10th International IFAC Symposium on
Computer Applications in Biotechnology
Preprints Vol.2, June 4-6, 2007, Cancún, Mexico

A GENERAL KINETIC MODEL STRUCTURE SIMULATION AND EXPERIMENTAL VALIDATION

Aline Grosfils[a], Alain Vande Wouwer[b] and Philippe Bogaerts[a]

[a]*Service de Chimie générale et Biosystèmes, Université Libre de Bruxelles, Belgium,*
[b]*Service d'Automatique, Faculté Polytechnique de Mons, Belgium*

Abstract: In this study, a general kinetic model structure is proposed, which can describe various effects such as activation, inhibition and saturation. Its main advantage lies in an associated identification procedure allowing identifiability problems to be (at least partly) alleviated. The usefulness of the proposed model is tested with simulated as well as with experimental data.

Keywords: biological systems, mathematical modelling, identification, nonlinear systems.

1. INTRODUCTION

Macroscopic models of bioprocesses consist of a system of mass balances for macroscopic species (biomass, main substrates and products of interest) involved in a reaction scheme describing the main phenomena occurring in the cell culture (Bastin and Dochain, 1990).

The derivation of a biologically sound model implies the *a priori* selection of a reaction scheme and a kinetic model structure. In recent years, several methods for the determination of macroscopic reaction schemes have been proposed. A first category of methods is based on model reduction procedures and start from a detailed metabolic network of intracellular reaction pathways (Haag *et al*, 2005; Provost and Bastin, 2004). If no information about metabolic pathways is available, a second category of methods attempt to directly determine a macroscopic reaction scheme linking the (external) substrates to the products of the reactions (Hulhoven *et al*., 2005; Bernard and Bastin, 2005).

Regarding the kinetic model structure, its choice is made particularly difficult by the lack of detailed knowledge about the cell culture under consideration and the profusion of kinetic laws, which can be of various types, ranging from biologically-inspired to purely black-box models.

The most widespread black-box models are the artificial neural networks (Oliveira, 2004). These models are often preferred when bioprocess knowledge is missing. However, the selection of the appropriate type of neural networks and the choice of the network size are difficult tasks.

There is a vast array of biologically-inspired kinetic laws, among which it is not always easy to discriminate the appropriate model structure for the problem under consideration. These laws are nonlinear and, in most of cases, non linearizable (with respect to the parameters), so that time-consuming optimization and local optima are common problems.

In order to keep the physical interpretation of the parameters while avoiding the classical problems associated with biologically-inspired laws, (Bogaerts *et al*., 1999) have proposed a kinetic model structure allowing the description of activation and /or inhibition effects of macroscopic species in the culture. This structure is linearizable with respect to the parameters, a property which allows the determination of unique estimates through the solution of a least-squares problem.
The objectives of this study are to further generalize this model structure to include saturation effects (besides activation and inhibition), and to carefully test the generalized model structure with simulated and real-life experimental data. Especially, attention is focused on a systematic identification procedure, alleviating identifiability problems.

2. BIOPROCESS MODELING

A general approach to describe the dynamics of a bioprocess has been proposed in (Bastin and Dochain, 1990). It consists of the system of mass balances for the macroscopic species involved in a reaction scheme. The expression of such a reaction scheme is the following:

$$\sum_{i\in R_k}(-\nu_{i,k})\,\xi_i \xrightarrow{\varphi_k} \sum_{j\in P_k}\nu_{j,k}\,\xi_j \qquad k\in[1,M] \tag{1}$$

where M is the number of reactions, φ_k the kth reaction rate, ξ_i the ith component, $\nu_{i,k}$ the corresponding pseudo-stoichiometric (or yield) coefficients (positive when associated to a component which is produced, negative when it is consumed), R_k the kth set of reactants and catalysts indices and P_k the kth set of products indices.

The system of mass balances for each of the N components ξ_i can be written in the following matrix form:

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

where $\xi\in\Re^N$ is the vector of concentrations, $\mathbf{K}\in\Re^{\mathbf{NxM}}$ is the pseudo-stoichiometric coefficients matrix ($N\geq M$), $\boldsymbol{\varphi}\in\Re^M$ is the vector of reaction rates, $D\in\Re$ is the dilution rate, $\mathbf{F}\in\Re^N$ is the vector of external feed rates, $\mathbf{Q}\in\Re^N$ is the vector of gaseous outflow rates.

