IChemE

www.icheme.org/ece
doi: 10.1205/ece06016

1749–7728/06/$30.00+0.00

Trans IChemE, Part D, 2006
Education for Chemical Engineers, 1: 139–145

THE INFLUENCE OF EXPERIMENTAL DATA QUALITY AND QUANTITY ON PARAMETER ESTIMATION ACCURACY
Andrews Inhibition Model as a Case Study

A. GUISASOLA[1,*], J. A. BAEZA[1], J. CARRERA[1], G. SIN[2], P. A. VANROLLEGHEM[2] and J. LAFUENTE[1]

[1]*Department d'Enginyeria Química, ETSE, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain*
[2]*BioMATH, Department of Applied Mathematics, Biometrics and Process Control, Ghent University, Coupure Links 653, B-9000 Gent, Belgium*

Model parameters are usually estimated through minimization algorithms with respect to experimental data. However, students should realize that the values obtained in the classical minimization approach are not always correct and need critical evaluation though the minimum of the cost function is attained. For this purpose, a typical example of a substrate inhibition model in activated sludge processes (Andrews' model) was used. Once the parameters were estimated, the confidence intervals were assessed through a numerical method based on the Fisher Information Matrix. Both procedures were implemented in MATLAB® (software available on request). With this exercise, the student can easily observe how the reliability of the estimated parameter value increases with the increase of data quantity and with the decrease of the data measurement error.

Keywords: confidence interval; Fisher Information Matrix; parameter estimation; substrate inhibition.

INTRODUCTION

The utilization of modelling tools in view of process design and characterization has become very widespread in all engineering fields. Model parameters are estimated through minimization algorithms with respect to experimental data and, afterwards, the calibrated model can be used for process improvement (e.g., in process design or process control) (Petersen *et al.*, 2002). However, for a reliable posterior utilization of the estimated parameters, the assessment of their confidence interval should be as important as the estimation of the parameter values themselves (Brun *et al.*, 2002; Gernaey *et al.*, 2002).

Recently, a lot of research is being conducted in assessing the precision of the parameters estimated from experimental data (Vanrolleghem *et al.*, 1995; Omlin and Reichert, 1999; Walter and Pronzato, 1999; Dochain and Vanrolleghem, 2001; Brun *et al.*, 2002; Marsili-Libelli *et al.*, 2003). Confidence interval assessment is not a straightforward task since many different factors are involved such as the experimental data, the inherent structure of the model or the minimization approach used (Beck, 1987).

The exercise proposed in this work was planned so that the student can understand how the model parameters are estimated from experimental data and help the student to realize that each parameter has an uncertainty which needs to be considered. The teacher can choose whether students need to implement the whole procedure by themselves or, simply, give them the software already implemented in MATLAB® (MATLAB, 2002) and design an appropriate application for it. The MATLAB scripts used in this work will be freely provided by the authors on request.

Mathematical models based on first principles can be divided into two categories: linear versus non-linear models. Although linear models are easier to solve, a wide range of processes and phenomena occurring in nature as well as in engineered systems are non-linear and therefore require a non-linear model structure. Hence, we use here an example of a non-linear model application. In particular, we focus on studying the relation between the quality and quantity of experimental measurements and the confidence intervals of the parameter estimates. For this purpose, a typical substrate inhibition model within microbial growth processes (Andrews' model) was used.

The parameter estimation was carried out with the classical simplex Nelder and Mead minimization algorithm (Nelder and Mead, 1965), using as a cost function the norm of the difference between the experimental data and the model. The confidence intervals were assessed through a numerical method based on the Fisher Information Matrix (FIM). Once the procedure is implemented, several changes can be made to analyse their influence on the parameter estimation and the confidence interval assessment. These changes can be related to the model itself, to the quantity and quality of the experimental data used or to the initial guesses in the minimization algorithm.

[*]*Correspondence to*: Dr A Guisasola, Department d'Enginyeria Química, ETSE, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain.
E-mail: Albert.Guisasola@uab.cat

The general procedure for parameter estimation error assessment using the FIM is detailed in the methodology section after a brief description of model linearity and normal distribution. The case study used in this work (i.e., Andrews' substrate inhibition model) is also described in this section together with the experimental data generation procedure. The results section uses the Andrews' model to assess the effect of quantity and quality of this experimental data on the value of the confidence interval of parameter estimates.

