A FAST AND SYSTEMATIC PROCEDURE TO DEVELOP DYNAMIC MODELS OF BIOPROCESSES - APPLICATION TO MICROALGAE CULTURES

J. Mailier[1] and A. Vande Wouwer[2*]

Control Department, Engineering Faculty, University of Mons (UMONS),
31 Boulevard Dolez, B-7000 Mons, Belgium.
[1]Phone: +32-(0)65374133, Fax: +32-(0)065374136
E-mail: johan.mailier@umons.ac.be
[2]Phone: +32-(0)65374141, Fax: +32-(0)065374136
E-mail: alain.vandewouwer@umons.ac.be

Abstract - The purpose of this paper is to report on the development of a procedure for inferring black-box, yet biologically interpretable, dynamic models of bioprocesses based on sets of measurements of a few external components (biomass, substrates, and products of interest). The procedure has three main steps: (a) the determination of the number of macroscopic biological reactions linking the measured components; (b) the estimation of a first reaction scheme, which has interesting mathematical properties, but might lack a biological interpretation; and (c) the "projection" (or transformation) of this reaction scheme onto a biologically-consistent scheme. The advantage of the method is that it allows the fast prototyping of models for the culture of microorganisms that are not well documented. The good performance of the third step of the method is demonstrated by application to an example of microalgal culture.
Keywords: Reaction networks; Mathematical modeling; Parameter estimation; Bioprocesses.

INTRODUCTION

The increasing demand for process monitoring, control, and optimization encourages the development of various techniques for modeling bioprocesses. Among these techniques, the one proposed by Bastin and Dochain (1990) relies on the concept of a macroscopic reaction scheme, i.e., a set of M reactions that represent macroscopic relationships between the N key-components of the system (biomass, substrates, and products). The general form of a macroscopic reaction scheme can be expressed as:

$$\sum_{i \in \mathcal{R}_k} (-\nu_{ik})\xi_i \xrightarrow{\varphi_k} \sum_{j \in \mathcal{P}_k} \nu_{jk}\xi_j, \text{ for } k \in [1, M], \quad (1)$$
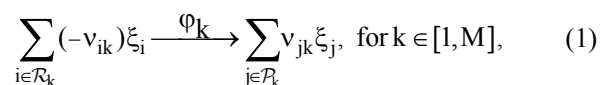
where $\mathcal{R}_k$ and $\mathcal{P}_k$ are, respectively, the sets of reactants and products of reaction k, and φ_k is the kth reaction rate, generally expressed as a non-negative, nonlinear function of the reacting species concentrations ξ_i and ξ_j. The coefficients ν_{*k} are the pseudo-stoichiometric (or yield) coefficients of reaction k, which are positive when related to a product and negative for a reactant. A necessary condition for independence between the reactions is to force their number to be less than or equal to the number of components, i.e., $M \leq N$.

This formulation provides a representation of the main mechanisms underlying any bioprocess without dealing with complex metabolic pathways. In addition, Equation (1) attempts to preserve the

---

.

essential biological aspects that are of interest for further process monitoring and optimization. Assuming for the sake of clarity no gaseous outflow, a simple mass balance thus gives the following set of ordinary differential equations, written in matrix form:

$$\frac{d\xi(t)}{dt} = K\varphi(\xi) - D(t)\xi(t) + F(t), \qquad (2)$$

where $\xi(t) \in \mathbb{R}^N$ is the vector of species concentrations, $K \in \mathbb{R}^{M \times N}$ the pseudo-stoichiometric matrix, $\varphi(\xi) \in \mathbb{R}^M$ the reaction rate vector, $D(t) \in \mathbb{R}$ the input dilution rate, and $F(t) \in \mathbb{R}^N$ the vector of external feed rates. From Equation (2), it is obvious that the model reactions are linearly independent if and only if matrix K is full rank:

$$\text{rank}(K) = M. \qquad (3)$$

The next section introduces the three successive stages of the identification methodology and the subsequent section illustrates its use in a simulation example of microalgal growth, using a mechanistic reformulation of the well-known Droop model (Lemesle and Mailleret, 2008).