In (Bogaerts *et al.*, 1999) a kinetic model structure allowing the representation of the activation and inhibition effects is proposed:

$$\boldsymbol{\varphi}_j(\xi_1,\ldots,\xi_N) = \boldsymbol{\alpha}_j \prod_{h\in R_j^*}\xi_h^{\gamma_{hj}}(t)\prod_{l\in I}e^{-\boldsymbol{\beta}_{lj}\xi_l(t)} \qquad j\in[1,M] \tag{3}$$

where $\boldsymbol{\alpha}_j>0$ is a kinetic constant, $\boldsymbol{\gamma}_{hj}>0$ the activation coefficient associated to the component h in reaction j, $\boldsymbol{\beta}_{lj}\geq 0$ the inhibition coefficient of the component l in reaction $j\in[1,M]$. I is the set of all the components indices while R_j^* is the set of reactants, catalysts and auto-catalysts components.

The main advantage of this structure is the possibility to develop a systematic identification procedure, which proceeds in 2 steps.

The first step is a unique least squares estimation based on the linearization of the kinetic model structure (3)

$$\ln\boldsymbol{\varphi}_j(t_{s,k}) = \ln\boldsymbol{\alpha}_j + \sum_{h\in R_j^*}(\ln\xi_h(t_{s,k}))\,\boldsymbol{\gamma}_{hj} - \sum_{l\in I}\xi_l(t_{s,k})\boldsymbol{\beta}_{lj} \tag{4}$$

where estimates $\hat{\boldsymbol{\varphi}}_j(t_{s,k})$ can be obtained by differentiating smoothing splines applied to the measured concentration vector $\hat{\xi}$.

In the second step, the kinetic coefficients together with the initial conditions are identified starting from the initial estimates determined in the previous step. This step relies on the complete simulation model {(2), (3)} and a maximum-likelihood estimator taking the measurement errors into account (Bogaerts *et al.*, 2003).

3. A GENERAL KINETIC MODEL

The kinetic model structure (3) is not able to represent saturation effects. The model is forced to mimic this behaviour by an inhibition compensating a stronger activation. This, of course, alters the physical interpretation of the parameters. To represent saturation by a culture component, the idea is to replace the activation function $\xi_h^{\gamma_{hj}}$ by $\left(1-e^{-\boldsymbol{\kappa}_{hj}\xi_h(t)}\right)^{\gamma_{hj}}$, where $\boldsymbol{\kappa}_{hj}\geq 0$ is the saturation coefficient of the component h in reaction j. The kinetic model structure therefore becomes

$$\boldsymbol{\varphi}_j(\xi,t) = \boldsymbol{\alpha}_j^* \prod_{h\in R_j^*}\left(1-e^{-\boldsymbol{\kappa}_{hj}\xi_h(t)}\right)^{\gamma_{hj}}\prod_{l\in I}e^{-\boldsymbol{\beta}_{lj}\xi_l(t)} \qquad j\in[1,M] \tag{5}$$

where $\boldsymbol{\alpha}_j^*>0$ is a kinetic constant, $\boldsymbol{\gamma}_{hj}\geq 0$ the activation coefficient associated to the component h in reaction j, $\boldsymbol{\kappa}_{hj}\geq 0$ the saturation coefficient of the component h in reaction j, and $\boldsymbol{\beta}_{lj}\geq 0$ the inhibition coefficient of the component l in reaction j.

Note that the original activation function $\xi_h^{\gamma_{hj}}$ is actually a particular case of the generalized structure. Indeed, when the saturation coefficient is small ($\boldsymbol{\kappa}_{hj}<<$), the Taylor series expansion of $\left(1-e^{-\boldsymbol{\kappa}_{hj}\xi_h(t)}\right)$ around 0 and limited to the first order, is given by

$$\left(1-e^{-\boldsymbol{\kappa}_{hj}\xi_h(t)}\right)\approx\boldsymbol{\kappa}_{hj}\xi_h(t) \tag{6}$$

