t) = \sum_{i=1}^{k} (1 - Y_i) r_{Si}(t) \tag{3}$$

where it is assumed that X, the biomass concentration remains constant during the batch experiment due to the low S/X ratio. Notice that in the output equation (3) the effect of the endogenous respiration rates has already been subtracted from the oxygen uptake rate. Hence, $OUR_{ex}$ only represents the exogenous or substrate induced oxygen uptake rate. In the given model structure there are a number of unknown parameters such as $Y_i$, $_{max,i}$, X, $K_{m,i}$ and the initial conditions $S_i(0)$. From a theoretical identifiability study (also indicated as structural or *a priori* identifiability) one can deduce that for a single Monod model only three parameter combinations can be identified from given $OUR_{ex}$ data (Dochain *et al.*, 1995):

$$\theta_1 = \frac{\mu_{max,1} X (1 - Y_1)}{Y_1} \qquad \theta_2 = (1 - Y_1) S_1(0) \qquad \theta_3 = (1 - Y_1) K_{m1} \tag{4}$$

Unlike the identification of biodegradation parameters from batch experiments, the parameters can also be estimated from CSTR experiments with continuous input of biodegradable matter. Eqn (1) for one pollutant is then expanded to,

$$\frac{dS(t)}{dt} = - r_S(t) + \frac{Q(t)}{V} (S_{in}(t) - S(t)) \tag{5}$$

In one of our applications we model the degradation of readily biodegradable matter in industrial influent in an on-line respiration meter (see Spanjers and Klapwijk, 1990). Then, Eqn (5) is rewritten as:

$$\frac{dS_S(t)}{dt} = - \left[ \frac{Q(t)}{V} + \frac{r_{max} X(t)}{K_S + S_S(t)} \right] S_S(t) + \frac{Q(t)}{V} (1 - Y) S_{in}(t) \tag{6}$$

where $S_S$ is the substrate concentration expressed as short-term BOD in mg $O_2$/l, i.e. $S_S$=(1-Y)S, and $S_{in}$ is the substrate concentration of the influent in mg COD/l. Furthermore, the actual respiration rate, expressed in mg $O_2$/l.h, is

$$r_{act}(t) = r_{max}\frac{S_S(t)}{K_S + S_S(t)}X(t) + bX(t) \tag{7}$$

the unknown parameters are now: maximum specific oxygen consumption rate $r_{max}$, Monod constant for substrate ($K_S$), specific decay rate (b) and yield coefficient (Y).

For the identification from discrete-time measurements it is significant to express the finite-dimensional, continuous-discrete time systems in a more or less general nonlinear state-space model structure which includes a measurement equation with output-error term:

$$\mathbf{dx(t,\theta)/dt = f(x(t,\theta), u(t), t; \theta)} \tag{8a}$$

$$\mathbf{y(t,\theta) = g(x(t,\theta), u(t), t; \theta)} \tag{8b}$$

$$\mathbf{z(t_k) = h(x(t,\theta), u(t), t_k; \theta) + e(t_k)} \quad \text{for } \mathbf{k=1,....,N} \tag{8c}$$

in which we distinguish a state, output and discrete-time measurement equation. In what follows, we assume that $\theta \in \mathbf{D_M} \subset R^m$, which implies that the parameter vector belongs to a prespecified set $\mathbf{D_M}$ in the parameter space $R^m$.

It should be realized that the output-error e contains errors from both the modelling and the measurement process. In the next section, nonlinear parameter estimation methods for the identification of the unknown parameters from given data will be presented. However, since both the model structures and the measurements contain errors, it is of paramount importance to have at least an indication of the uncertainties in the estimates as a result of uncertainty propagation through the parameter estimation methods. Therefore, also attention will be given to the processing of estimate uncertainties and the prior information needed for this.

Nonlinear parameter estimation (NPE)

In essence, a parameter estimation method is a mapping from a given data set $\mathbf{Z^N}$ to the set $\mathbf{D_M}$:

$$\mathbf{Z^N \rightarrow \theta_N \in D_M} \tag{9}$$

(see, for instance, Ljung, 1987). Before being able to further specify this mapping, assumptions about the error characteristics must be made. Conventionally, one assumes that *e* is a serially uncorrelated sequence with zero mean and constant variance, also indicated as the statistical approach. We will first present some methods which are essentially based on this kind of assumption. Then, a natural way of parameter estimation is to choose $\theta_N$ such that there is a small misfit between data and model output. Formally speaking, and selecting a 2-norm that measures the misfit, the estimate can be found from,

$$\mathbf{\theta_N = \arg\min_{\theta \in D_M} \|\varepsilon(\theta)\|_2^2} \tag{10}$$

where the residual vector contains elements $\varepsilon(\theta,t_k)=z(t_k)-z(t_k|\theta)$ and the criterion function $J(\theta,Z^N)=\|\varepsilon(\theta)\|_2^2 = \sum_k \varepsilon(\theta,t_k)^2$. The predicted systems response, given parameter vector q and input/output data, takes the general form

$$\mathbf{z(t_k|\theta) = \phi(t_k, Z^{tk-1}; \theta)} \tag{11}$$

in which $\mathbf{Z^{tk-1}}$ denotes the set of input and output measurements available at $t_k$. The function $\imath(.,.,;.)$ can be viewed as the result of a black-box or mechanistic-based modelling procedure with or without filtering; in the latter case of straightforward simulation from time instant $t_0$ $Z^{tk-1}$ contains only available input measure-

ments. Clearly, in the case studied here, the prediction $z(t_k|\theta)$ can be constructed by evaluating the model Eqns (8a-b) up to $t_k$ for given (measured) input u(t) and parameter vector $\theta$. Hence, (8) can be written as a nonlinear regression model, relating the measured output to the parameter vector

$$z(t_k) = \phi(t_k, Z^{tk-1}; \theta) + e(t_k) \tag{12}$$

Herein, $e(t_k)$ can be considered now as the multi-step-ahead prediction error. This regression model forms then the starting point for our further discussions on parameter estimation methods.