ROBUST METHODOLOGY OF IDENTIFICATION

The present work concerns the identification of model (2) from a limited set of noisy measurements of the state vector:

$$\xi_m\left(t_k^e\right), \qquad \text{with} \begin{cases} e = 1, \dots, N_e \\ k = 1, \dots, N_s^e \end{cases}, \qquad (4)$$

where N_e is the number of informative experiments and N_s^e is the number of samples taken during the eth experiment. Discrete instant t_k^e refers to the kth measurement time of the eth experiment. It is assumed that the initial concentrations of all the experiments are measured with special care, so that they can be regarded as perfectly known:

$$\xi_m\left(t_0^e\right) = \xi(0), \forall e. \qquad (5)$$

This assumption can eventually be relaxed by estimating initial conditions, together with the

unknown parameters of Equation (2). In our methodology, the distinction is made between three types of such parameters:

- M^*, the minimum number of reactions to be taken into account in order to represent the experimental data with a given accuracy;
- ϑ_K^*, the optimum stoichiometric parameters, which are the unknown elements of the yield matrix K; and
- ϑ_φ^*, the optimum kinetic parameters appearing in the reaction rate vector $\varphi(\xi)$.

It is assumed that the biological information available on the process encompasses the structure of the reaction rate vector. Also, full-state measurement is assumed at discrete times, so that a continuous, noise-reduced signal $\xi_m(t)$ is obtained via standard techniques of interpolation and linear filtering. Finally, the time-evolution of the dilution rate and external feed rates are regarded as perfectly known.

Number of Reactions

As described by Bernard and Bastin (2005a, 2005b), the number of reactions required for the model to reproduce the observations with a given error threshold can be assessed by a principal component analysis of the interpolated datasets. Their estimation approach is briefly outlined in what follows.

After placing the measured terms on the left-hand side of the equation, Equation (2) can be written as follows for every measurement instant:

$$\frac{d\xi(t)}{dt}\bigg|_{(t=t_k^e)} + D\left(t_k^e\right)\xi\left(t_k^e\right) - F\left(t_k^e\right) = K\varphi\left(\xi\left(t_k^e\right)\right). \quad (6)$$

The interpolation and linear filtering of both sides of this expression gives

$$u(t) = Kw(t), \qquad (7)$$

which can be written in matrix form after sorting column-wise the vectors related to the various instants:

$$U = KW. \qquad (8)$$

Using Equation (3), together with the fact that matrix W has full row rank by construction, linear matrix algebra states that the square matrix UU^T has rank M. The eigenvalue decomposition of this matrix then gives

$$UU^T = P^T \Sigma P, \tag{9}$$

where Σ is a diagonal, positive semi-definite matrix whose nonzero elements represent the variance related to each eigenvector, i.e., each column of matrix P. The first stage of the methodology thus consists of selecting the set of the M largest eigenvalues that represent a total variance larger than a given confidence threshold.

Identification of C-Identifiable Schemes

A systematic, data-driven approach was suggested by Hulhoven et al. (2005) to identify those reaction schemes that satisfy a mathematical property known as C-identifiability. This procedure relies on a state transformation operated via the following partition of the pseudo-stoichiometric matrix:

$$K = \begin{pmatrix} K_a \\ K_b \end{pmatrix}, \quad \text{with} \begin{cases} K_a \in \mathbb{R}^{M \times M} \\ K_b \in \mathbb{R}^{(N-M) \times M} \end{cases}. \tag{10}$$

The matrix C is chosen such that

$$CK_a + K_b = 0, \tag{11}$$

which makes the dynamics of the transformed state vector $z = C\xi_a + \xi_b$ independent of the reaction rate vector:

$$\frac{dz(t)}{dt} = -D(t)z(t) + CF_a(t) + F_b(t). \tag{12}$$

Using Equation (12), the measurement of $z(t)$ and the knowledge of the transport terms enable the estimation of C independently of the reaction kinetics (Bogaerts et al., 2003). The property of C-identifiability denotes those schemes whose pseudo-stoichiometric matrix can be computed as the unique solution of Equation (11) when an estimate of C is known (Chen and Bastin, 1996).

The identification approach recommended by Hulhoven et al. (2005) ensures model C-identifiability by systematically screening those matrices K_a that have the following specific form:

$$K_a = \begin{pmatrix} \pm 1 & 0 & \dots & 0 \\ 0 & \pm 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \pm 1 \end{pmatrix}. \tag{13}$$

By determining the value of K_a that offers the smallest cost after identification of its corresponding matrix C, an estimate of the whole pseudo-stoichiometric matrix is deduced from Equation (11). This identification procedure is repeated for every possible partition $[\xi_a, \xi_b]^T$ of the state vector and the pseudo-stoichiometric matrix leading to the smallest cost function is eventually adopted.

The second step of our methodology makes the most of the observation that real bioprocess models are usually not C-identifiable because their submatrix K_a is rarely of the form (13). Rather than spending computational effort in screening all the possible values of K_a suggested by (13), which would lead anyway to reaction schemes lacking biological interpretation, our approach keeps K_a equal to identity the first time. Doing so speeds up the algorithm by restricting the screening of many candidate matrices K_a to the single choice of the best state partition. It is important to note that the issue of conferring biological interpretability to the model is voluntarily left to the last stage of the methodology.