Hence, when $\boldsymbol{\kappa}_{hj}\xi_h<<$, the generalized kinetic model can be expressed as follows:

$$\boldsymbol{\varphi}_j(\xi,t)\approx\boldsymbol{\alpha}_j^*\prod_{h\in R_j^*}\left(\boldsymbol{\kappa}_{hj}\xi_h(t)\right)^{\gamma_{hj}}\prod_{l\in I}e^{-\boldsymbol{\beta}_{lj}\xi_l(t)} \tag{7}$$

$$\text{or}\quad \boldsymbol{\varphi}_j(\xi,t)\approx\left(\boldsymbol{\alpha}_j^*\prod_{h\in R_j^*}\left(\boldsymbol{\kappa}_{hj}\right)^{\gamma_{hj}}\right)\prod_{h\in R_j^*}\xi_h^{\gamma_{hj}}(t)\prod_{l\in I}e^{-\boldsymbol{\beta}_{lj}\xi_l(t)} \tag{8}$$

which corresponds to the expression (5) where

$$\boldsymbol{\alpha}_j\approx\boldsymbol{\alpha}_j^*\prod_{h\in R_j^*}\left(\boldsymbol{\kappa}_{hj}\right)^{\gamma_{hj}} \tag{9}$$

This property is used in the following section in order to initialize the identification procedure, i.e., to systematically determine a first estimation of the parameters.

4. IDENTIFICATION PROCEDURE

Besides the ability of model structure (5) to describe various effects (activation, inhibition, saturation), it is

desirable to establish a systematic parameter identification procedure (in constrast with Monod-like or neural network models, for which such a procedure is quite difficult to suggest). The idea is to build upon the procedure described in Section 2, and originally developed for model (3).

If we first assume weak saturation effects ($\boldsymbol{\kappa}_{hj} <<$), the generalized kinetic model structure (5) can be approached by the original model structure (3). Hence, the procedure of Section 2 can be used to determine a first estimation of the original parameters $\boldsymbol{\alpha}_j, \boldsymbol{\gamma}_{hj}, \boldsymbol{\beta}_{lj}$. Based on these parameter values and a selection of values for $\boldsymbol{\kappa}_{hj}$, the parameters $\boldsymbol{\alpha}_j^*$ can be computed according to (9). Starting from these initial parameter estimates (which hopefully, are not so far away from the optimum), nonlinear identification of the parameters $\boldsymbol{\alpha}_j^*, \boldsymbol{\gamma}_{hj}, \boldsymbol{\beta}_{lj}$. and $\boldsymbol{\kappa}_{hj}$ as well as the initial component concentrations (initial conditions of the mass balance equations) can then be achieved.

The next subsections describe in more details these several estimation steps.

4.1 Initial estimation

The kinetic model structure (3) can be linearized w.r.t. its parameters thanks to a logarithmic transformation (4). The resulting formulation allows a linear least squares estimation of the kinetic coefficients (which necessarily exists, is unique and independent of any initial guess):

$$\hat{\boldsymbol{\theta}}_{cin}^{(j)} = \underset{\boldsymbol{\theta}_{cin}^{(j)}}{ArgMin} \frac{1}{2} \sum_{s=1}^{S} \sum_{k=1}^{N_s} \left(\mathbf{Y}_{m,s,k}^{(j)} - \boldsymbol{\varphi}_{s,k}^{(j)T} \boldsymbol{\theta}_{cin}^{(j)} \right)^2 \tag{10}$$

where

$$\mathbf{Y}_{m,s,k}^{(j)} = \ln \hat{\boldsymbol{\varphi}}_j(t_{s,k}) \tag{11}$$

$$\boldsymbol{\varphi}_{s,k}^{(j)T} = \begin{bmatrix} 1 & \ln \boldsymbol{\xi}_{h\ldots}(t_{s,k}) & -\boldsymbol{\xi}_{l\ldots}(t_{s,k}) \end{bmatrix} \tag{12}$$

$$\boldsymbol{\theta}_{cin}^{(j)T} = \begin{bmatrix} \ln \boldsymbol{\alpha}_j & \boldsymbol{\gamma}_{h\ldots j} & \boldsymbol{\beta}_{l\ldots j} \end{bmatrix} \tag{13}$$

$$\text{and under the constraints } \begin{bmatrix} \boldsymbol{\gamma}_{h\ldots j} & \boldsymbol{\beta}_{l\ldots j} \end{bmatrix} \geq 0 \tag{14}$$