METHODOLOGY

Linear Versus Non-Linear Models

Linearity is a basic characteristic of the model which has a very high impact on the properties of the solution. Two different kinds of linearity can be distinguished (Walter and Pronzato, 1999):

- *Input linearity:* a model is linear in its inputs when it satisfies the superposition principle with respect to its inputs (u) [equation (1)].

$$y_M(t, \theta, \alpha \cdot u_1 + \beta \cdot u_2) = \alpha \cdot y_M(t, \theta, u_1) + \beta \cdot y_M(t, \theta, u_2) \quad (1)$$

  where y_M = measured output, $\theta =$ parameter (θ: vector of parameters), $u =$ input (u: vector of inputs).
- *Parameter linearity:* a model is linear in its parameters when it satisfies the superposition principle with respect to its parameters (θ) [equation (2)].

$$y_M(t, \alpha \cdot \theta_1 + \beta \cdot \theta_2, u) = \alpha \cdot y_M(t, \theta_1, u) + \beta \cdot y_M(t, \theta_2, u) \quad (2)$$

Examples of all the linear possibilities are shown in Table 1.

Linear models are preferred since an analytical solution can easily be found and many mathematical tools have already been developed to this end. On the other hand, non-linear systems may require complex numerical solutions and that is why they tend to be linearized. This linearization is performed either because the model can be considered as linear in the context of this study or because we have been able to transform it to a linear form by a proper variable manipulation. For example, the system can be linearized around some equilibrium point (or steady state). However, the results obtained from the linearized model will only be valid close to these parameter and variables values. Therefore, the utilization of non-linear models [where classical parameter uncertainty assessment tools can not be used (see below)] is increasing because of the narrow range of applicability of linearized non-linear models.

Table 1. Description of the different existing linearities.

Model example	Input linearity	Parameter linearity
$\theta \cdot u(t)$	Yes	Yes
$\log(\theta) + u(t)$	Yes	No
$\theta \cdot u^2(t)$	No	Yes
$\log(\theta) + u^2(t)$	No	No

Normal Distribution

The normal distribution [equation (3)] is a two-parameter distribution that describes the distribution of the population in terms of probability versus value. In short, for a certain population with a certain mean (λ) and a certain standard deviation (σ), the following describes the probability of appearance of a certain value around μ. Figure 1 shows an example of normal distribution for $\lambda = 10$ and $\sigma = 2$.

$$f(\lambda,\sigma) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-(x-\lambda)^2/2\sigma^2} \quad (3)$$

Parameter Estimation and Confidence Interval Assessment

The parameters can be estimated through a minimization algorithm where the weighed sum of squared errors, J [equation (4)], between model outputs $\vec{y}(k, \vec{\theta})$ and the measured outputs $\vec{y}_M(k)$ is minimized (k represents a certain sampling point). $\vec{Q}_k$ is a weighting matrix to balance the effect of each kind of measurement.

$$J = \sum_{k=1}^{N} [\vec{y}(k, \vec{\theta}) - \vec{y}_M(k)]^T \vec{Q}_k [\vec{y}(k, \vec{\theta}) - \vec{y}_M(k)] \quad (4)$$

where N is the number of measurements and $\vec{\theta}$ is the parameter set used to calculate the model outputs.

The presented confidence interval assessment procedure is based on the FIM (Dochain and Vanrolleghem, 2001). This matrix is regarded as an indicator of the amount of information contained in the experimental data. The FIM is calculated using a linearization of each one of the output signals in the neighbourhood of the optimal vector

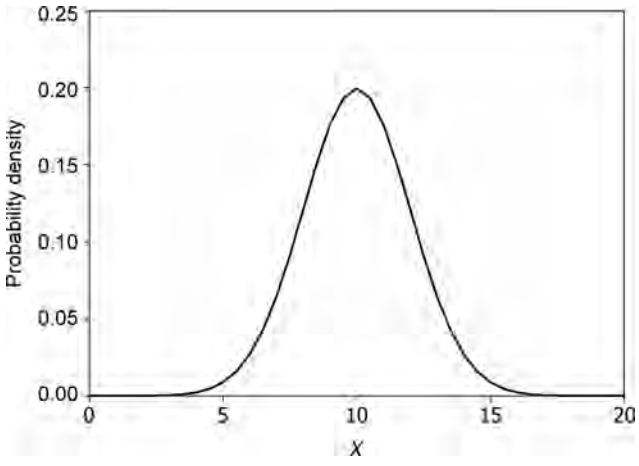

Figure 1. Example of normal distribution with $\lambda = 10$ and $\sigma = 2$.

of parameters $\vec{\theta}_O$. The linearization is conducted for each parameter as expressed in equation (5).

$$y(t, \vec{\theta}_O + d\theta) = y(t, \vec{\theta}_O) + \left[\frac{\partial y(t,\vec{\theta})}{\partial \theta}\right]_{\theta_O} \cdot d\theta$$
$$= y(t, \vec{\theta}_O) + Y^y_\theta(t)\, d\theta \quad (5)$$

