On the estimation of the pseudo-stoichiometric matrix for mass balance modeling of biotechnological processes

Olivier Bernard, Georges Bastin

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

On the estimation of the pseudo-stoichiometric matrix for mass balance modeling of biotechnological processes

Olivier Bernard — Georges Bastin

N° 4977

October 2003

THÈME 4

ISSN 0249-6399 ISRN INRIA/RR--4977--FR+ENG

INRIA
SOPHIA ANTIPOLIS

On the estimation of the pseudo-stoichiometric matrix for mass balance modeling of biotechnological processes

Olivier Bernard* , Georges Bastin†

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet COMORE

Rapport de recherche n° 4977 — October 2003 — 25 pages

Abstract: In this paper we propose a methodology to determine the structure of the pseudo-stoichiometric coefficient matrix $\mathcal{K}$ in a mass balance based model. The first step consists in estimating the number of reactions that must be taken into account to represent the main mass transfer within the bioreactor. This provides the dimension of $\mathcal{K}$. Then we discuss the identifiability of the components of $\mathcal{K}$ and we propose a method to estimate their values. Finally we present a method to select among a set of possible reaction networks those which are in agreement with the available measurements. These methods are illustrated with real data of the growth of filamentous fungi *Pycnoporus cinabarinnus*, and with a process of lipase production from olive oil by *Candida rugosa*.

Key-words: Modeling, Nonlinear systems, Biotechnology, Identification, Identifiability, Validation.

* INRIA-COMORE; BP93, 06902 Sophia-Antipolis Cedex, France; e-mail: olivier.bernard@inria.fr
† UCL-CESAME, av. G. Lemaître 4-6, 1348 Louvain-La-Neuve, Belgium; e-mail: bastin@auto.ucl.ac.be

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : +33 4 92 38 77 77 — Télécopie : +33 4 92 38 77 65

Vers l'estimation de la matrice pseudo-stoechiométrique pour la modélisation par bilan de masse des procédés biotechnologiques

Résumé : Dans ce papier nous proposons une méthode pour déterminer la structure d'une matrice de pseudo coefficients stoechimétriques $\mathcal{K}$ d'un modèle basé sur un bilan de masse. La première étape consiste à estimer le nombre de réactions qui doivent être prises en compte pour représenter les principaux transferts de masse au sein du bioréacteur. Cela fournit la dimension de $\mathcal{K}$. Nous discutons ensuite de l'identifiabilité des coefficients de $\mathcal{K}$ et proposons une méthode pour les identifier. Finalement, nous voyons comment sélectionner parmi un ensemble de schémas réactionnels possibles ceux qui sont adéquation avec les données disponibles. Ces méthodes sont illustrées avec des données réelles de la croissance et de la biotransformation du champignon filamenteux *Pycnoporus cinabarinnus*, ainsi que par un procédé de production de lipase à partir d'huile d'olive par *Candida rugosa*.

Mots-clés : Modélisation, Systèmes non linéaires, Biotechnologie, Validation.

Contents

1 Introduction and motivation

When the set of n_r microbiological and biochemical reactions taking place in a stirred tank bioreactor is known, a mass balance model can be derived. It describes the dynamical evolution of the n_ζ biological or chemical species involved in the reactions, such as microorganisms, substrates, metabolites, enzymes...

The dynamical behavior of a stirred tank bioreactor is then often described by a general mass-balance model of the following form (see e.g. [1, 2]):

$$\frac{d\zeta}{dt} = \mathcal{K}\, r(\zeta, \psi) + D(\zeta_{in} - \zeta) - \mathcal{Q}(\zeta, \psi), \tag{1}$$

In this model, the vector $\zeta = (\zeta_1, \zeta_2, \ldots, \zeta_{n_\zeta})^T$ is made-up of the concentrations of the various species inside the liquid medium. The term ζ_{in} represents the influent concentrations.

The matrix D is the dilution rate matrix representing the hydraulics mechanisms (inflows and outflows and possible retention) associated with the various species in the reactor. The exchange of matter in gaseous form between the surrounding and the reaction medium is represented by the gaseous flow rate $\mathcal{Q}(\zeta, \psi)$.

The term $\mathcal{K}\, r(\zeta, \psi)$ represents the biological and biochemical conversions in the reactor (per unit of time) according to the underlying reaction network. The $(n_\zeta \times n_r)$ matrix $\mathcal{K}$ is a constant stoichiometric-like coefficient matrix. $r(\zeta, \psi) = (r_1(\zeta, \psi), r_2(\zeta, \psi), \ldots, r_{n_r}(\zeta, \psi))^T$ is a vector of **reaction rates** (or conversion rates). $\mathcal{Q}(\zeta, \psi)$ and $r(\zeta, \psi)$ are supposed to depend on the state ζ and on external environmental factors (represented by ψ) such as temperature, light, aeration rate, etc.