To estimate the reaction rate vector $\hat{\boldsymbol{\varphi}}(t_{s,k})$, it is useful to partition the vector $\boldsymbol{\xi}^T = \begin{bmatrix} \boldsymbol{\xi}_a^T & \boldsymbol{\xi}_b^T \end{bmatrix}$, so that the corresponding partition of $\mathbf{K}^T = \begin{bmatrix} \mathbf{K}_a^T & \mathbf{K}_b^T \end{bmatrix}$ involves a matrix $\mathbf{K}_a \in \Re^{MxM}$ of full rank ($\text{rank}(\mathbf{K}_a) = M$, assuming $\text{rank}(\mathbf{K}) = M$). On the basis of this partition, $\hat{\boldsymbol{\varphi}}(t_{s,k})$ can be computed as

$$\hat{\boldsymbol{\varphi}}(\boldsymbol{\xi}(t_{s,k})) = \hat{\mathbf{K}}_a^{-1} \left(\left(\frac{d\boldsymbol{\xi}_a(t_{s,k})}{dt} \right)^{\wedge} + D(t_{s,k}) \boldsymbol{\xi}_a(t_{s,k}) - \mathbf{u}_a(t_{s,k}) \right) \tag{15}$$

where the estimate of the derivative $d\boldsymbol{\xi}_a(t_{s,k})/dt$ can be evaluated by numerical differentiation of a smoothing spline (generally speaking, an interpolation model taking the measurement noise into account) of the vector $\boldsymbol{\xi}_a(t)$.

4.2 A special treatment for $\boldsymbol{\kappa}_{hj}$ and $\boldsymbol{\alpha}_j^*$

As mentioned previously, $\boldsymbol{\kappa}_{hj}$ values have to be sufficiently small in order to allow the approximation of the activation function $\left(1 - e^{-\boldsymbol{\kappa}_{hj} \boldsymbol{\xi}_h(t)}\right)$ by its Taylor series expansion limited to the first order around 0 (6). Hence, we make an initial guess of these parameters by considering that we accept $(100-x)\%$ of disparity between the nonlinear activation function and its linear version. So, we compute $\boldsymbol{\kappa}_{hj}$ values on the basis of the following equation

$$\left(1 - e^{-\boldsymbol{\kappa}_{hj} \boldsymbol{\xi}_h^{\max}}\right) = \frac{x}{100} \quad \boldsymbol{\kappa}_{hj} \boldsymbol{\xi}_h^{\max} \tag{16}$$

Approaching $\left(1 - e^{-\boldsymbol{\kappa}_{hj} \boldsymbol{\xi}_h(t)}\right)$ by its Taylor series expansion limited to the second order around 0, we obtain

$$\boldsymbol{\kappa}_{hj} \boldsymbol{\xi}_h^{\max} - \frac{\left(\boldsymbol{\kappa}_{hj} \boldsymbol{\xi}_h^{\max}\right)^2}{2} \approx \frac{x}{100} \quad \boldsymbol{\kappa}_{hj} \boldsymbol{\xi}_h^{\max} \tag{17}$$

which leads to the following expression of $\boldsymbol{\kappa}_{hj}$

$$\boldsymbol{\kappa}_{hj} = \frac{2\left(1 - \frac{x}{100}\right)}{\boldsymbol{\xi}_h^{\max}} \tag{18}$$

In turn, a first estimation of $\boldsymbol{\alpha}_j^*$ can be deduced from (9):

$$\boldsymbol{\alpha}_j^* = \frac{\boldsymbol{\alpha}_j}{\prod_{h \in R_j^*} \left(\boldsymbol{\kappa}_{hj}\right)^{\gamma_{hj}}} \tag{19}$$

4.3 Markov Estimation

At this stage, a first estimation of all the parameters has been obtained. These parameter values are usually not accurate enough to be accepted as such, but can serve as starting point in a nonlinear identification procedure. This identification step concerns both the parameter set and the initial conditions of the experimental runs used to build the database. Indeed, these initial conditions are necessary to integrate in time the simulation model (2), (5), which consists of a nonlinear differential equation system of the form

$$\frac{d\mathbf{x}(t)}{dt} = f(\mathbf{x}(t), \mathbf{u}(t); \boldsymbol{\theta}) \tag{20}$$