$Y^y_\theta(t)$ represents the so called output sensitivity function, which is the derivate of an output variable (y) with respect to one parameter (θ) and can be approximated by a central difference expression [equation (6)]:

$$Y^y_\theta = \frac{\partial y(t, \vec{\theta})}{\partial \theta}(k)$$
$$= \lim_{\Delta\theta \to 0} \frac{y(t, \vec{\theta} + \Delta\theta) - y(t, \vec{\theta} - \Delta\theta)}{2\Delta\theta} \quad (6)$$

The information related to the uncertainties and dependencies among the estimated parameters can also be summarized in the FIM (Mehra, 1974). If $\vec{\vec{Q}}_k$ is calculated as the inverse of the covariance matrix of the measurement noise (error), the FIM is defined as equation (7).

$$FIM = \sum_{k=1}^{N} \vec{\vec{Y}}^T_\theta(k) \vec{\vec{Q}}_k \vec{\vec{Y}}_\theta(k) \quad (7)$$

The FIM is a square matrix with the same number of columns and rows as the number of parameters to estimate. $\vec{\vec{Q}}_k$ is also a square matrix with the same number of columns and rows as output measurements used. If the measurement error in each sampling time was known, $\vec{\vec{Q}}_k$ would be different for each sampling time. However, in reality, measurement errors depend on the measurement technique/sensor/instrument itself. Often one finds the standard errors of the measurements to be absolute, i.e., constant. In such a case, $\vec{\vec{Q}}_k$ becomes a scalar for each output variable for all sampling times.

An example of the FIM calculation, for a system with three estimated parameters (θ_1, θ_2 and θ_3) and two output variables ($y1$, $y2$), is given below in equation (8):

$$FIM = \sum_{k=1}^{N} \begin{bmatrix} a(k) & b(k) & c(k) \\ d(k) & e(k) & f(k) \\ g(k) & h(k) & i(k) \end{bmatrix} \quad (8)$$

where the FIM elements for a certain time (k) correspond to

$$\begin{bmatrix} a(k) & b(k) & c(k) \\ d(k) & e(k) & f(k) \\ g(k) & h(k) & i(k) \end{bmatrix}$$
$$= \begin{bmatrix} Y^{y1}_{\theta1}(k, \vec{\theta}_O) & Y^{y2}_{\theta1}(k, \vec{\theta}_O) \\ Y^{y1}_{\theta2}(k, \vec{\theta}_O) & Y^{y2}_{\theta2}(k, \vec{\theta}_O) \\ Y^{y1}_{\theta3}(k, \vec{\theta}_O) & Y^{y2}_{\theta3}(k, \vec{\theta}_O) \end{bmatrix} \cdot \begin{bmatrix} \vec{Q}_1(k) & 0 \\ 0 & \vec{Q}_2(k) \end{bmatrix}$$
$$\times \begin{bmatrix} Y^{y1}_{\theta1}(k, \vec{\theta}_O) & Y^{y1}_{\theta2}(k, \vec{\theta}_O) & Y^{y1}_{\theta3}(k, \vec{\theta}_O) \\ Y^{y2}_{\theta1}(k, \vec{\theta}_O) & Y^{y2}_{\theta2}(k, \vec{\theta}_O) & Y^{y2}_{\theta3}(k, \vec{\theta}_O) \end{bmatrix}$$

$\vec{Q}_i(k)$ corresponds to the inverse of the covariance of the measurement noise of the output variable y_i for each sampling point (k). When the measurement error is considered constant along the experiment, $\vec{Q}_i$ becomes a scalar instead of a vector.

The FIM matrix summarizes the quantity and quality of information obtained in each experiment as it considers the output sensitivity functions and the measurement errors of the experimental data (i.e., precision of an experiment). Assuming no model mismatch, no data autocorrelation, white measurement noise (i.e., independent and normally distributed with zero mean) and uncorrelated errors (i.e., the measurement error covariance matrix is a diagonal matrix), the inverse of the FIM provides the lower bound of the parameter estimation error covariance matrix, which can be used for assessing the estimation uncertainty of $\vec{\theta}_O$ (Dochain and Vanrolleghem, 2001) as shown in equation (9).

$$COV(\vec{\theta}_O) \geq FIM^{-1} \quad (9)$$

Then, approximate standard errors for the estimated parameters can be calculated as the square root of the diagonal elements of the inverse of the FIM [equation (10)]:

$$\sigma(\theta_i) = \sqrt{COV(i,i)} \quad \text{where } COV = FIM^{-1} \quad (10)$$

Hence, the higher the FIM values, the lower the standard errors estimated. This is understandable if one analyses which factors make the FIM be higher [equation (7)]: high parameter sensitivity (high Y^y_θ) and low measurement errors (high $\vec{\vec{Q}}_k$).