Newton-like method

Many methods for unconstrained minimization, i.e. $D_M=R^m$, but also for constrained optimization problems as e.g. the SQP (Sequential Quadratic Programming) method, are derived from a quadratic model of the criterion function. In Fletcher (1987) a large number of practical algorithms based on this quadratic model have been described and discussed. A quadratic model can be obtained from a truncated Taylor series expansion of $J(\theta)=J(\theta,Z^N)$ about $\theta^{(i)}$, the estimate for iteration i,

$$J(\theta^{(i)}+\delta) \approx J^{(i)} + \gamma^{(i)}\delta + \tfrac{1}{2}\delta^T \Gamma^{(i)} \delta \tag{13}$$

where $\delta=\theta-\theta^{(i)}$ and $J^{(i)}$, $\gamma^{(i)}$, $G^{(i)}$ are zero, first and second derivatives of $J(\theta)$. Define the Jacobian matrix: $A(\theta) = [\nabla\varepsilon_1, \nabla\varepsilon_2, ...., \nabla\varepsilon_m]$, the columns of which are the first derivative vectors of $\mathfrak{h}(\theta)$ and
.Since $\Gamma(\theta) = 2AA^T + 2\ddot{}\varepsilon_k \nabla^2\varepsilon_k$ with $\varepsilon_k=\varepsilon(t_k,\theta)$ and $\varepsilon_k$ is small for a least-squares estimate of $\theta$, a good approximation of ˇ might be given by

$$\Gamma(\theta) \approx 2AA^T \tag{14}$$

This approximation is essential for the class of Gauss-Newton methods, so that

$$\theta^{(i+1)} = \theta^{(i)} + [A^{(i)} A^{(i)T}]^{-1} A^{(i)} \epsilon(\theta^{(i)}) \tag{15}$$

To avoid a number of problems encountered in the Gauss-Newton methods and to ensure progress in reducing $J(\theta)$ an alternative class of methods based on a line search procedure has been developed, where the direction of search is given by

$$s^{(i)} = -\Gamma(i)-1\, \gamma^{(i)} \tag{16}$$

that is the local downhill gradient direction, so that $\theta^{(i+1)} = \theta^{(i)} - \alpha s^{(i)}$ (17)

Methods which include a line search and a direct approximation of $\Gamma^{-1}$ are called quasi-Newton methods. The well-known BFGS (Broyden, Fletcher, Goldfarb and Shanno) method (see Fletcher, 1987) is a member of this class, and it approximates the inverse of $\Gamma(i)$ by a symmetric positive definite matrix $H^{(i)}$ (the so-called Hessian), which is updated from iteration to iteration. Once the matrix $\Gamma(\theta)$ or $H(\theta)$ is known, it is relatively easy to calculate the covariance matrix of the estimates $\Sigma(\theta)$ from these. That is, $\Sigma(\theta) = 2\sigma^2 [\Gamma(\theta)]^{-1}$ or $\Sigma(\theta) = 2\sigma^2 [H(\theta)]^{-1}$.

Direction set method

An interesting nonlinear least-squares algorithm which avoids computing the first derivatives is Brent's method (Brent, 1973). This method basically contains a finite difference approach of the gradients. Furthermore, it is an example of a direction set method in which a set of independent directions $s^{(1)}, s^{(2)}, ...., s^{(m)}$ is maintained, and successive line searches are taken along $s^{(j)}$ in a cyclical manner. The directions are

chosen to be conjugate to make the search as efficient as possible. In the absence of any explicit derivative information the parallel subspace property is applied to generate conjugate directions. The set of conjugate directions are the principal directions of the Hessian matrix H. This set of directions can be found without an explicit evaluation of H because the terms $s^{(i)T} H s^{(i)}$ are available from the line search along $s^{(i)}$. Essentially their calculation involves an eigenvalue decomposition. Subsequently, the Hessian obtained after convergence can be used to calculate the covariance matrix as given above.

Simplex method

Another no-derivative method is the simplex search method of Nelder and Mead (1964). A (regular) simplex is a set of m+1 (equidistant) points in $R^m$, also called vertices. This method for minimizing a function of m variables depends then on the comparison of function values at the (m+1) vertices of a general simplex, followed by the replacement of the vertex with the highest value by another point. This new point is found by reflection, contraction or expansion with respect to the previous simplex. As in all NPE methods, the search is stopped when a certain stopping criterion is satisfied. Since there are no gradients of the criterion function with respect to the parameters involved in the computation, this search method is rather easy to implement. And, because in practice derivatives are not conveniently available most of the time, it is also very popular. In order to obtain the estimate uncertainties the Hessian matrix of second derivatives at the minimum is approximated numerically by fitting a quadratic surface through a number of well-chosen points.

Bootstrap and Monte Carlo methods

If we are really interested in an interval estimate of $\theta$ associated with a prescribed probability, then knowledge about the distribution of the parameter estimates must be available. In practice, most often a normal distribution is assumed, but this assumption is only valid if the errors $e(t_k)$ for k=1,...,N originate from a stochastic process with normal distribution, the model is exact and it is linear in the parameters. Clearly, under the normality assumption the parameter estimation error covariance matrices obtained from the preceding methods and resulting uncertainty intervals only give a rough approximation of the uncertainty in the estimates. The idea of using an empirical distribution of the parameters is therefore attractive. In this study, two approaches were used to estimate such a distribution.

One is based on an extensive number of Monte Carlo simulations. The procedure is to draw random numbers from an appropriately chosen distribution (typically a uniform distribution) around the best parameter estimates and evaluate the objective function for each of these new parameter vectors. Rather than present all details of the obtained probability distributions, one normally summarizes the results in the form of confidence regions. Such regions can be described as a hypervolume bound by a hypersurface in the parameter space. According to Beale (1960), the $(1-^{-})$ confidence region corresponds with the set of parameter vectors for which the objective function MSE (mean square error, $J_{min}/(N-m)$) is below the threshold given by

$$\mathbf{MSE * (1 + m/(N-m) * F_{\alpha,m,N-m})} \tag{18}$$

in which F is the value of the F-distribution with m and N-m degrees of freedom at a confidence level $\alpha$. It is evident that the approach is very computer intensive because the number of evaluations increases as a power function of m (Lobry & Flandrois, 1991).