Subsequently, the assumption is made that an estimate $\hat{C}$ of matrix C has been computed independently of the reaction rates (Bogaerts et al., 2003), so that the pseudo-stoichiometric matrix of the following C-identifiable model involves an identity submatrix. Indeed, using $K_a = I_M$ with the estimate of C in Equation (11) gives $K_b = -\hat{C}$, so the partitioned version of (2) becomes:

$$\frac{d}{dt}\begin{pmatrix} \xi_a(t) \\ \xi_b(t) \end{pmatrix} = \begin{pmatrix} I_M \\ -\hat{C} \end{pmatrix} \varphi_C(\xi, \vartheta_K^*, \vartheta_\varphi^*) - \\ D(t)\begin{pmatrix} \xi_a(t) \\ \xi_b(t) \end{pmatrix} + \begin{pmatrix} F_a(t) \\ F_b(t) \end{pmatrix}. \tag{14}$$

In this expression, $\varphi_C(\xi, \vartheta_K^*, \vartheta_\varphi^*)$ is an unknown reaction rate vector that generally lacks a biological interpretation. For instance, elements of φ_C may become negative for some state values.

The concept of equivalence between reaction schemes can be exploited here to facilitate the biological interpretation of Equation (14) and it is therefore important to introduce it before going deeper into the last stage of the approach. Two reaction schemes (or their associated dynamic

models) are equivalent if, for any possible input, they describe identical time-trajectories of the system state (Delcoux et al., 2001). A necessary and sufficient condition for two dynamic models to be equivalent is the existence of a linear mapping that transforms one model into the other. To illustrate this, let $\mathcal{A}$ and $\mathcal{B}$ be two equivalent models. The equivalence between $\mathcal{A}$ and $\mathcal{B}$ can be highlighted via the respective replacement of $K_{\mathcal{A}}$ and $\varphi_{\mathcal{A}}$ in the dynamic expression of $\mathcal{A}$ by the following quantities:

$$K_{\mathcal{A}} = K_{\mathcal{B}}Q \quad \text{and} \quad \varphi_{\mathcal{A}} = Q^{-1}\varphi_{\mathcal{B}}, \tag{15}$$

where $Q \in \mathbb{R}^{M \times M}$ can be any regular transformation matrix. This illustrates the existence of an infinite number of equivalent models to represent the same time-evolution of the state.

By comparing equivalent models (2) and (14), and by taking (10) into account, the unknown vector φ_C can be expressed as a linear combination of the true reaction rate vector φ, in which the coefficients are the actual reaction yields of the components associated to the identity submatrix:

$$\varphi_C\left(\xi, \vartheta_K^*, \vartheta_\varphi^*\right) = K_a\left(\vartheta_{K_a}^*\right)\varphi\left(\xi, \vartheta_\varphi^*\right). \tag{16}$$

The only way to free up the biological interpretation of (16), i.e., to separate the influence of the reaction stoichiometry from its kinetics, is thus by identifying the remaining unknown parameters of the model, i.e., $\vartheta_{K_a}^*$ and ϑ_φ^*. In the last stage of our methodology, the submatrix K_a is therefore regarded as an unknown transformation matrix that maps the previously-identified, C-identifiable model (14) onto a biologically-consistent one.

Identification of Biologically Relevant Schemes

As explained in the previous section, the third stage of our identification procedure aims at determining a biologically-consistent scheme by identifying the remaining stoichiometric and kinetic parameters of a C-identifiable model in order to transform it. While the estimation approach is basically measurement-driven, the availability of biological information may strongly influence its convergence towards the optimum value of the parameters. Biological knowledge about the system should therefore be included into the problem in the form of mathematical constraints on the

stoichiometric and kinetic parameters, while kinetic considerations should, in addition, specify the structure of the reaction rate vector as accurately as possible. The use of too general forms of the kinetic functions can indeed lead to a lack of model identifiability.

The present identification stage is split up into two steps, namely a quick parameter assessment step followed by a general refinement of the estimates, e.g., in the maximum-likelihood sense. Since parameter identification of dynamic models is, in general, a well-documented approach (Walter and Pronzato, 1997), the present paper focuses instead on the description of the quick assessment approach.