where $\mathbf{x}^T(t) = \boldsymbol{\xi}^T(t)$ is the state vector containing the concentrations of the components involved in the reaction scheme; $\mathbf{u}^T(t) = \begin{bmatrix} D(t) & \mathbf{F}(t) & \mathbf{Q}(t) \end{bmatrix}$ is the input vector containing the dilution rate, the external feed rates and the gaseous outflow rates; $\boldsymbol{\theta}^T = \begin{bmatrix} \boldsymbol{\alpha}_j^* & \boldsymbol{\gamma}_{h\ldots j} & \boldsymbol{\beta}_{l\ldots j} & \boldsymbol{\kappa}_{h\ldots j} & \boldsymbol{\xi}^T(0) \end{bmatrix}$ $j \in [1, M]$ is the vector of the parameters to be identified (kinetic coefficients and initial concentrations); $\mathbf{f}$ is the model structure corresponding to relations (2)and (5). Let $\boldsymbol{\xi}(t) = g(t, \mathbf{u}(t), \mathbf{x}(t); \boldsymbol{\theta})$ be the solution (generally obtained by a numerical integration algorithm) of the

ordinary differential equation system (20) starting from the initial concentrations $\xi(0)$.

On the basis of sampled measurements

$$\mathbf{y}_{m,s,k} = g(t_{s,k}, \mathbf{u}(t_{s,k}), \mathbf{x}_s(0); \boldsymbol{\theta}) + \varepsilon_{y,s,k} \quad (21)$$

corrupted by a white measurement noise $\varepsilon_{y,s,k}$, normally distributed with zero mean and covariance matrix $\mathbf{Q}_{s,k}$, the maximum-likelihood estimate of $\boldsymbol{\theta}$ can then be deduced from a non linear Markov estimator

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{ArgMin} \frac{1}{2} \sum_{s=1}^{S} \sum_{k=1}^{N_s} \left(\mathbf{y}_{m,s,k} - g(t_{s,k}, \mathbf{u}(t_{s,k}), \mathbf{x}_s(0); \boldsymbol{\theta})\right)^T \quad (22)$$
$$\mathbf{Q}_{s,k}^T \left(\mathbf{y}_{m,s,k} - g(t_{s,k}, \mathbf{u}(t_{s,k}), \mathbf{x}_s(0); \boldsymbol{\theta})\right)$$

under the constraints $\hat{\boldsymbol{\theta}} \geq 0$ (23)

The initial guess of $\boldsymbol{\theta}$ consists, on the one hand, of the first estimate of the kinetic parameters deduced from the previous estimation steps and, on the other hand, of the measurements of ξ at the initial time (which are not "exact" since there are measurement errors).

The covariance matrix of the parameter estimation errors can also be estimated in this last step (Bogaerts *et al.*, 2003):

$$\hat{E}[\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^T] \approx \left(\sum_{s=1}^{S} \sum_{k=1}^{N_s} \mathbf{G}_\theta^T(t_{s,k}, \mathbf{u}(t_{s,k}), \mathbf{x}_s(0); \hat{\boldsymbol{\theta}}) \right.$$
$$\left. \mathbf{Q}_{s,k}^{-1} \, \mathbf{G}_\theta(t_{s,k}, \mathbf{u}(t_{s,k}), \mathbf{x}_s(0); \hat{\boldsymbol{\theta}}) \right)^{-1} \quad (24)$$

where

$$\mathbf{G}_\theta(t_{s,k}, \mathbf{u}(t_{s,k}), \mathbf{x}_s(0); \hat{\boldsymbol{\theta}}) = \left. \frac{\partial g(t_{s,k}, \mathbf{u}(t_{s,k}), \mathbf{x}_s(0); \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad (25)$$

This Jacobian is obtained by solving (together with the simulation model (21)) the sensitivity equations

$$\frac{\partial}{\partial t} \mathbf{G}_\theta(t, \mathbf{u}(t), \mathbf{x}(0); \boldsymbol{\theta}) = \quad (26)$$
$$\frac{\partial f(\mathbf{x}, \mathbf{u}(t); \boldsymbol{\theta})}{\partial \mathbf{x}} \mathbf{G}_\theta(t, \mathbf{u}(t), \mathbf{x}(0); \boldsymbol{\theta}) + \frac{\partial f(\mathbf{x}, \mathbf{u}(t); \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

with the initial condition

$$\mathbf{G}_\theta(0, \mathbf{u}(0), \mathbf{x}(0); \boldsymbol{\theta}) = \frac{\partial \mathbf{x}(0)}{\partial \boldsymbol{\theta}} = \mathbf{G}_{\theta 0} \quad (27)$$

where $\mathbf{G}_{\theta 0}$ is a matrix whose elements are all equal to zero except the ones corresponding to the partial derivative of the elements of $\mathbf{x}(0) \in \boldsymbol{\theta}$, these partial derivatives being equal to 1. The Jacobian $\mathbf{G}_\theta(t_{s,k}, \mathbf{u}(t_{s,k}), \mathbf{x}_s(0); \hat{\boldsymbol{\theta}})$ involved in relation (24) is thus obtained by evaluating the numerical solution $\mathbf{G}_\theta(t, \mathbf{u}(t), \mathbf{x}(0); \boldsymbol{\theta})$ of the system (20), (26) for $t = t_{s,k}$ and $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$.