Matrix $\mathcal{K}$ plays a key role in the mass balance modeling. It is associated with a reaction network which is supposed to govern the considered process. Each line of the matrix corresponds to one (bio)chemical species involved in the process. Each column of the matrix corresponds to a (bio) chemical reaction between some of the species. A positive entry k_{ij} means that the i^{th} species is a product of the j^{th} reaction while a negative entry $k_{ij} < 0$ means that it is a reactant or a substrate of the reaction. If $k_{ij} = 0$, the i^{th} species is not involved in the j^{th} reaction.

In some instances the reaction network and thus the matrix $\mathcal{K}$ are unknown or incompletely specified. One can for instance hesitate between several plausible reaction networks that can be supposed to underlie the process. The objective of this paper is to propose a method to guide the user in the identification of the pseudo-stoichiometric matrix $\mathcal{K}$. It is worth noting that the determination of matrix $\mathcal{K}$ is a problem equivalent to that of determining the reaction network. The usual approach dedicated to the determination of reaction networks relies on the linearization of the dynamics around a reference solution [3, 4]. Here, in the spirit of [5, 6], we exploit the structure of the bioprocesses (equation (1)) and our arguments are not based on any linearization.

We will show how to use a set of available data consisting of measurements of ζ, $\mathcal{Q}$, D and ζ_{in} at sampling instants, to determine the size of the matrix $\mathcal{K}$ (i.e. the number of reactions that must be taken into account) and to address the problem of the identification and validation of its coefficients.

Note that it is quite rare for bioprocesses that all the involved variables are measured (sometimes it is even unclear which variables are involved). For this reason we will focus on the estimation of K the submatrix of $\mathcal{K}$ associated with the available measurements ξ.

We stress the fact that the methodology that we discuss is the first modeling stage. The second stage in the modeling, which is not discussed here, would consist in determining the reaction rates as functions of the state variables. This second problem is difficult and suffers as well from a lack of tools to assist the modeler. But this delicate step can be avoided for a large number of applications, where the knowledge of the mass balance (*i.e* matrix $\mathcal{K}$) is sufficient to design controllers or observers [1].

The paper will address the three following problems:

- How many reactions (i.e. how many columns for matrix $\mathcal{K}$) must be taken into account to reproduce the available data set ?
- Which reactions must be taken into account ? Which are the most plausible reaction networks ?
- What are the values of the pseudo-stoichiometric coefficients ?

We will successively consider these 3 problems, without any *a priori* knowledge on the reaction rates $r(\zeta, \psi)$. The approaches will be illustrated with two examples of significant complexity: real data of the growth and biotransformation of the filamentous fungi *Pycnoporus cinnabarinus* and a process of lipase production from olive oil by *Candida rugosa*.

2 Determination of the number of reactions

2.1 Statement of the problem

In this section, we address the first problem, consisting in determining n_r the minimum number of reactions to explain the observed dynamics of the fermenter. We assume that we measure a subset ξ of n_ξ components of ζ that are involved in the systems (*i.e.* which present significant variations along time). Indeed the measurements of the other state components (denoted $\tilde{\xi}$) may not be available, but we assume however that we measure more variables than the number of reactions: $n_\xi > n_r$. If these components have a gaseous phase, we assume that the associated gaseous flow rates $Q(\xi, \tilde{\xi}, \psi)$ are measured.

The equation associated with ξ is thus:

$$\frac{d\xi}{dt} = K\, r(\xi, \tilde{\xi}, \psi) + D(\xi_{in} - \xi) - Q(\xi, \tilde{\xi}, \psi), \tag{2}$$

The matrices K and Q are submatrices of $\mathcal{K}$ and $\mathcal{Q}$, respectively, associated with ξ. As a consequence, in the expression of the mass balance model (2), only the term $K\, r(\xi, \tilde{\xi}, \psi)$ needs to be mathematically expressed.