After substituting Equation (16) into Equation (14) and gathering the measured terms on the left-hand side of the equation, the following expression is obtained at every measurement instant:

$$\left.\frac{d\xi(t)}{dt}\right|_{(t=t_k^e)} + D\left(t_k^e\right)\xi(t_K^e) - F(t_k^e) =$$

$$\begin{pmatrix} I_M \\ -\hat{C} \end{pmatrix}K_a\left(\vartheta_{K_a}^*\right)\varphi\left(\xi(t_k^e), \vartheta_\varphi^*\right). \tag{17}$$

The interpolation and linear filtering of both sides of this equation gives

$$u(t) = \begin{pmatrix} I_M \\ -\hat{C} \end{pmatrix}K_a\left(\vartheta_{K_a}^*\right)w\left(t, \vartheta_\varphi^*\right), \tag{18}$$

which can be written in matrix form after column-wise sorting the vectors related to the various instants:

$$U = \begin{pmatrix} I_M \\ -\hat{C} \end{pmatrix}K_a\left(\vartheta_{K_a}^*\right)W(\vartheta_\varphi^*). \tag{19}$$

In the absence of any stoichiometric constraint, a least-squares estimator of the pseudo-stoichiometric submatrix K_a is easily derived from (19). Its value varies as a function of the kinetic parameters since it depends on matrix $W(\vartheta_\varphi)$:

$$\left.\hat{C}_a\right|_{(\vartheta_\varphi = \hat{\vartheta}_\varphi)} = \left(I_M + \hat{C}^T\hat{C}\right)^{-1}\begin{pmatrix} I_M \\ -\hat{C} \end{pmatrix}^T UW(\hat{\vartheta}_\varphi)^T$$

$$\left(W(\hat{\vartheta}_\varphi)W(\hat{\vartheta}_\varphi)^T\right)^{-1}. \tag{20}$$

In the presence of biological information on the stoichiometry, the estimation of K_a for a given value $\hat{\vartheta}_\varphi$ of the kinetic parameters can be done by solving a quadratic optimization problem of order M^2 with linear constraints.

The quick assessment approach simply consists of a possibly constrained, nonlinear optimization problem, in which the kinetic parameters are adjusted so as to minimize the following cost function derived from (19) and (20):

$$\mathcal{J}\left(\vartheta_\varphi\right) = \sum_{\forall i, \forall j} \left[U - \begin{pmatrix} I_M \\ -\hat{C} \end{pmatrix} \hat{K}_a \bigg|_{(\vartheta_\varphi = \hat{\vartheta}_\varphi)} W\left(\vartheta_\varphi\right) \right]^2_{(i,j)} . \quad (21)$$

This parameter assessment approach is expected to work faster than classical identification methods, since it requires no integration of the model to evaluate the cost function. However, the harmful effect of measurement noise on the convergence of the optimization algorithm towards the true parameter values is not taken into account by this quick step. It may thus be necessary to reinject the set of computed estimates to improve the quality of the initial guess for a more accurate identification procedure. This results in a lower bias of the estimates, as well as in a decreased estimation time.
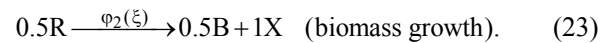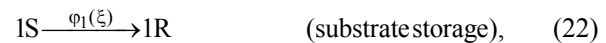
SIMULATION EXAMPLE

The continuous cultivation of microalgae in a photobioreactor is investigated as a simulation example to illustrate the performances of the last stage of our identification procedure. As previously mentioned, this stage is split up into a quick parameter assessment, requiring no integration of the model equations, followed by a more classical, nonlinear identification of the whole parameter set devoted to bias reduction. The usefulness of the quick assessment step is clearly highlighted by the simulation results.

Macroscopic Model

Plant cells generally behave differently from animal cells. Indeed, rather than directly depending on the external substrate concentrations, their growth rate varies as a function of the concentration of intracellular nutrients. The dynamic model considered in this example was developed by Lemesle and Mailleret (2008) as a mechanistic

reformulation of the well-known Droop model that represents the limitation of growth by a single substrate.

The scheme at the root of this model relies on two macroscopic reactions. The first one is the absorption and storage of extracellular substrate S into the algal cell, while the second one represents the transition from the pool of the stored nutrients R to the pool of the metabolized nutrients B, producing along the way a certain amount of biomass X. For the sake of ease, capital letters will refer to both the component and to its concentration. The reaction scheme can be written as follows:

$$1S \xrightarrow{\varphi_1(\xi)} 1R \qquad\qquad (\text{substrate storage}), \quad (22)$$

$$0.5R \xrightarrow{\varphi_2(\xi)} 0.5B + 1X \quad (\text{biomass growth}). \quad (23)$$

As usual, the reaction rates of the system are expressed as nonlinear functions of its state variables. In this model, Monod-like reaction rates were chosen:

$$\varphi_1\left(\xi\right) = \frac{S}{15 + S} X \quad \text{and} \quad \varphi_2\left(\xi\right) = \frac{R/X}{1 + R/X} X. \quad (24)$$

In the case of a continuous culture of microalgae with dilution rate D and substrate input concentration S_{in}, a mass balance applied to the macroscopic scheme gives the following matrix system of ordinary differential equations:

$$\frac{d}{dt} \begin{pmatrix} S \\ R \\ B \\ X \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 1 & -0.5 \\ 0 & 0.5 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \dfrac{S}{15 + S} X \\ \dfrac{R/X}{1 + R/X} X \end{pmatrix} - D \begin{pmatrix} S - S_{in} \\ R \\ B \\ X \end{pmatrix}, \quad (25)$$

which will then be numerically integrated to simulate laboratory experiments.