In the second approach an empirical distribution is obtained from resampling methods such as the bootstrap (Roy, 1994). The resampling strategy applied here is the regular or exact jackknife procedure (Fox *et al.*, 1980). The procedure is as follows:

1) Estimate the parameter vector $\theta_0$ using all available measurement data
2) Perform N estimations of the parameter vector $\theta_{-i}$ using all data except the ith data point
3) Calculate the pseudo-values as the vectors satisfying: $_i = N\theta_0-(N-1)\ \theta_{-i}$

With these N pseudo-values the jackknife estimate of the parameter set and its covariance matrix can be calculated:

$$\begin{aligned} \boldsymbol{\theta} &= \Sigma_i\, \boldsymbol{\theta}_i /N \\ \Sigma(\boldsymbol{\theta}) &= \Sigma_i\, \Sigma_j\, (\boldsymbol{\theta}_i - \boldsymbol{\theta})(\boldsymbol{\theta}_j - \boldsymbol{\theta})/N/(N-1) \end{aligned} \tag{19}$$

With these two methods, empirical distributions of the parameter estimates are obtained that are insensitive to both the error distribution specification and linearizing approximations.

Maximum likelihood method

Recall that the basis equation for all these optimization methods is the nonlinear regression model (12). For the model structures Eqn. (1)-(3) or (5)-(6) the predictor ($\eta$) can simply be evaluated by simulation for given parameter vector $\theta$. However, this implicitly assumes that the biodegradation model structures are exact, since S(t) is reconstructed as a function of $\theta$. Conversely, for the single Monod model (1)-(3) with i restricted to 1, S(t) could also have been estimated from measured OUR or $r_{act}$ data. This approach has been followed by Spanjers and Keesman (1994) for the selection of a wastewater biodegradation kinetics model. Since S(t) contains errors, due to both measurement and modelling uncertainty, the problem turns into an "error-in-variables" problem. Hence, two kinds of residuals can be distinguished, which are adjoined in one residual vector,

$$\begin{aligned} \epsilon(\boldsymbol{\theta}, S^*(t_k)) &= [\epsilon_{S,k} \quad \epsilon_{r,k}]^T \\ &= [S(t_k) - S^*(t_k) \quad r_S(t_k) - f(S^*(t_k), \boldsymbol{\theta})]^T \end{aligned} \tag{20}$$

where $S^*(t_k)$ denotes the true value at time instant $t_k$. Consequently, on the basis of maximum likelihood estimation theory (Bard, 1974), the parameter estimation problem can be formulated as,

$$[\boldsymbol{\theta} \;\; \mathbf{S}]^T = \arg\min \Sigma_k\, \epsilon(\boldsymbol{\theta}, \mathbf{S}(t_k))^T\, \mathbf{V}(t_k)^{-1}\, \epsilon(\boldsymbol{\theta}, \mathbf{S}(t_k)) \tag{21}$$

where $\mathbf{S} = [S(t_1), S(t_2), ...., S(t_{N-1})]^T$ and $V(t_k)$ is the covariance matrix of the residuals at $t_k$, which basically depends on the integration time interval. The uncertainty in the parameter estimates can be estimated from

$$\Sigma(\boldsymbol{\theta}) \simeq [\; \Sigma_k \;\; \mathbf{B}(t_k)^T\, \mathbf{V}(t_k)^{-1}\, \mathbf{B}(t_k) \;]^{-1} \tag{22}$$

where the elements of the matrix B are defined as:

Set-membership method

So far, it has been assumed that the errors are of random nature, and thus can be described in terms of statistical properties as mean, variance and probability density functions. If, however, such detailed characterization of the errors is not possible due to a limited length of the data records, or if the error has non-random components as a result of model inadequacy or systematic measurement errors, a statistical approach will give unreliable results. Under these circumstances, a deterministic error characterization in terms of lower and upper bounds only will be a good alternative. Overviews of (nonlinear) time-domain estimation methods within the so-called set-membership or bounded-error context can be found in Norton (1987), Walter *et al.* (1991) and Milanese and Vicino (1991). The basic assumption for symmetric bounds within this bounded-error approach is that

$$\|\mathbf{e}\|_\infty = \max_k\, |\mathbf{e}(t_k)| \leq \eta \tag{23}$$

It should be noted that symmetry is usually assumed due to lack of more detailed information. On the basis of this error interval model we define the behaviour set as:

$$\Omega_z = \cup_{k=1}^{N} \{ z \in \mathbb{R}^m : z = z(t_k) - e(t_k);\ e(t_k) \in \Omega_e(t_k) \} \quad (24)$$

where $\Omega_e(t_k)$, the error set, contains all errors which fulfil the condition of Eqn. (23). Thus, the effect of model uncertainty is directly taken into account in the error term. In addition to this behaviour set we define the prior parameter set $D_M$, which contains a priori parametric information from, for instance, preceding estimation procedures or literature. The ultimate aim is then to identify the set of parameter vectors $\Omega_\theta$, which is consistent with model, behaviour set and prior parameter set. This set with feasible parameter vectors, which explicitly represents the parametric uncertainty, is called the posterior or feasible parameter set.

$$\Omega_\theta = \{\theta \in \mathbb{R}^m : z(t_k) - h[x(t,\theta),\mu(t),t_k;\theta] \in \Omega_e(t_k);\ \theta \in \Omega_\theta,\ k = 1,...,N\} \quad (25)$$