At the end of this identification step, all the parameters have been estimated as accurately as possible but the model has to be validated through direct and cross validation tests and the study of the covariance matrix of the parametric errors. This latter study usually leads to parameters reduction. This is the subject of the following subsection.

4.4 Parameter reduction

The study of the covariance matrix (24) can help reducing the number of parameters. Indeed, hardly assessable coefficients can be suppressed after checking that their information is covered by other parameters. In concrete terms, coefficients with high variance and a sufficient correlation with other parameters can be cancelled out (the thresholds have to be chosen by the user). This cancellation reduces the number of parameters, and in turn the effect of a component (e.g. activation, saturation or inhibition).

However, considering the structure of the activation function $\left(1 - e^{-\kappa_{hj} \xi_h(t)}\right)^{\gamma_{hj}}$, the cancellation of a saturation coefficient implies the absence of activation. So, if a component activates a reaction with no saturation effect, the saturation coefficient should be cancelled out while keeping an activation effect. The easiest way to achieve this is to come back to the original model structure, i.e. to replace the activation/saturation function $\left(1 - e^{-\kappa_{hj} \xi_h(t)}\right)^{\gamma_{hj}}$ by the original activation function $\xi_h^{\gamma_{hj}}$.

However, all the parameters cannot be cancelled out without a careful check. Saturation and inhibition coefficients are usually strongly correlated. When saturation and inhibition parameters of a component display high variance and correlation, the most inaccurate parameter has to be eliminated.

The following parameter reduction procedure is therefore proposed:

- First, among all the above-mentioned parameters, the parameters with a variance and a covariance respectively superior to 10^3 and 10^6 are cancelled out.
- Then, the remaining parameters are re-estimated and the parameters with a variance and a covariance respectively superior to 100 and 50 are cancelled out.
- The same operation is repeated with variance and covariance levels of 4 and 2, respectively.
- Finally, a last round is achieved with variance and covariance levels of 1.5 and 1, respectively.

Note that the thresholds, which seem, at first sight, to have been determined rather arbitrarily, are dimensionless. Indeed, the parameter positivity is ensured through a logarithmic transformation of the parameters and the elements of (22) can therefore be regarded as relative errors.

In conclusion, the overall procedure has a number of steps, which, in our experience, are necessary to efficiently isolate hardly assessable parameters. It is important to note that an inaccurate estimation of one parameter can have a significant influence on the estimation of other parameters, i.e., it can lead to significant variance and correlation of parameters, which are essential in the model. We now turn our attention to a few case studies.

5. CASE STUDIES

In this section, three applications are considered, two in simulation, and one based on real experimental data.

5.1 A simple microbial growth

The following reaction scheme is considered:

$$k\,S \xrightarrow{\varphi} X \tag{28}$$

where S denotes the substrate concentration, X the biomass concentration, k the pseudo-stoichiometric coefficient and φ the reaction rate which follows a theoretical Monod-type kinetic law:

$$\varphi(S(t), X(t)) = \mu_{\max} \frac{S(t)}{K_m + S(t)} \frac{X(t)}{K_X + X(t)} \tag{29}$$

The numerical values of the model parameters are the following: $k = 0.5g(10^{11}cell)^{-1}$, $K_m = 12\,gl^{-1}$, $K_X = 10^{11}cell/l$, and $\mu_{\max} = 1.4\,h^{-1}$.

Simulation of this model in various conditions allows the creation of a database, which consists of 4 batches of 50 hours which have the same initial concentration in biomass ($X_0 = 1.4\ 10^{11} cell\,l^{-1}$) and different initial concentrations in substrate ($S_0 = [18 \quad 24 \quad 12 \quad 30] \quad (gl^{-1})$). These latter concentrations are chosen in order to ensure significant substrate saturation at the beginning of the experiments and a strong biomass saturation at the end. The sampling time for the simulated substrate and biomass concentration measurements is 2 hours. Among the different experiments, the first two are used for the identification while the others are kept for cross- validation, to test the generalization of the models.