Identification Problem

The identification problem focuses on the last stage of the methodology, i.e., the one that confers on the model its biological interpretability. More precisely, it aims at illustrating the advantages of using a quick assessment of the parameters as an initial guess for the global, nonlinear optimization. For this purpose, both stoichiometric and kinetic parameters are regarded as unknown. In order to keep the algorithm comparison free from model

errors, the structure of the reaction rate vector is assumed to be perfectly known. It is also considered that an equivalent C-identifiable model of the form (14) is known in advance, where matrix C is derived from (11) with known K_a and K_b:

$$\frac{d}{dt}\begin{pmatrix} S \\ R \\ B \\ X \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \\ -2 & -2 \end{pmatrix}\begin{pmatrix} \vartheta^*_{K_a,1} & \vartheta^*_{K_a,2} \\ \vartheta^*_{K_a,3} & \vartheta^*_{K_a,4} \end{pmatrix}$$

$$\begin{pmatrix} \dfrac{S}{\vartheta^*_{\varphi,1}+S}X \\ \dfrac{R/X}{\vartheta^*_{\varphi,2}+R/X}X \end{pmatrix} - D\begin{pmatrix} S-S_{in} \\ R \\ B \\ X \end{pmatrix}. \tag{26}$$

In this expression, the optimum stoichiometric and kinetic parameter vectors are respectively defined as follows:

$$\vartheta^*_{K_a} = \begin{pmatrix} -1 \\ 0 \\ 1 \\ -0.5 \end{pmatrix} \quad \text{and} \quad \vartheta^*_{\varphi} = \begin{pmatrix} 15 \\ 1 \end{pmatrix}. \tag{27}$$

In order to generate data for parameter identification, four experiments are simulated with a constant dilution rate ($D = 0.1$) and a constant input concentration ($S_{in} = 10$), starting from the following known initial conditions:

$$\xi(t_0^1) = \begin{pmatrix} 2 \\ 0.75 \\ 5.9 \\ 3 \end{pmatrix}, \xi(t_0^2) = \begin{pmatrix} 30 \\ 0.25 \\ 5.3 \\ 2.6 \end{pmatrix},$$

$$\xi(t_0^3) = \begin{pmatrix} 10 \\ 0.5 \\ 2.6 \\ 1.3 \end{pmatrix}, \text{and } \xi(t_0^4) = \begin{pmatrix} 4 \\ 1 \\ 2.2 \\ 1 \end{pmatrix}. \tag{28}$$

Each experiment lasts 20 days and the whole state vector is measured daily. Measurements are affected by a zero-mean, Gaussian noise, characterized by a relative standard deviation of $\sigma_r = 10\%$. The experimental datasets are systematically interpolated by cubic smoothing splines via the *csaps* function of Matlab, choosing $p = 0.15$ as the smoothing parameter (Fig. 1).

The experimental data and their interpolation are then treated by the two-step stage of our identification procedure without any extra filtering. In order to make performance comparisons, these data are also treated with a classical optimization algorithm that attempts to identify all the unknown parameters in a single step.
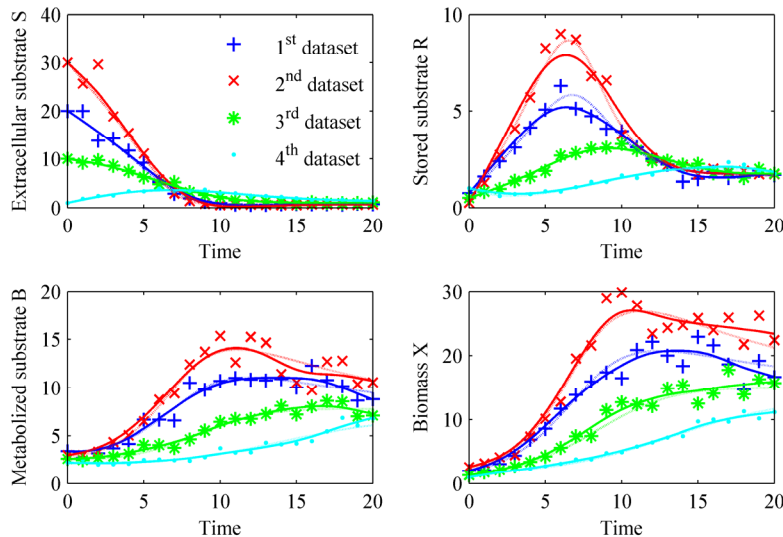