The MCSM (Monte Carlo Set-Membership) algorithm (Keesman, 1990; Keesman and van Straten, 1990) can handle this nonlinear parameter estimation problem. The key idea behind this algorithm is that randomly selected parameter vectors which result in a model response consistent with the behaviour set belong to the feasible set. By choosing an appropriate error bound, the feasible parameter set can be reduced to a singleton. This so-called min-max estimate is found from:

$$\theta_N = \arg\min_{\theta \in D_M} \|\epsilon(\theta)\|_\infty \quad (26)$$

Model quality assessment by diagnostic checking of residuals

To evaluate the model quality, the assumed properties of the noise have been compared with those of the calculated residual sequences. Two tests were applied to test whether the residuals could be assumed to be a realization of independent random variables with zero mean. The autocorrelation test (Söderström & Stoica, 1989) evaluates whether the covariance function for a white noise sequence $\varepsilon_k$:

$$r_\epsilon(\tau) = \sum_{k=1}^{N-\tau} \epsilon(k-\tau)\epsilon(k) \quad (27)$$

is zero except for $\tau$=0. To make objective decisions on the whiteness of the residuals, one compares the covariance for each lag with the limit value $N(0,1)/\sqrt{N}$. For =0.05, five percent of the covariances may be larger than $1.96/\sqrt{N}$. Another residuals test applied was the runs test (Söderström & Stoica, 1989). It is a non-parametric test in which the number of sign changes $R$ in the residuals sequence is compared with the expected number of runs, $E(R)=N/2$.

## RESULTS

### General

The parameter estimation methods were applied to identify corresponding models (Eqn 1-3 or Eqn 6-7) to $OUR_{ex}$ and $r_{act}$ data sets (the maximum likelihood method was only applied to the single Monod data). As Figures 1, 2 and 3 illustrate the fit of the models as identified by the different methods are nearly identical. Before focusing on the parameter values, the model adequacies are checked. The figures show that the residual sequences have rather constant variance but can hardly be considered to correspond to a white noise sequence. The autocorrelation function in Figure 1 confirms this for the Single Monod data set as more than five percent of the correlations exceeds the critical value of 0.134. Similar results were obtained for the other data-sets (results not shown). The results of the runs test (Tables 1, 4 and 7) also deviate significantly from the expected value. Hence, information is still left in the residual sequence. From this point one can

choose either adjustment of the system dynamics or modelling of the noise, but these steps are out of the scope of this paper.

The parameter sets obtained with the different methods are summarized in Tables 1, 4 and 7 for the respective models, data-sets and NPE methods. For the Monte Carlo method only a confidence region is obtained and not an unique "best" parameter set. In case of the Set-Membership method, one observes from (26) that the min-max estimate is in fact the parameter vector for which the largest residual is minimized, i.e. a minmax solution is pursued instead of a minimal MSE as in (10). For the Single Monod and Double Monod $OUR_{ex}$ data-set, these min-max error bounds are 0.0383 and 0.0476 mg $O_2$/l.min. In case of the CSTR $r_{act}$ data, the min-max error bound is 18.973 mg $O_2$/l.h. One observes that these bounds correspond with the extreme values found in the residuals sequences of Fig. 1-3.

The $r_{act}$ data-set proved, however, to be insufficiently informative to allow simultaneous estimation of the four parameters of the model described by Eqn (6-7). It appears that only the combination of $r_{max}$ and $K_s$ can be identified. This is also apparent from the nearly constant ratio $r_{max}/K_s$=15 found in Table 7. Consequently, the correlations between the other parameters in Table 8 cannot be considered as accurate. A particularity of our study was that for the Brent, Simplex and Jackknife methods bounds were imposed on the $r_{max}$ parameter based on physical reasoning. The upper bound was set at 200 mg $O_2$/mg.h. The consequences of this choice will be discussed below.

Table 1. NPE results for single Monod model to $OUR_{ex}$ data-set (N=284, m=3)

| | $\mu_{max1}$ (/d) | $K_{s1}$ (mg/l) | $S_1(0)$ (mg/l) | MSE (*$10^6$) | Runs |
|---|---|---|---|---|---|
| BFGS | 0.409 | 0.649 | 17.017 | 196.907 | 24 |
| Brent | 0.405 | 0.574 | 17.281 | 115.141 | 28 |
| Simplex | 0.413 | 0.674 | 17.394 | 124.587 | 24 |
| Jackknife | 0.423 | 0.729 | 17.514 | 168.211 | 24 |
| Set-Memb | 0.401 | 0.552 | 17.179 | 118.338 | 28 |
| ML | 0.397 | 0.446 | 17.017 | 166.469 | 24 |

Table 2 Correlation matrix and standard deviations of parameters obtained from BFGS

| | $\mu_{max1}$ | $K_{s1}$ | $S_1(0)$ |
|---|---|---|---|
| $\mu_{max1}$ | 1.0000 | | |
| $K_{s1}$ | 0.8331 | 1.0000 | |
| $S_1(0)$ | 0.5665 | 0.2214 | 1.0000 |
| S.D. | 0.0022 | 0.0214 | 0.0584 |

Table 3. Correlation matrix and standard deviationsof parameters obtained by the Jack knife technique

| | $\mu_{max1}$ | $K_{s1}$ | $S_1(0)$ |
|---|---|---|---|
| $\mu_{max1}$ | 1.0000 | | |
| $K_{s1}$ | 0.8644 | 1.0000 | |
| $S_1(0)$ | 0.4346 | 0.2179 | 1.0000 |
| S.D. | 0.0023 | 0.0398 | 0.0660 |

Table 4. NPE results for double Monod model to $OUR_{ex}$ data-set (N=323, m=6)

| | $\mu_{max1}$ (/d) | $K_{m1}$ (mg/l) | $\mu_{max2}$ (/d) | $K_{m2}$ (mg/l) | $S_1(0)$ (mg/l) | $S_2(0)$ (mg/l) | MSE (*$10^6$) | Runs |
|---|---|---|---|---|---|---|---|---|
| BFGS | 0.344 | 0.449 | 0.173 | 1.414 | 17.859 | 31.984 | 92.01 | 39 |
| Brent | 0.340 | 0.385 | 0.172 | 1.419 | 17.930 | 32.051 | 74.22 | 37 |
| Simplex | 0.358 | 0.603 | 0.171 | 1.423 | 18.387 | 31.771 | 111.86 | 26 |
| Jackknife | 0.337 | 0.358 | 0.171 | 1.361 | 17.823 | 31.948 | 75.59 | 37 |
| Set-Memb | 0.413 | 0.860 | 0.148 | 1.215 | 20.940 | 28.723 | 729.06 | 8 |