Note that since output sensitivities of parameters with respect to measurements are calculated using a model, the FIM also depends on the structure of the model. The model structure is a very important issue to take into account; since the FIM procedure aforementioned is based on no model mismatch (i.e., the model can describe correctly the experimental observations with the correct parameters). This assumption of 'correct model' should not always be instantaneously accepted without examination.

Software Required

The modelling software used is MATLAB® 6.5. The differential equations are solved using the function ode45. This solver is based in an explicit Runge–Kutta formula, the Dormand-Prince pair. Parameter optimization to fit to experimental data is carried out by using the heuristic method implemented in the MATLAB® function *fminsearch* (Nelder–Mead-Simplex) (Nelder and Mead, 1965), using equation (4) as the cost function.

Case Study: Andrews Substrate Inhibition Model

The response of a microbial population to external substrate presence is generally modelled using Monod kinetics. However, different patterns of dependence on substrate concentration have been described in the literature. These different responses are commonly modeled using variants of the Monod kinetics. Substrate inhibition is a frequent phenomenon observed in the literature and several

modifications of Monod kinetics have appeared for its description (Luong, 1987; Meriç *et al.*, 2002; Han and Levenspiel, 1988). Andrews' modification (Andrews, 1968) is probably the most often used substrate inhibition model [equation (11)].

$$\mu = \frac{\mu_{MAX} \cdot S}{(S + K_S)(1 + (S/K_I))} \tag{11}$$

where μ_{MAX} = maximum specific growth rate (d^{-1}), S = substrate concentration (g L^{-1}), K_S = half-saturation constant (g L^{-1}), K_I = inhibition constant (g L^{-1}).

This model derives from the Haldane equation (Haldane, 1965), which described enzyme inhibition by the formation of an inactive complex of the enzyme with two substrate molecules. In most cases, the ratio K_S/K_I is considered to be negligible and, hence, the equation (11) can be simplified in terms of equation (12):

$$\mu = \frac{\mu_{MAX} \cdot S}{S + K_S + (S^2/K_I)} \tag{12}$$

Many applications of the Andrews inhibition model can be found in the literature. An example is the nitrification process, which is a two-step process where ammonia is first oxidized to nitrite and subsequently to nitrate. Both steps are considered to be inhibited by their own substrate: ammonia and nitrite respectively (Anthonisen *et al.*, 1976; Carrera *et al.*, 2004).

Case Study: Dynamic Model

The output measurement considered in this study was not the substrate concentration but the oxygen uptake rate (OUR) profile. This output variable has high frequency and accuracy and it is widely used for modelling purposes because it is relatively easy to measure and meanwhile provides a lot of information (Spanjers *et al.*, 1998). The amount of oxygen taken up is stoichiometrically linked to the substrate consumption rate by means of the biomass substrate yield. Equation (13) shows [in terms of mass basis of chemical oxygen demand (COD) and nitrogen] the biomass growth process that relies on consumption of a defined organic substrate.

$$\frac{1}{Y} CH_yO_z + \left(1 - \frac{1}{Y}\right) O_2 + i_{NB} \cdot NH_3 \longrightarrow CH_aO_bN_c + (\cdots)CO_2 + (\cdots)H_2O \tag{13}$$

where CH_yO_z is the substrate (mg COD_S L^{-1}), $CH_aO_bN_c$ is the biomass (mg COD_B L^{-1}), Y is the biomass growth yield (g COD_B g^{-1} COD_S) and i_{NB} is the fraction of nitrogen in biomass (g N g^{-1} COD_B).

Equations (14) show a simplified set of ordinary differential equations to describe the dynamics of the batch reactor and Table 2 shows the parameters involved.

$$\frac{dX}{dt} = \frac{\mu_{MAX} \cdot S}{(S + K_S + (S^2/K_I))} \cdot X \quad \frac{dS}{dt} = -\frac{1}{Y} \cdot \frac{dX}{dt}$$
$$OUR(t) = \frac{(1 - Y)}{Y} \frac{dX}{dt} \tag{14}$$

Table 2. Model parameters.

Known parameters		Unknown parameters
Biomass yield: Y	0.67 (g COD_B g^{-1} COD_S)	μ_{MAX} (d^{-1})
Initial heterotrophic biomass: $X(0)$	2000 (mg COD_B L^{-1})	K_S (mg COD_S L^{-1})
Initial substrate: $S(0)$	200 (mg COD_S L^{-1})	K_I (mg COD_S L^{-1})
Final time	50 min	