Figure 1: The four noisy experimental datasets (symbols) and their continuous interpolations (solid lines), together with their originating simulation curves (dotted line)

Benchmark Statement and Results

Initial guesses of the parameters are randomly generated at increasing distances from their optimum value ϑ^* through the following formula:

$$\vartheta_0 = \vartheta^* (1 + \Delta\vartheta\alpha), \tag{29}$$

where α is a standard normally distributed random vector and $\Delta\vartheta$ is an initial level of uncertainty that successively takes 1000 increasing values that are equally-spaced over the range $[0, 1.5[$. Numerical problems are prevented by simply forcing the positivity of the initial guesses of the kinetic parameters.

Since no prior information is available on the stoichiometry of the system, the stoichiometric matrix K_a is estimated through (20). All the optimizations are performed with the Nelder-Mead simplex algorithm implemented in the Matlab function *fminsearch*. This optimizer indeed offers the advantage of being more robust than classical gradient-based algorithms with regard to the presence of local minima in the cost function.

In order to illustrate the performances of our identification procedure, the distance that separates the estimates from the true parameter value has been computed and is represented as histograms in Fig. 2. For both approaches, i.e., the direct identification of all the parameters and the two-step procedure, the following relative Euclidian distance is chosen:

$$d(\vartheta, \vartheta^*) = \sqrt{\sum_i \left(\frac{\vartheta_i - \vartheta_i^*}{\vartheta_i^*} \right)}. \tag{30}$$

While the estimates resulting from the direct identification approach are widely distributed over the range of relative distances, the quick assessment step visibly has a gathering effect on their distribution. This highlights the presence of a strongly-attractive, global minimum in the cost function (21), which leads to convergence even from poor initial guesses. The second step of our procedure then clearly refines the result by removing the remaining bias from the solution vector. Fig. 3 confirms this explanation by showing the influence of the initial uncertainty level on the optimality of the estimates for both approaches.

Since it operates on a restricted set of parameters and requires no numerical integration of the model differential equations, the quick assessment step usually converges within a fraction of a second. This is extremely fast in comparison with the seconds or minutes required for other classical optimization methods to converge. A plot of the computation time as a function of the initial uncertainty level illustrates these observations for both identification approaches (Fig. 4).
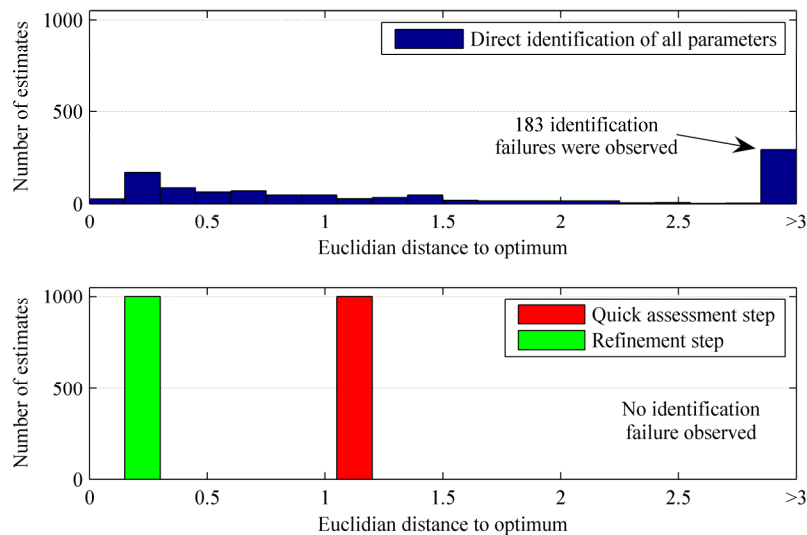