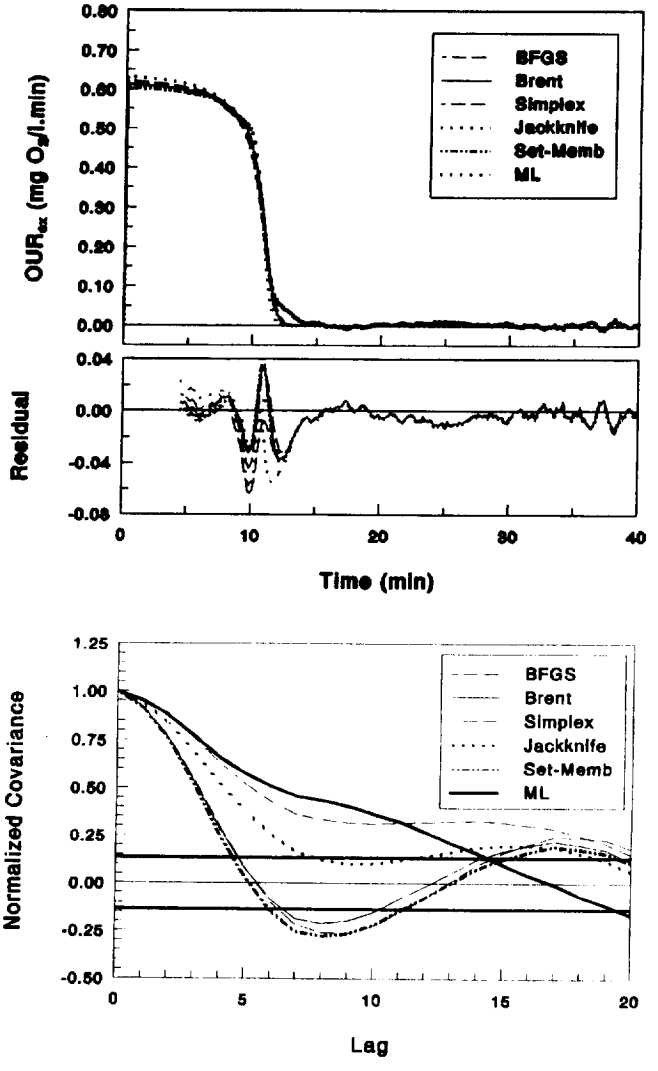


Figure 1. Results of Single Monod type dataset identification with 6 estimation methods (top). Residual sequence (middle) and autocorrelation function (bottom) allow diagnostic checking.

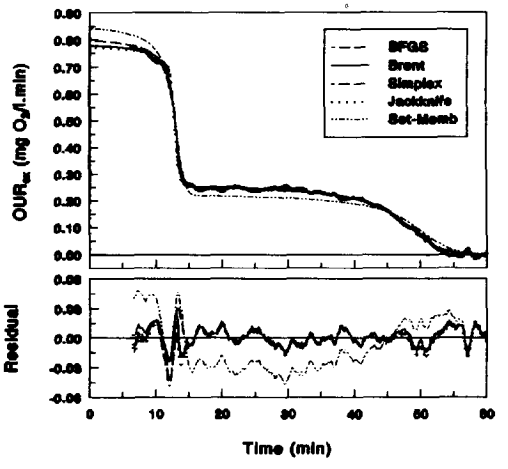


Figure 2. Results of identification of Double Monod model to $OUR_{ex}$ dataset with 5 estimation methods (top). The residual sequence is given for diagnostic checking (bottom).

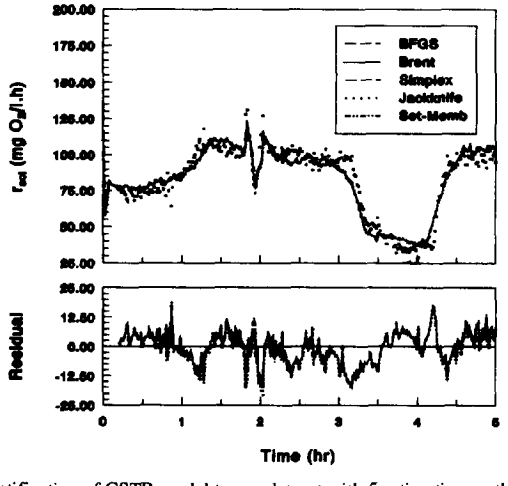


Figure 3. Results of identification of CSTR model to $r_{act}$ dataset with 5 estimation methods (top). The residual sequence is given for diagnostic checking (bottom).

Table 5. Correlation matrix and standard deviations of parameters obtained from BFGS

| | $\mu_{max1}$ | $K_{m1}$ | $\mu_{max2}$ | $K_{m2}$ | $S_1(0)$ | $S_2(0)$ |
|---|---|---|---|---|---|---|
| $\mu_{max1}$ | 1.0000 | | | | | |
| $K_{m1}$ | 0.7813 | 1.0000 | | | | |
| $\mu_{max2}$ | -0.4047 | -0.1465 | 1.0000 | | | |
| $K_{m2}$ | -0.2822 | -0.1024 | 0.8345 | 1.0000 | | |
| $S_1(0)$ | 0.9077 | 0.5061 | -0.5321 | -0.3686 | 1.0000 | |
| $S_2(0)$ | -0.3622 | -0.1310 | 0.7854 | 0.4424 | -0.4784 | 1.0000 |
| S.D. | 0.0017 | 0.0138 | 0.0008 | 0.0397 | 0.0668 | 0.0892 |