Only a few of the parameters of the model [equation (14)] can be reliably estimated using a single measurement such as OUR because of identifiability problems (Petersen *et al.*, 2001; Gernaey *et al.*, 2002). Identifiability analysis has often shown that parameter estimation from output measurements depends on the model structure and the available measurements used for calibration (Dochain and Vanrolleghem, 2001). It is often shown that only a few of the total parameters of the model can be estimated using a single measurement (e.g., respirometric data). Structural identifiability refers to the assessment of identifiable model parameters considering the measurements available, the model characteristics and assuming ideal experimental data. If the quantity and quality of data is also taken into account, this step is called practical identifiability. Both steps can be learned from other sources dedicated for it e.g., Dochain and Vanrolleghem, 2001). As this is out of the scope of this current contribution, the parameters chosen for estimation from the OUR measurements (μ_{MAX}, K_S and K_I) were selected for illustrative purposes, whereas the biomass yield and the initial values of substrate and biomass were assumed known. The selection of the known and unknown parameters can be changed according to the user needs. For example, if the kinetic parameters were considered to be known, the initial values of the state variables (S, X) would become the unknown parameters and could be estimated.

Case Study: Experimental Data Generation

The 'experimental data' used in this work were previously generated with the model described in

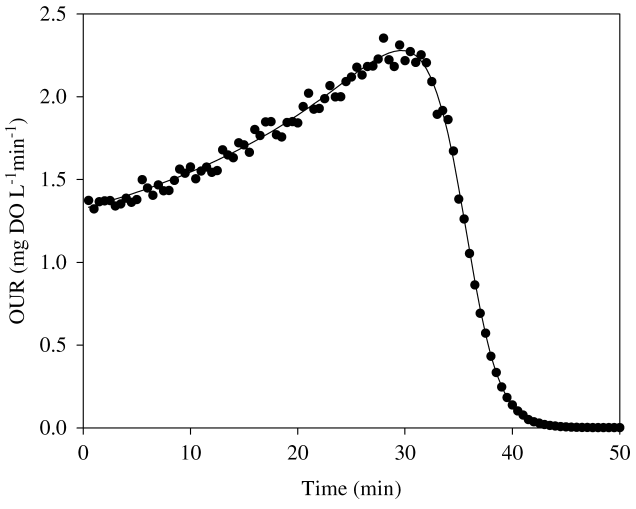

Figure 2. OUR profiles: reference (solid line) and experimental (•) (3% of measurement error).

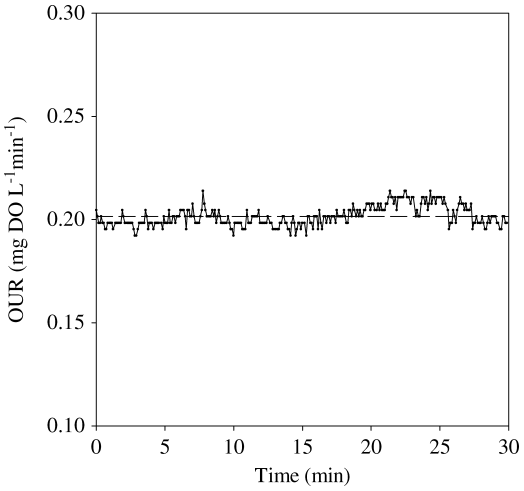

Figure 3. Example of Q_{OUR} estimation.

equations (14) with $\mu_{MAX} = 6\ d^{-1}$, $K_S = 20$ mg COD_S L^{-1} and $K_I = 100$ mg COD_S L^{-1}. The OUR profile obtained (from this point on termed reference OUR) is depicted in Figure 2. An error was added to each of the experimental measurements so that the data became more realistic. The measurement error noise was considered to be white, so it followed a normal distribution with zero mean. The MATLAB® function *normrnd* was used to add random experimental error to the reference OUR. This function returned a matrix of random numbers chosen from the normal distribution with the input parameters λ (reference OUR) and σ (measurement error). This function is only available with the statistics toolbox. However, MATLAB® also contains the function *random* in the normal package which can be as well used for the 'experimental data' generation. The experimental data obtained (using 3% of relative measurement error) are also depicted on Figure 2.

RESULTS AND ANALYSES

Procedure Implementation. Program Structure

The structure of the software implemented is described next. First of all, the 'experimental' OUR measurements were generated as indicated in the paragraph above. Then, μ_{MAX}, K_S and K_I were estimated through the minimization function *fminsearch.* The cost function was defined as the norm of the vector resulting of the difference between the experimental and the modelled data [equation (4)]. The OUR profile obtained using the optimal parameters (which minimize the cost function) was used for the calculation of the sensitivity functions, which are the basis for the calculation of the FIM. The sensitivity function of the output measurements with respect to each parameter and the FIM were calculated by equations (6) and (7), respectively.