2.2 Theoretical determination of $\dim(\mathcal{I}m(K))$

Let us integrate equation (2) between 2 time instants $t-T$ and t (T denotes the considered time window):

$$\xi(t) - \xi(t-T) - \int_{t-T}^{t} D(\xi_{in}(\tau) - \xi(\tau)) + Q(\xi(\tau), \psi(\tau))d\tau = K \int_{t-T}^{t} r(\zeta(\tau), \psi(\tau))d\tau\,, \tag{3}$$

Let us denote:

$$\eta(t) = \xi(t) - \xi(t-T) - \int_{t}^{t-T} D(\xi_{in}(\tau) - \xi(\tau)) + Q(\xi(\tau), \psi(\tau))d\tau \tag{4}$$

and

$$\epsilon(t) = \int_{t-T}^{t} r(\zeta(\tau), \psi(\tau)) d\tau$$

Equation (3) can then be rewritten:

$$\eta(t) = K\,\epsilon(t) \tag{5}$$

The vector $\eta(t)$ can be estimated along time from the available measurements. The value of the integral in (4) can be computed *e.g.* with a trapeze approximation.

Remark: More generally, in order to improve the cleaning of the data (noise reduction and diminution of autocorrelation) any linear scalar filter can be applied to (2) and will lead to a linear relationship of the same type as (5). The moving average (3) that we have presented for sake of simplicity is of course only one example of such a filtering.

Indeed, if $G(s,\theta)$ is any transfer function (i.e. any combination of integration, differentiation and delay θ), we have:

$$\mathcal{Y}(s) = G(s,\theta)\mathcal{U}(s) = K\,G(s,\theta)\mathcal{W}(s)$$

where $\mathcal{U}(s)$ and $\mathcal{W}(s)$ respectively denote the Laplace transforms of $\eta(t)$ and $\epsilon(t)$. $\mathcal{Y}(s)$ is the Laplace transform of the signal after filtering.

We denote respectively by $u(t)$ and $w(t)$ the signal derived from $\eta(t)$ and $\epsilon(t)$ after filtration, they verify:

$$u(t) = K\,w(t)$$

Now the question of the dimension of matrix K can be formulated as the *determination of the dimension of the image of* K or in other words, of the dimension of the space where $u(t)$ lives. Note that we assume K to be a full rank matrix. Otherwise, it would mean that the same dynamical behavior for $u(t)$ could be obtained with a matrix K of lower dimension.

Determining the dimension of the $u(t)$ space is a classical problem in statistical analysis. It can be solved by a principal component analysis (see e.g. [7]) that determines the dimension of the vector space spanned by the vectors k_i, rows of K. In order to reach this objective, we consider N time instant $t_1, \ldots, t_N$ (we choose $N > n_\xi$). The way to select these time instants will be discussed in the sequel. We build then the $n_\xi \times N$ matrix U made of N vectors $u(t)$ at these time instants:

$$U = (u(t_1), \ldots, u(t_N))$$

We will also consider the associated matrix of reaction rates, which is unknown:

$$W = (w(t_1), \ldots, w(t_N))$$

We assume that matrix W is full rank. It means that the reactions are independent (none of the reaction rates can be written as a linear combination of the others).

Property 1 *For a matrix K of rank n_r, if W has full rank, then the $n_\xi \times n_\xi$ matrix $M = UU^T = KWW^TK^T$ has rank n_r. Since it is a symmetric matrix, it can be written:*

$$M = P^T \Sigma P$$

where P is an orthogonal matrix ($P^TP = I$) and

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & & & \dots & & 0 \\ 0 & \sigma_2 & 0 & & & & 0 \\ \vdots & & \ddots & & & & \\ & & & \sigma_{n_r} & & & \\ & & & & 0 & & \\ & & & & & \ddots & \vdots \\ 0 & & \dots & & & & 0 \end{pmatrix}$$

with $\sigma_{i-1} \geq \sigma_i > 0$ for $i \in \{2, ..., n_r\}$.

It is a direct application of the singular decomposition theorem [8]. Since rank (M) = rank (KW) = rank (K) = rank $(\Sigma) = n_r$, it provides the result.

Now from a theoretical point of view it is possible to determine the number of reactions in the reaction network: it corresponds to the number of non zero singular values of UU^T. This theoretical approach must however be adapted in the real case where the available measurements are discrete data points perturbed by a noise.

2.3 Practical determination of the number of reactions

In the reality, the ideal case presented in the previous paragraph is perturbed for four main reasons:

- The reaction network that we are looking for is a first approximation of chemical or biochemical reactions which can be very complex. The "real" matrix K is probably a very large matrix. The reactions which are fast or of low magnitude can be considered as perturbation of a dominant reaction network. It is this central (perturbed) reaction network that we want to estimate.
- The measurements are corrupted by noise. This noise can be very important, especially for the measurement of biological quantities for which reliable sensors are rarely available.
- The measurements are performed on a discrete basis. Moreover they are rarely all available exactly at the same time instant t_i, and therefore they must be interpolated if we need a state estimate $\xi(t_i)$ at N time instants t_i in order to build vector U.
- In order to estimate $u(t)$ in equation (4) we need to compute the approximate value of an integral. This may generate additional perturbations.