The direct and cross-validation results are presented in Table 1. We observe that the original kinetic structure gives good results in validation, but that it artificially represents the saturation effects by a strong activation factor compensated by a non negligible inhibition. When using the generalized structure, we observe that an initial 95% confidence in the absence of saturation leads back to the original structure, whereas for x = 70% and 50%, we observe saturation effects without any inhibition and the same minimum of the cost function. Hence, these two starting points lead to the same model, which gives significantly better results in direct and cross validation as well as better standard deviations in parameters than the original structure.

5.2 An animal cell culture

The second simulation example involves a more complex model of animal cell cultures presented in (Perrier *et al*., 2000), i.e. human embryo kidney cell cultures. The corresponding macroscopic reaction scheme is given by

$$\nu_{21} G \xrightarrow{\varphi_1} X \tag{30}$$
$$\nu_{22} G \xrightarrow{\varphi_2} X + \nu_{32} L$$

where G denotes the glucose concentration, X the biomass concentration and L the lactate concentration. ν_{ij} are the pseudo-stoichiometric coefficients and φ_i the reaction rates which follow Monod-type laws:

$$\varphi_1 = \mu_{m1} X \frac{G}{K_R + G} \frac{K_L}{K_L + L} \tag{31}$$

$$\varphi_2 = \mu_{m2} X \frac{G}{K_F + G} \tag{32}$$

The numerical values of the model parameters are the following: $\nu_{21} = 1.7 mmol_G / mmol_X$, $\nu_{22} = 18.5 mmol_G / mmol_X$ $\nu_{32} = 12 mmol_L / mmol_X$, $K_R = 10 mmoll^{-1}$, $K_L = 50 mmoll^{-1}$, $K_F = 10 mmoll^{-1}$, $\mu_{m1} = 0.055\,h^{-1}$ and $\mu_{m2} = 0.045\,h^{-1}$.

The simulated database consists of 9 fed-batch cultures of 110 hours with the same initial concentrations in glucose and lactate ($G_0 = 21 mmol/l, L_0 = 0.13 mmol/l$) but different initial concentrations in biomass ($X_0 = [0.18 \quad 0.4 \quad 0.8]$ $mmol/l$) and different profiles of the external feed rate **F** ($[0.5t \quad 0.1t \quad 0.01t]\,(l/h)$). The sampling time for the simulated measurements is 10h. Among the different experiments, only three are considered for the identification while the others are kept for cross-validation.

The direct and cross-validation results are presented in Table 2. We see that the generalized model gives much better results than the original one. However, we observe three different local minima. An initialization with x = 95% or 70% does not capture the two-substrate saturation, whereas x = 50% does. This demonstrates the existence of local minima and the need for a multistart strategy (note that in contrast with a general multistart strategy, we can use a single parameter x to investigate various scenarios).

5.3 An experimental process (*B. subtilis*)

This bioprocess consists of a culture of bacteria (*B. subtilis*), which produce enzyme by consumption of two substrates (carbon and nitrogen sources). Seven fed-batch experiments with constant feed rate have been carried out to build an experimental data base. The culture time is generally 39h and the sampling period is 4h. All the experiments present similar initial concentrations in substrates (Table 3) but differ in the substrate concentrations in the feed solution (fresh medium). Available experimental data contain the concentration measurements of the four main components: biomass (g/l), nitrogen (g/l), enzyme (axu/ml i.e. activity of xylanase/ml) and carbon (expressed in (g/l) of equivalent glucose). A reaction scheme has been determined in (Grosfils *et al*., 2004):

$$
\begin{aligned}
&1C \xrightarrow{\varphi_1} 0.2\,X_{non-lysed} \\
&1N \xrightarrow{\varphi_2} 15\,X_{non-lysed} + 120\,E \\
&1X_{non-lysed} \xrightarrow{\varphi_3} X_{lysed}
\end{aligned}
\qquad (33)
$$

using a systematic procedure for selecting the most-likely C-identifiable reaction scheme (Hulhoven *et al.*, 2005). $X_{non-lysed}$, X_{lysed}, *E, C* and *N* are the non lysed cells, lysed cells, enzymes, carbon and nitrogen concentrations and the cell lysis reaction rate is assumed to be proportional to the non lysed cells concentration, i.e., $\varphi_3 = 0.051\,X_{non-lysed}$.