The measurement error covariance matrix ($\vec{\vec{Q}}_k$) for the respirometric measurements (OUR) was calculated following Petersen's method (Petersen *et al.*, 2001). Petersen's method used the data from a period when the value of the output variable is assumed known. For example, Q_{OUR} can be estimated using a phase without external substrate (i.e., endogenous respiration phase) in which OUR is assumed to be relatively constant during a short-term period e.g., 30 min (Figure 3). In this constant OUR period, the average of the data and the resulting residuals (the difference between the average OUR value and the experimental data) are calculated. The measurement error (s^2) is then calculated as follows [equation (15)]:

$$s^2 = \frac{SSE}{N - p} \tag{15}$$

where SSE = sum squared srrors, N = number of OUR measurements, p = number of parameters.

Then, the measurement error weighting matrix ($\vec{\vec{Q}}_k$) is calculated taking the inverse of s^2. This choice of $\vec{\vec{Q}}_k$

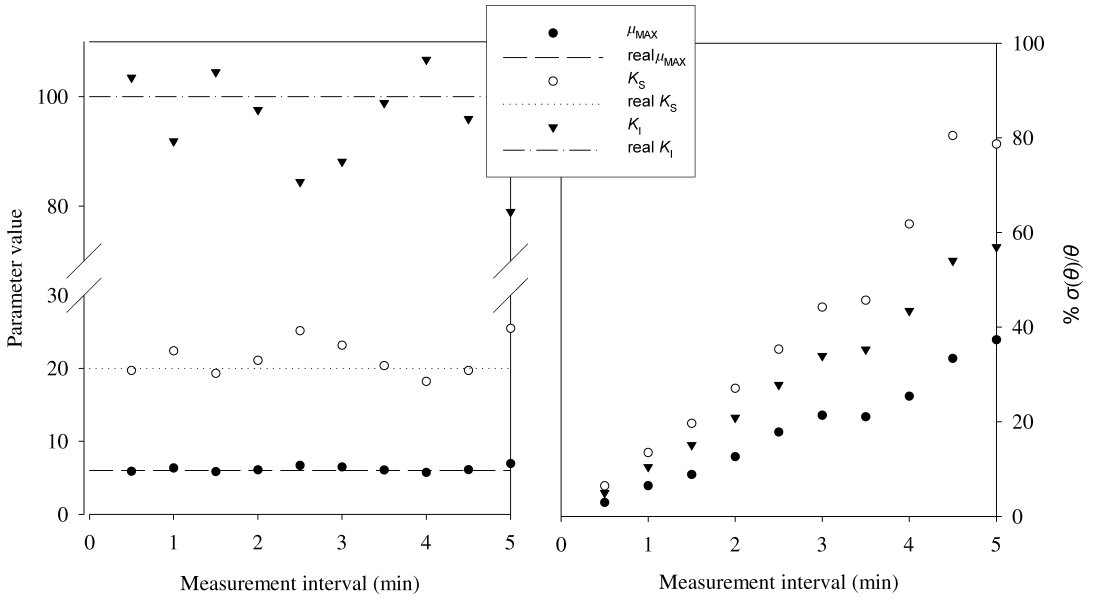

Figure 4. Influence of the measurement interval on the estimated parameter value (left) and on the confidence interval assessment (right).

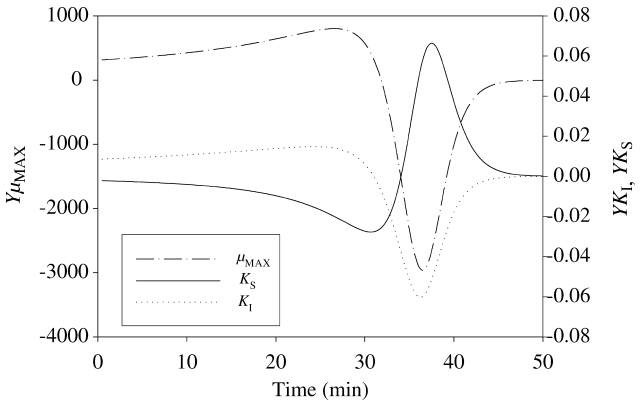

Figure 5. Sensitivity functions of OUR with respect to μ_{MAX}, K_S and K_I.

means that the more a measurement error is noise corrupted, the less it will count in the FIM. Finally, approximate standard errors for the estimated parameters can be calculated using equation (10).

Once the procedure is implemented, several changes can be developed to study their influence on the parameter estimation and the confidence interval assessment. These changes can be related to the model itself (variations in the model equations), to the experimental data used (number and quality) or to the initial guesses in the minimization algorithm. In this study, the influence of the number of data used for parameter estimation and the influence of the measurement error of this data were analysed.