2.3.1 Data processing: interpolation and smoothing

The data collected on a biotechnological process often result from various sampling strategies carried out with various devices. As a consequence the data are seldom sampled simultaneously. In order to apply the proposed transformations vector U has to be computed with values of the state variables at the same time instants t_i. A large number of tools are available in the literature to interpolate and smooth the data. We suggest here to use spline functions [9] which will at the same time interpolate and smooth a signal. The trade-off between interpolation and smoothing can be chosen by the user.

In the sequel we assume therefore that the set of measurements is available at the time instants τ_i, and that after a smoothing and interpolation process all the variable estimates are available at the time instants t_i.

We hypothesize that the estimates $\xi(t_i)$ are of reasonably good quality and in particular that the sampling frequency is in adequation with the time constants of the signal.

2.3.2 Data normalization

To avoid conditioning problems and to give the same weighting to all the state variables, we normalize each component u_i of the vector u as follows:

$$\tilde{u}_i(t_j) = \frac{u_i(t_j) - a(u_i)}{\sqrt{N}s(u_i)}$$

where $a(u_i)$ is the average value of the $u_i(t_k)$ for $k \in \{1..N\}$, and $s(u_i)$ their standard deviation.

2.3.3 Conclusion for the determination of the number of reactions

In the reality, the noises due to model approximations, measurement errors or interpolation perturb the analysis. Therefore in practice there are no zero eigenvalues for the matrix $M = UU^T$.

The question is then to determine the number of eigenvectors that must be taken into account in order to represent a reasonable approximation of the data $u(t)$. To solve this problem, let us remark that the eigenvalues σ_i of M correspond to the variance associated with the corresponding eigenvector (inertia axis) [10].

The method will then consist in selecting the n_r first principal axis which represent a total variance larger than a fixed threshold.

For instance, in the next example, we have fixed a threshold (depending on the information available on noise measurements) at 95% of the variance. This led to the selection of 6 axes, and therefore $n_r = 6$.

Remark: if rank $(M) = n_\xi$ it means that $rank(W) \geq n_\xi$. In such a case we cannot estimate n_r and measurements of additional variables are requested to apply the proposed method.

2.4 Example 1

We consider here the production of vanillin from vanillic acid by the filamentous fungus *Pycnoporus cinnabarinus* [11]. The species involved in this biotransformation process are the carbon sources (maltose and glucose), the nitrogen source (ammonium), oxygen, carbon dioxide, fungal biomass and phenolic compounds (vanillic acid, vanillin, vanillic alcohol and methoxyhydroquinone). This results from a complex set of reactions [12], most of them being ill known.

The process generally proceeds in two steps. In a first step (which generally lasts the 3 first days), the fungus uses the available substrates (nitrogen, maltose and glucose) to grow. The growth is aerobic, and therefore oxygen is consumed and CO_2 produced.

In a second step, the biosynthesis is triggered with addition of cellobiose 2 hours before continuous addition of vanillic acid. Then the fungus transforms the vanillic acid either in methoxyhydroquinone, or in vanillin. In this last case, vanillin can also be degraded into vanillic alcohol.

For illustration purpose, Figure 1 presents the typical evolution of some of the key variables during the fermentation. The figure presents also the splines used to smooth and interpolate the data in order to build the vector $u(t)$ made of the 10 measured species. The final data set consists in 9 experiments which have been resampled to get 4 time instant t_i per day. Finally, 619 data points $u(t_i)$ were considered.

Figure 2 represents the cumulated variance associated with the number of reactions. Four reactions are sufficient to explain 80% of the observed variance. Five reactions explain 95% of the total variance. This analysis motivated the structure of the model presented in [13].

3 Validation of a reaction network

3.1 Statement of the problem

In the previous section we have shown how to determine the number of reactions which must be considered in order to explain the available data. Let us now assume that a plausible reaction network, with this number of reactions, is postulated with the aim of describing the process. In this section, we shall now show how such a candidate reaction network can be validated from the data.

One additional difficulty in comparing a reaction network, via its stoichiometric matrix K, with a set of data is that some pseudo-stoichiometric coefficients may be *a priori* unknown. We shall propose a method which will allow at the same time to test the validity of the reaction network and to identify the missing pseudo-stoichiometric coefficients.