Table 6. Correlation matrix and standard deviation of parameters obtained by Jackknife technique

| | $\mu_{max1}$ | $K_{m1}$ | $\mu_{max2}$ | $K_{m2}$ | $S_1(0)$ | $S_2(0)$ |
|---|---|---|---|---|---|---|
| $\mu_{max1}$ | 1.0000 | | | | | |
| $K_{m1}$ | 0.6346 | 1.0000 | | | | |
| $\mu_{max2}$ | -0.3162 | -0.0477 | 1.0000 | | | |
| $K_{m2}$ | -0.2151 | -0.0342 | 0.5055 | 1.0000 | | |
| $S_1(0)$ | 0.7789 | 0.1963 | -0.3858 | -0.2459 | 1.0000 | |
| $S_2(0)$ | -0.3240 | -0.0558 | 0.7609 | 0.2282 | -0.3777 | 1.0000 |
| S.D. | 0.0017 | 0.0186 | 0.0003 | 0.0133 | 0.0748 | 0.0435 |

Table 7. NPE results for CSTR model to $r_{max}$ dataset (N=289, m=4)

| | $r_{max}$ (mg/mg.h) | $K_s$ (mg/l) | Y (mg/mg) | b (/h) | MSE | Runs |
|---|---|---|---|---|---|---|
| BFGS | 57.03 $10^5$ | 3.60 $10^5$ | 0.194 | 4.780 | 45.577 | 67 |
| Brent | 199.690 | 13.823 | 0.194 | 4.791 | 46.325 | 65 |
| Simplex | 99.826 | 6.859 | 0.196 | 4.612 | 47.114 | 57 |
| Jackknife | 191.581 | 12.933 | 0.194 | 4.850 | 46.365 | 59 |
| Set-Memb | 65.125 | 4.000 | 0.192 | 5.000 | 47.733 | 59 |

Table 8. Correlation matrix and standard deviations of parameters obtained from BFGS

| | $r_{max}$ | $K_s$ | Y | b |
|---|---|---|---|---|
| $r_{max}$ | 1.0000 | | | |
| $K_s$ | 1.0000 | 1.0000 | | |
| Y | 0.1543 | 0.1543 | 1.0000 | |
| b | -0.1672 | -0.1672 | -0.9419 | 1.0000 |
| S.D. | 2.87 $10^{11}$ | 1.81 $10^{10}$ | 0.0033 | 0.3060 |

Table 9. Correlation matrix and standard deviations of parameters obtained by the Jackknife technique

| | $r_{max}$ | $K_s$ | Y | b |
|---|---|---|---|---|
| $r_{max}$ | 1.0000 | | | |
| $K_s$ | 0.8945 | 1.0000 | | |
| Y | 0.7069 | 0.8253 | 1.0000 | |
| b | -0.7199 | -0.8270 | -0.9974 | 1.0000 |
| S.D. | 220.7974 | 27.8286 | 0.0249 | 2.3752 |

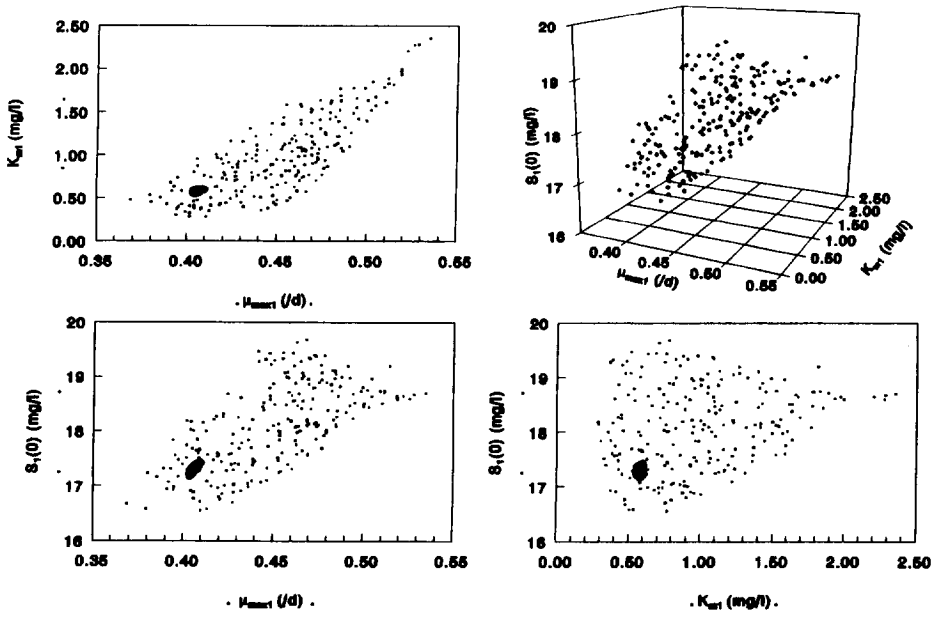


Figure 4. Set membership parameter estimation results from 2000 evaluations of the Single Monod model and the $OUR_{ex}$ dataset. The error bound on $OUR_{ex}$ was set to 0.1 mg $O_2$/l.min. The spot in the 2D-figures is the 95 percent parameter confidence region obtained from the Monte Carlo study.

Differences among NPE methods

Although the similar parameter estimates obtained give the impression that the performance of all evaluated methods is similar, clear differences were noticed. These differences will be classified according to the sensitivity to local minima and the rate of convergence of the methods.

Sensitivity to local minima

Tables 1,4 and 7 show that the estimates are close but not equal to each other. These differences also surface in the minimal MSE's found by the different methods. It appears that the Brent algorithm converges to the lowest MSE. This is not the case for the $r_{act}$ data-set where a parameter bound of 200 mg/mg.h was imposed on $r_{max}$ in case of the Brent parameter estimation.

For the three data-sets tested, but also from previous studies, it was obvious that the BFGS method was the most sensitive to local minima. Only with the values of the other estimation routines, good initial values for the parameters could be given which allowed it to converge close to the set obtained with the other methods. The simplex method was substantially less sensitive to the initial parameter estimate while the Brent method was most robust in this respect. Still, careful inspection of the residual sequence is advisable as local minima can clearly not always be avoided. Evidently, the Jackknife method is also sensitive to local minima as it uses non-linear parameter estimation to calculate the pseudo-values. The Set-membership and Monte Carlo methods are not sensitive to local minima from a numerical point of view, but are clearly dependent on the user-specified initial parameter bounds which may affect the outcome of the estimation procedure.