Influence of Quantity of Experimental Data

In order to analyse the influence of the data quantity, the whole procedure was run several times varying the 'experimental' measurement interval (sampling frequency): this was done by varying the number of elements of the same 'experimental' OUR vector. The parameter estimation results and standard deviation assessment are depicted in Figure 4. On the one hand, as can be seen in Figure 4 (left), there is a high variability in the parameter estimates depending on the measurement interval used, particularly in the case of K_I. The general trend observed is the lower the sampling frequency (the higher the number of measurements), the closer the final estimated values to the real ones. When the measurement interval used was higher than one measurement every 2 min (in this case, less than 20 measurements), the values of the parameter estimation were not reliable and differ a lot from the real ones. However, there is also the possibility to obtain a good set of parameters with a fortunate set of experimental data, as for example the parameter values obtained with a measurement interval of 3.5 min. In addition, if the parameter fitting procedure was repeated with another experimental data set, different parameter values could be obtained though the same measurement interval was used. On the other hand, Figure 4 (right) shows the ratio of the standard deviation of each parameter [calculated through equation (10)] over the value of the parameter versus the measurement interval. The graph shows that the lower the measurement interval is, the more reliable the values of the estimated parameters become.

An important fact to highlight is that even though the measurement error was unique, the relative parameter estimation errors [Figure 4(b)] obtained are different for the three parameters. The reason can be found by analysing the sensitivity of OUR to each parameter. As an example, let us compare μ_{MAX} and K_I. As can be seen in Figure 5, in the experiment proposed in this work, the sensitivity of OUR to μ_{MAX} was higher than the sensitivity of OUR to K_I in terms of absolute value. In short, small variations on the μ_{MAX} value would have stronger effects on the OUR profile than small variations on the K_I value. Hence, the value of μ_{MAX} would be always more easily assessed and the relative estimation error of μ_{MAX} would be always lower than the relative estimation error of K_I.

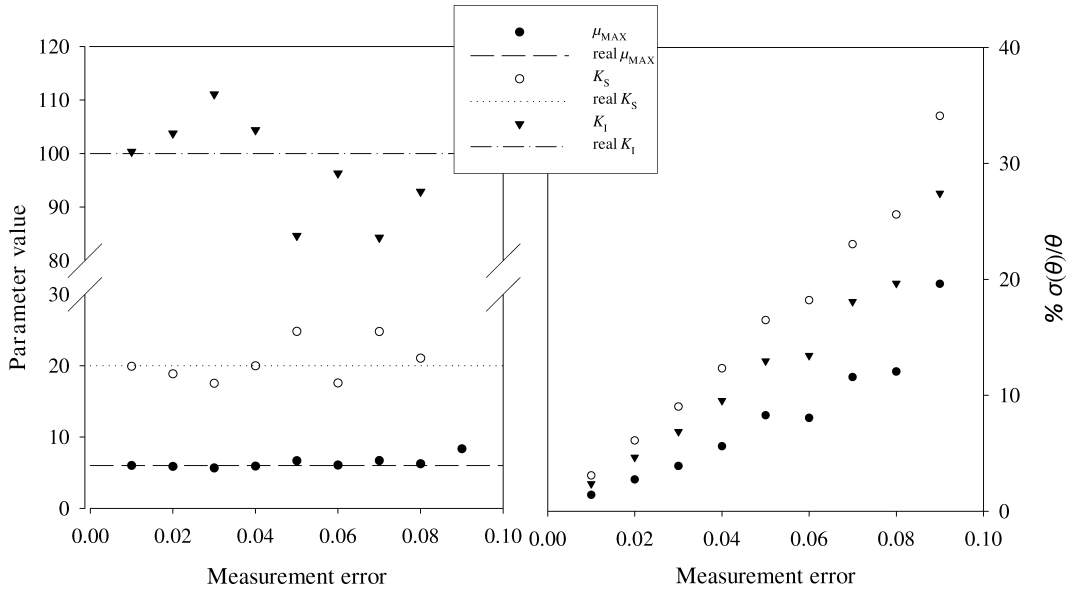

Figure 6. Influence of measurements error on the parameter estimation (left) and on the confidence interval assessment (right).

Influence of Quality of Experimental Data

The influence of the data quality was studied through variation of the experimental data measurement error. This measurement error appears twice in this example: firstly when generating the 'experimental' data and secondly when calculating the measurement error covariance matrix. The results obtained using different measurement errors are depicted in Figure 6. At first sight, the parameter estimation results obtained agree with what should be expected: the increase of measurement error implied obtaining different parameter estimates [Figure 6 (left)] and an increase in the confidence intervals [Figure 6 (right)]. Although a good set of estimated parameter values can be obtained, even with a high measurement error (see for example error 0.06), the estimates have a high uncertainty around 10–20% [Figure 6 (right)]. The repetition of the experiment with the same measurement error would probably lead to different estimated parameter values.