Figure 2: Repartition of the relative Euclidian distances between the estimates and their optimum value, over a set of one thousand identification runs by the direct and the two-step methods.
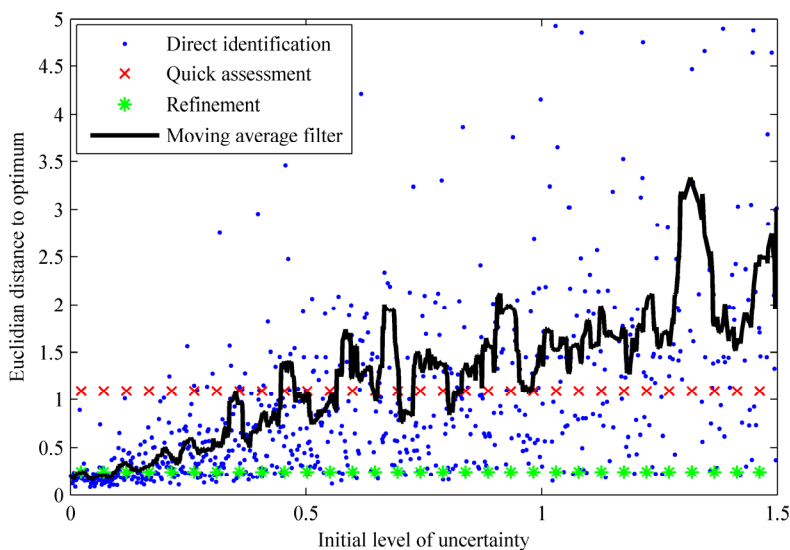
Figure 3: Evolution of the distance separating the estimates from their optimum value as a function of the initial uncertainty level $\Delta\vartheta$ of the parameters, for both identification approaches. A moving average filter with a 20-sample time window shows the increasing tendency for the direct approach.

Figure 4: Evolution of the computation time as a function of the initial uncertainty level $\Delta\vartheta$ of the parameters, for both identification approaches. A moving average filter with a 20-sample time window shows the increasing tendency for the direct approach.
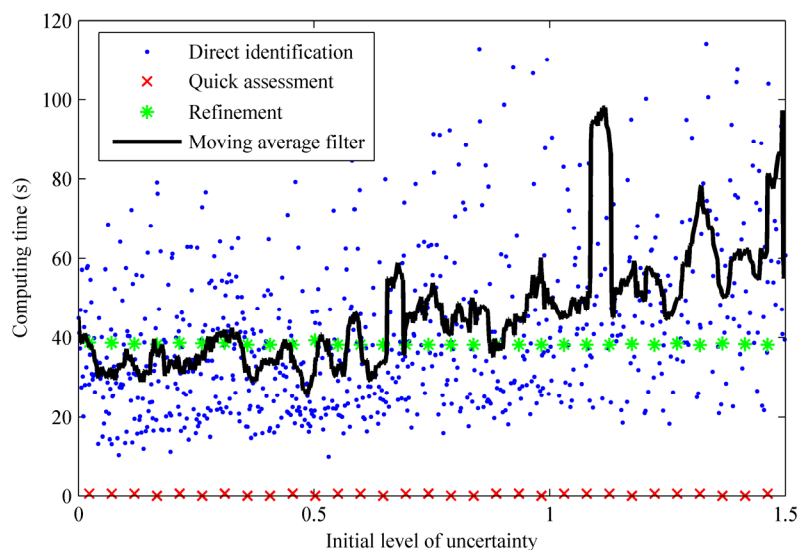
CONCLUSIONS

The systematic identification procedure presented in this paper uses and extends previous research results to the transformation of C-identifiable reaction schemes into biologically-consistent schemes. A strong point of the new approach lies in the quick assessment step of the transformation stage that has a fast computation time because it integrates no differential equation and operates on only a limited number of parameters, i.e., the set of the kinetic parameters. A simulation example of

continuous culture of microalgae highlights the robustness of the new procedure with regard to uncertainties in the initial guess.

Further research is still required, though, to characterize better the influence of biological constraints on the convergence of the identification algorithm. In particular, the choice of an appropriate structure for the expression of the reaction rates is a challenging issue that could still improve the applicability of our methodology.