Rate of convergence

While the robustness against local minima is a more important trait of an estimation method, the rate of convergence is important as it determines the time necessary to complete an identification task. Especially in more complex modelling exercises (such as model selection where parameter estimation may be a necessary subtask) or in case an on-line requirement is set (Vanrolleghem *et al.*, 1994), it may become the most important feature of a method.

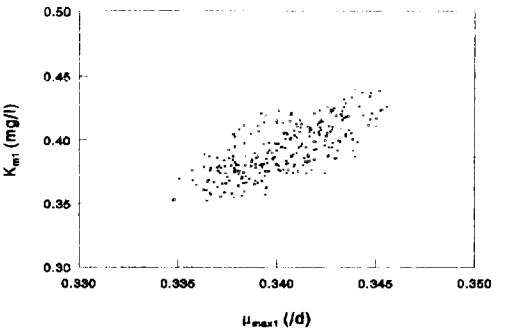


Figure 5. Monte Carlo based 95 % confidence region for the biokinetic parameters of the degradation of the first substrate in the Double Monod model.

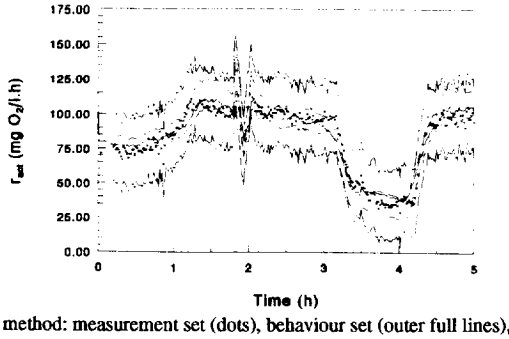


Figure 6. Set-membership method: measurement set (dots), behaviour set (outer full lines), feasible model output set (dashed line) and min-max trajectory (center full line).

The identification studies performed showed that the BFGS method was more efficient than the Brent method (typically for a Double Monod model: 150 (BFGS) and 250 (Brent) evaluations needed). The Simplex method typically required a threefold number of function evaluations than the latter to converge. The Jackknife, Set-Membership and Monte Carlo methods require substantially more computations. The Jackknife requires (N+1) parameter estimations so that the amount of simulations increases equally.

However, if the estimation is started from the estimates obtained on the basis of the whole data-set, an important reduction in necessary computations can be achieved (typically for a Double Monod case: 30000 evaluations). Probably further reductions can be obtained if a robust method (e.g. Brent) is used for the first NPE, while a more efficient NPE (e.g. BFGS) is applied for the estimation of the N subsequent pseudo-values. The methods based on random sampling require a large number of evaluations to allow sufficient details in the parameter confidence region (for the Monte Carlo method) or the feasible parameter set (for the Set-Membership method: 2000 evaluations with approx. 650 feasible vectors for $\eta$=0.2 mg O2/l.min). Fig. 4 illustrates the results of both random sampling methods for the Single Monod case with $OUR_{ex}$ data. One thousand Monte Carlo simulations were sufficient to determine the 95 percent confidence region. For the Double Monod case this has increased to 50000 evaluations to obtain the confidence region depicted in Fig. 5.

Parameter confidence regions and correlation

The correlation matrices and the standard deviations given in Tables 2-3, 5-6 and 8-9 are a concise summary of the parameter confidence regions. Tables 2, 5 and 8 are based on a linear approximation of the objective function in the neighbourhood of the optimal parameter estimates. In Tables 3, 6 and 9 on the other hand more reliable confidence regions are calculated using the Jackknife technique. Comparison shows that for the case studies of this paper little difference exists between the approaches. Finally, the Monte Carlo approach is illustrated in Figs 4 and 5. Here too, it appears that the parameter confidence regions obtained in this way are close to the ones obtained using the other methods. As an example, consider the 95%

confidence region of $\mu_{max1}$ in both the Single Monod (Fig. 4) and the Double Monod case (Fig. 5), which are nearly 0.1 /d large. This corresponds to a 4*S.D. interval, making an estimate for the standard deviation from the Monte Carlo situation to be approximately 0.025 /d. This compares favourably with the S.D.-values obtained from the linearization and Jackknife approaches.

The correlation between parameters in case of Monod-type models is known to be problematic (Vanrolleghem *et al.*, 1995). The high correlation between the maximum growth rate $_{max}$ and the affinity constant $K_m$ are confirmed in the batch experiments (Tables 2-3 and 5-6). In the CSTR experiment, it is found that these parameters are completely correlated so that the model is not identifiable. This makes the evaluation of the parameter confidence region difficult as the Hessian is singular and can therefore not be inverted to yield the covariance matrix. The Jackknife method does allow this as it calculates the matrix directly. A peculiarity of the study was that the 200 mg/mg.h boundary imposed on the $r_{max}$ in the Jackknife estimation made the correlation appear "broken" so that the correlation between $r_{max}$ and $K_S$ was not complete. Finally, an important correlation between Y and b is observed for this model. It is evident that the data are not sufficiently informative for the model that is to be identified.

Model uncertainty

The residual sequences given in Fig. 1 to 3 allow one to point to periods in which the model is not able to describe the behaviour of the system. In Fig. 1 and 2 it is apparent that the models used do not allow one to describe the important transients observed between 8 and 13 minutes after injection of the substrate in the batch reactor.

For the CSTR experimental data, the usefulness of the set-membership method for model adequacy testing can be illustrated. To this end one can examine the relation between the data and the feasible output set (Fig. 6). If all data are contained within the output set obtained with realistic values for the error bounds, one can state that the model is an adequate description of the process. Data that are not within the output set can be considered outliers, or, if a series of such data exist, the corresponding behaviour cannot be described by the applied model. Hence, the model is inadequate to describe this particular process behaviour. In Fig. 6 almost all data are contained in the output set obtained with an error on $r_{act}$ of 25 mg $O_2$/l.h, except for the data associated with the large transients (t=3 and 4 hours). Here, a delay between model predictions and data is observed. Careful inspection showed that a calculation error was made for the input data ($S_{in}$), explaining the modelling error found.