The main notion derived form these two studies is that parameter estimation should be a two-step process: parameter optimization + parameter error assessment. Otherwise, if only the first step was developed, the student would not be able to realize whether the obtained parameters were far from reality or not. For example, one should be able to know that the parameter estimation values obtained with the higher measurement error and the lower frequency sampling are not as reliable as the ones obtained with lower measurement error and higher sampling frequency. As shown on Figures 4 and 5, these parameter estimation values obtained in both cases are very different.

CONCLUSIONS

An easily implementable procedure for parameter estimation and confidence interval assessment is described using the common Andrews substrate inhibition model as an example. In addition, to understand the parameter estimation and confidence interval assessment, it is shown to students that a critical evaluation on the parameter estimation values is always required. The student will also realize the importance of the confidence interval assessment in view of future usage of the parameter estimation values (for example for process design or control). For this purpose, the effect of the quantity (number of measurements) and quality (measurement error) of data on the parameter estimation and confidence interval assessment is also clearly depicted. In general, the results show that the increase of the measurement frequency and the decrease of the measurement error imply a more accurate parameter estimation (in terms of proximity of the estimated value to the real one and in terms of confidence interval assessment).

REFERENCES

Andrews, J.F., 1968, A mathematical model for the continuous culture of microorganisms utilizing inhibitory substrate, *Biotechnol Bioeng*, 10: 707–723.

Anthonisen, A.C., Loehr, R.C., Prakasan, T.B.S. and Srineth, E.G., 1976, Inhibition of nitrification by ammonia and nitrous acid, *J Water Pollut Control Fed*, 48(5): 835–852.

Beck, M.B., 1987, Water quality modelling: a review of the analysis of uncertainty, *Water Resour Res*, 23: 1393–1442.

Brun, R., Kuhni, M., Siegrist, H., Gujer, W. and Reichert, P., 2002, Practical identifiability of ASM2d parameters—systematic selection and tuning of parameter subsets, *Wat Res*, 36(16): 4113–4127.

Carrera, J., Jubany, I., Carvallo, L., Chamy, R. and Lafuente, J., 2004, Kinetic models for nitrification inhibition by ammonium and nitrite in a suspended and an immobilised biomass systems, *Process Biochem* 39: 1159–1165.

Dochain, D. and Vanrolleghem, P.A., 2001, *Dynamical Modelling and Estimation in Wastewater Treatment Processes* (IWA Publishing, London, UK).

Gernaey, K., Petersen, B., Dochain, D. and Vanrolleghem, P.A., 2002, Modelling aerobic carbon source degradation processes using titrimetric data and combined respirometric-titrimetric data: Structural and practical identifiability, *Biotech Bioeng*, 79: 754–769.

Haldane, J.B.S., 1965, *Enzymes* (MIT Press, Cambridge, MA, USA).

Han, K. and Levenspiel, O., 1988, Extended Monod kinetics for substrate, product and cell inhibition, *Biotechnol Bioeng*, 32: 430–437.

Luong, J.H.T., 1987, Generalisation of Monod kinetics for analysis of growth data with substrate inhibiton, *Biotechnol Bioeng*, 24: 242–248.

Marsili-Libelli, S., Guerrizio, S. and Checchi, N., 2003, Confidence regions of estimated parameters for ecological systems, *Ecol Model*, 165: 127–146.

MATLAB, 2002, *User's Guide. Version 6 (Release 13)* (The MathWorks, Natick, MA, USA).

Mehra, R.K., 1974, Optimal input signals for parameter estimation in dynamic systems: survey and new results, *IEEE Transactions on Automatic Control*, 19(6): 753–768.

Meriç, S., Tünay, O. and Ali San, H., 2002, A new approach to modelling substrate inhibition, *Env Technol*, 23: 163–177.

Nelder, J.A. and Mead, R., 1965, A simple method for function minimization, *The Computer Journal*, (7): 308–313.

Omlin, M. and Reichert, P., 1999, A comparison of techniques for the estimation of model prediction uncertainty, *Ecol Model*, 115: 45–59.

Petersen, B., Gernaey, K., Henze, M. and Vanrolleghem, P.A., 2002, Evaluation of an ASM1 model calibration procedure on a municipal–industrial wastewater treatment plant, *J Hydroinformatics*, 4: 15–38.

Petersen, B., Gernaey, K. and Vanrolleghem, P.A., 2001, Practical identifiability of model parameters by combined respirometric-titirimetric measurements, *Water Sci Technol*, 43(7): 647–655.

Spanjers, H., Vanrolleghem, P.A., Olsson, G. and Dold, P.L., 1998, *Respirometry in Control of the Activated Sludge Process: Principles*. Scientific and technical report, no. 7 (IWA Publishing, London, UK).

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

Walter, E. and Pronzato, L., 1999, *Identification of Parameteric Models from Experimental Data* (Springer Verlag, Heidelberg, Germany).

The manuscript was received 19 July 2006 and accepted for publication after revision 3 October 2006.