We observe better results in direct and cross validation (Table 4) for the proposed structures (3) and (5), as compared to a classical Monod-type representation of the kinetics. The proposed structures give very similar results in this application, where saturation effects are not significant.

6. CONCLUSION

In this study, a general kinetic model structure is presented, which describes activation, saturation and inhibition effects of the main components of a culture. Besides its generality (flexibility in representing a range of behaviours), the main advantage of this model structure (in contrast with classical Monod-type laws) is that it is amenable to a systematic identification procedure. This latter procedure combines various ingredients including initialization, a step-by-step nonlinear estimation (including a natural multistart strategy based on the exploration of a single parameter) and model reduction based on the analysis of the variance/covariance of the model parameters.

ACKNOWLEDGEMENTS

This paper presents research results of the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office. The scientific responsibility rests with its author(s). The authors are grateful to Beldem for its collaboration in the work reported here. They also gratefully acknowledge support from F.R.I.A.

Table 1 Case study 1: Direct and cross validations (DV & CV)

		DV	CV
Model (3)		0.03	0.57
Model (5)	x=0.95	0.03	0.57
	x=0.7 or x=0.5	0.01	0.14

Table 2: Case study 2: Direct & cross validations

		DV	CV
Model (3)		1.00	22.97
Model (5)	x=0.95	0.17	2.60
	x=0.7	0.23	2.69
	x=0.5	0.10	0.80

Table 3: Case study 3: Initial concentrations and feeding concentrations in substrates (g/l)

(g/l)	[C]init	[N]init	[C]feed	[N]feed
1	11.6	0.4	166.9	10.8
2	11.3	0.5	100	6.7
3	10.1	0.5	33.8	2.7
4	11.4	0.6	100	6.7
5	11.1	0.5	33.8	10.8
6	11.7	0.6	67.5	6.7
7	11.4	0.5	33.8	10.8

Table 4 : Case study 3: Direct cross validations

		DV	CV
Model (3)		481	112
Model (5)	x=0.95	475	100
	x=0.7	469	106
	x=0.5	467	100

REFERENCES

Bastin, G. and D. Dochain (1990). On-line estimation and adaptative control of bioreactor. Elsevier, Amsterdam.

Bernard, O. and G. Bastin (2005). On the estimation of the pseudo-stoichiometric matrix for mass balance modeling of biotechnological processes, *Mathematical Biosciences*, 193, 51-77, 2005.

Bogaerts Ph., Castillo J. and R. Hanus (1999) A general mathematical modelling technique for bioprocesses in engineering applications, *System Analysis Modeling Simulation*, 35, 87-113.

Bogaerts, Ph., Delcoux J.-L. and R. Hanus (2003). Maximum likelihood estimation of pseudo-stoichiometry in macroscopic biological reaction schemes. *Chemical Eng. Science*, 58, 1545-1563

Grosfils A., A. Vande Wouwer, A. Gaspar, T. Dauvrin and Ph. Bogaerts (2004). Systematic decoupled identification of pseudo-stoichiometry, lysis rate and kinetics for a xylanase production. *Proceedings of the 9th Symposium on Computer Applications in Biotechnology*, Nancy (France).

Haag J.E., A.Vande Wouwer and Ph. Bogaerts (2005). Dynamic modelling of complex biological systems: A link between metabolic and macroscopic description. *Mathematical Biosciences*, 193, 283-291.

Hulhoven X., Vande Wouwer A. and Ph. Bogaerts (2005). On a systematic procedure for the predetermination of macroscopic reaction schemes. *Bioprocess and Biosystems Engineering*, 27, 283-291.

Oliveira R. (2004). Combining first principles modelling and artificial neural networks: a general framework. *Computers and Chemical Engineering*, 28, 755-766.

Perrier M., S. Feyo de Azevedo, E.C. Ferreira and D. Dochain (2000). Tuning of observer-based estimator: theory and application to the on-line estimation of kinetic parameters. *Control Engineering Practice*, 8, 377-388.

Provost A. and G. Bastin (2004). Dynamic metabolic modelling under the balanced growth condition. *Journal of Process Control*, 14, 717-728.