## CONCLUSIONS

In practice, the identification of biodegradation models with Monod type kinetics is mostly hampered by badly identifiable or even unidentifiable model components. In the Monod type limitation function, for instance, the parameters $\mu_{max}$ and $K_m$ are strongly correlated. Furthermore, the residual sequences most often contain a structural component indicating model inadequacy. This notion is important when evaluating parameter estimation methods, since it implies that the assumptions, such as errors independent and identically distributed, are violated. This violation has some effect on the bias of the estimates, but its effect is more pronounced in the evaluation of the estimation uncertainties.

From the viewpoint of efficiency the following ordering for the estimation of model parameters is suggested: BFGS, Brent's and Simplex search method. It appeared, on the contrary, that BFGS is the most sensitive to local minima. To avoid the negative implications of wrong noise assumptions on the estimate uncertainty the bootstrap (jackknife) or Monte Carlo method can be applied. Furthermore, a maximum likelihood estimation procedure explicitly takes into account the measurement and modelling uncertainty in the predictor, resulting in an `error-in-variables' problem. Finally, a set-membership approach completely releases the statistical assumptions and is solely based on a deterministic characterization of the errors in terms of bounds. The effect of this is that not a single `optimal' parameter vector is found but a set of feasible parameter vectors, which directly represents the estimate uncertainty.

It appears that a systematic analysis of model, input and parameter uncertainty analysis is needed and can benefit from the interdisciplinary exchanges currently stimulated by the recent interest in integrated studies of sewer systems, wastewater treatment plants and receiving water bodies. In particular for uncertainty analysis it is felt that knowledge on the different techniques will be shared.

## REFERENCES


Bard Y. (1974) *Nonlinear Parameter Estimation.* Academic Press, New York.

Beale E.M.L. (1960) Confidence regions in non-linear estimation. *J. Roy. Stat. Soc.*, **B22**, 41-88.

Beck M.B. and Chen J. (1994) System identification, parameter estimation and the analysis of uncertainty: A review. In: *Proceedings IAWQ Specialised Seminar on Modelling and Control of Activated Sludge Processes.* Copenhagen, Denmark, August 22-24 1994.

Brent R.P. (1973) *Algorithms for Minimization without Derivatives.* Prentice-Hall, Englewood Cliffs, New Jersey. pp. 195.

Côté M., Grandjean B.P.A., Lessard P. and Thibault J. (1995) Dynamic modelling of the activated sludge process: Improving prediction using neural networks. *Wat. Res.*, **29**, 995-1004.

Dochain D., Vanrolleghem P. and Van Daele M. (1995) Structural identifiability of biokinetic models in activated sludge processes. *Wat. Res.*, **29**, 2571-2578.

Fletcher R. (1987) *Practical methods of Optimization.* John Wiley, Chichester.

Fox T., Hinkley D. and Larntz K. (1980) Jackknifing in nonlinear regression. *Technometrics*, **22**, 29-33.

Keesman K.J. (1990) Membership-set estimation using random-scanning and principal component analysis. *Math. Comp. Simul.*, **32**(5&6), 535-543.

Keesman K.J. and van Straten G. (1990) Set-membership approach to identification and prediction of lake eutrophication. *Water Res. Res.*, **26**(11), 2643-2652.

Ljung L. (1987) *System Identification: Theory for the User.* Prentice Hall, New Jersey.

Lobry J.R. and Flandrois J.P. (1991) Comparison of estimates of Monod's growth model parameters from the same data set. *Binary*, **3**, 20-23.

Milanese M. and Vicino A. (1991) Optimal estimation theory for dynamic systems with set membership uncertainty: an overview. *Automatica*, **27**(6), 997-1009.

Nelder J.A. and Mead R. (1964) A simplex method for function minimization. *Computer J.*, **7**, 308-313.

Norton J.P. (1987) Identification and application of bounded-parameter models. *Automatica*, **23**, 497-507.

Roy T. (1994) Bootstrap accuracy for non-linear regression models. *J. Chemometrics*, **8**, 37-44.

Reichert P., von Schulthess R. and Wild D. (1995) The use of Aquasim for estimating parameters of activated sludge models. *Wat. Sci. Tech.*, **31**(2), 135-147.

Schmidt K. and Isaacs S.H. (1995) An evolutionary algorithm for initial state and parameter estimation in a complex biochemical model. In: *Proceedings 6th IFAC Conference on Computer Applications in Biotechnology - CAB6. Garmisch-Partenkirchen, Germany, May 14-17 1995.* 239-242.

Söderström T. and Stoica P. (1989) Model validation and model structure determination. In: *System Identification.* Prentice- Hall, Englewood Cliffs, New Jersey. 422-467.

Spanjers H. and Keesman K.J. (1994) Identification of wastewater biodegradation kinetics. In: *Proc. 3rd IEEE Conf. on Control Appl.*, Glasgow, pp. 1011-1016.

Spanjers H. and Klapwijk A. (1990) On-line meter for respiration rate and short-term biochemical oxygen demand in the control of the activated sludge process. In: *Instrumentation control and Automation of Water and Wastewater Treatment and Transport Systems (Advances in Water Pollution Control Series),* R. Briggs (ed.), pp 67-77 Pergamon Press, Oxford.

Vanrolleghem P.A., Van Daele M., Van Overschee P. and Vansteenkiste G.C. (1994) Model structure characterization of nonlinear wastewater treatment systems. In: *Proceedings 10th IFAC Symposium on System Identification.* Copenhagen, Denmark, July 4-6 1994, Eds. Blanke M. and Söderström T., Vol. 1, 279-284.

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

Walter E. and Piet-Lahanier H. (1990) Estimation of parameter bounds from bounded-error data: a survey. *Math. Comp. Simul.*, **32**(5&6), 449-468.