ACKNOWLEDGEMENTS

NOMENCLATURE

$\mathcal{A}, \mathcal{B}$	Dynamic models of bioprocesses
B, B	Metabolized substrate (concentration)
C	Pseudo-stoichiometric submatrix that is part of a C-identifiable scheme
$\widehat{C}$	Estimation of matrix C
$d(\vartheta, \vartheta^*)$	Relative Euclidian distance between parameter vectors ϑ and ϑ^*
D	Input dilution rate
F	Vector of external feed rates
F_a, F_b	Subvectors of external feed rates
I_M	Identity matrix of size M
$\mathcal{J}$	Cost function
K	Pseudo-stoichiometric matrix
K_a, K_b	Pseudo-stoichiometric submatrices
$\widehat{K}_a$	Estimation of the pseudo-stoichiometric submatrices
$K_{\mathcal{A}}$	Pseudo-stoichiometric matrix of model $\mathcal{A}$
$K_{\mathcal{B}}$	Pseudo-stoichiometric matrix of model $\mathcal{B}$
M	Number of macroscopic reactions
M^*	Optimum number of macroscopic reactions
N	Number of key-components
N^e	Number of experiments
N_s^e	Number of samples taken during the eth experiment
p	Smoothing parameter for Matlab function csaps
P	Matrix of eigenvectors
$\mathcal{P}_k$	Set of the products of the kth reaction
Q	Regular transformation matrix
R, R	Intracellular substrate (concentration)
$\mathcal{R}_k$	Set of the reactants of the kth reaction
S, S	Extracellular substrate (concentration)
S_{in}	Substrate feed concentration
t_k^e	k^{th} time instant of the eth experiment
t_0^e	Initial time of the eth experiment
u(t)	Measured vector interpolated and filtered at time t
U	Matrix of measured vectors
w(t)	Non-measured vector interpolated and filtered at time t
W	Matrix of non-measured vectors
X, X	Biomass (concentration)
z	Transformed state vector
α	Standard normally distributed random vector
$\Delta\vartheta$	Initial level of uncertainty of the parameter vector
ϑ	Parameter vector
ϑ^*	Optimum parameter vector
ϑ_i	i^{th} element of vector ϑ
ϑ_K^*	Vector of the optimum stoichiometric parameters
ϑ_{K_a}	Vector of the parameters related to K_a
$\vartheta_{K_a}^*$	Vector of the optimum parameters related to K_a
$\vartheta_{K_a,i}^*$	i^{th} element of $\vartheta_{K_a}^*$
ϑ_0	Initial guess of the parameter vector
ϑ_φ	Vector of the kinetic parameters
$\widehat{\vartheta}_\varphi$	Estimation of the vector of the kinetic parameters
ϑ_φ^*	Vector of the optimum kinetic parameters
$\vartheta_{\varphi,i}^*$	i^{th} element of ϑ_φ^*
ν_{*k}	Pseudo-stoichiometric coefficient of the kth reaction
ξ	State vector composed of the reacting species concentrations
ξ_a, ξ_b	State subvectors of the reacting species concentrations
ξ_i	i^{th} element of the state vector
σ_r	Relative standard deviation
Σ	Diagonal matrix of eigenvalues
φ	Reaction rate vector

$\varphi_{\mathcal{A}}$	Reaction rate vector of model $\mathcal{A}$
$\varphi_{\mathcal{B}}$	Reaction rate vector of model $\mathcal{B}$
φ_C	Reaction rate vector of a C-identifiable scheme
φ_k	k^{th} element of the reaction rate vector

REFERENCES

Bastin, G. and Dochain, D., On-line Estimation and Adaptive Control of Bioreactors. Elsevier Science Ltd, New York (1990).

Bernard, O. and Bastin, G., On the estimation of the pseudo-stoichiometric matrix for macroscopic mass balance modelling of biotechnological processes. Mathematical Biosciences, 193, No. 1, 51 (2005a).

Bernard, O. and Bastin, G., Identification of reaction networks for bioprocesses: determination of an incompletely known pseudo-stoichiometric matrix. Bioprocess and Biosystems Engineering, 27, No. 5, 293 (2005b).

Bogaerts, Ph., Delcoux, J.-L. and Hanus, R., Maximum likelihood estimation of pseudo-stoichiometry in macroscopic biological reaction schemes, Chemical Engineering Science, 58, No. 8, 1545 (2003).

Bogaerts, Ph., Rooman, M., Vastemans, V. and Vande Wouwer, A., Determination of macroscopic reaction schemes: towards a unifying view. Proceedings of the 17th World Congress of IFAC. Seoul (2008).

Chen, L. and Bastin, G., Structural identifiability of the yield coefficients in bioprocess models when the reaction rates are unknown. Mathematical Biosciences, 132, No. 8, 35 (1996).

Delcoux, J.-L., Hanus, R. and Bogaerts, Ph., Equivalence of reaction schemes in bioprocesses modelling. Proceedings of the 4th International Symposium on Mathematical Modelling and Simulation in Agricultural and Bio-Industries. Haifa (2001).

Hulhoven, X., Vande Wouwer, A. and Bogaerts, Ph., On a systematic procedure for the predetermination of macroscopic reaction schemes. Bioprocess and Biosystems Engineering, 27, No. 5, 283 (2005).

Lemesle, V. and Mailleret, L., A mechanistic investigation of the algae growth "Droop" model. Acta Biotheoretica, 56, No. 1-2, 57 (2008).

Walter, E. and Pronzato, L., Identification of parametric models from experimental data. Springer Verlag, Berlin-Heidelberg (1997).