3.2 Finding the left kernel of the pseudo-stoichiometric matrix

Let us consider a vector $\lambda \in \mathcal{K}erK^T$:

$$\lambda^T K = 0$$

Assume moreover that λ is normalized such that one of its components λ_i is 1: $\lambda_{i_0} = 1$

Now let us consider the scalar quantity $\lambda^T u(t)$. From equation (5), it satisfies at any time t:

$$\lambda^T u(t) = 0$$

In other words, we have:

$$u_{i_0}(t) = -\sum_{j \neq i_0} \lambda_j u_j(t) \tag{6}$$

This means that the u_j are linked by a linear relation. The immediate idea that one can have is to check whether relationship (6) is in adequation with the data. This can be done by performing a linear regression between u_{i_0} and the u_j.

Nevertheless, we have to keep in mind that the u_j are *a priori* not independent, since they may be related by other relationships associated with other left kernel vectors of K. In particular, we have seen that rank $(U) = n_r$, and thus a regression (6) cannot involve more than $n_r + 1$ independent terms u_j.

We will therefore select the vectors λ of the left kernel such that they imply that only independent u_j are to be considered in (6).

It is worth noting that the vector λ involves three kinds of components:

1. entries which are structurally zero
2. entries that have an *a priori* known non-zero value (either 1 for the normalizing component, see above, or a known value related to the stoichiometry of the reaction network).
3. entries which are unknown because they depend on unknown coefficients of the pseudo-stoichiometric matrix. These entries have to be estimated from the data.

Remark that Equation (6) states a conservation between the variables u_i. This conservation is directly connected to the notion of reaction invariants [14, 15].

Definition 1 *We say that a set $\{u_{i_1} \dots u_{i_k}\}$ is associated with a left kernel vector λ if $\lambda_j = 0$ for all the indices $j \notin \{i_1 \dots i_k\}$. We say that λ is associated with the $k \times n_r$ submatrix $\tilde{K}$ which is the submatrix made of the raws $i_1 \dots i_k$ of K. Finally we call $\tilde{\lambda}$ the vector obtained by removing all the zeros entries in λ. The dimension of $\tilde{\lambda}$ (namely k) is called the regression dimension associated with λ and denoted $d(\lambda)$, the number of unknown components of λ is denoted $d_u(\lambda)$.*

We have therefore $\tilde{\lambda}^T \tilde{K} = 0$, and $\tilde{\lambda}$ has no zero component. Then, $\sum_{i_k} \lambda_{i_k} u_{i_k}(t) = 0$.

Note that, due to the normalization of λ, we have $d_u(\lambda) \leq d(\lambda) - 1$.

Definition 2 *We say that a left kernel vector λ is* **sound** *if its associated $d(\lambda) \times n_r$ matrix $\tilde{K}$ does not contain itself any $k \times n_r$ submatrix ($k < d(\lambda)$) whose rank is not full or — equivalently — if* $\dim(\mathcal{K}er\tilde{K}^T) = 1$.

Remark: For a sound vector λ we have $d(\lambda) \leq n_r + 1$

Indeed, if it has $k \geq n_r + 2$ non zero entries, then its associated submatrix $\tilde{K}$ is a $k \times n_r$ submatrix whose left kernel is at least of dimension 2.

3.3 Example 2

Let us consider the example of the competitive growth on two substrates [16] which could represent the production of lipase from olive oil by *Candida rugosa*. Here the microorganism is supposed to grow on two substrates that are produced by the hydrolysis of a primary complex organic substrate. We assume the following 3-step reaction network:

- Hydrolysis:
$$k_1 S_1 \ + \ E \ \longrightarrow \ S_2 \ + \ k_2 S_3 \ + E$$
- Growth on S_2:
$$k_3 S_2 \ + \ k_4 O \ \longrightarrow \ X \ + \ k_5 P$$
- Growth on S_3:
$$k_6 S_3 \ + \ k_8 O \ \longrightarrow \ X \ + \ k_7 E \ + \ k_9 P$$

where S_1 is the primary substrate (olive oil), S_2 (glycerol) and S_3 (fatty acid) are the secondary substrates, E is the enzyme (lipase), X the biomass (*Candida rugosa*), O the dissolved oxygen and P the dissolved CO_2. We assume that all the biochemical species are measured, except S_1 whose measurement is less straightforward since it is made of various compounds.

The associated pseudo-stoichiometric matrix $\mathcal{K}$ and the state vector ζ are therefore:

$$\mathcal{K} = \begin{pmatrix} -k_1 & 0 & 0 \\ 1 & -k_3 & 0 \\ k_2 & 0 & -k_6 \\ 0 & 0 & k_7 \\ 0 & 1 & 1 \\ 0 & -k_4 & -k_8 \\ 0 & k_5 & k_9 \end{pmatrix}, \quad \zeta = \begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ E \\ X \\ O \\ P \end{pmatrix}$$

Since S_1 is not measured, we will focus on the state ξ associated with the submatrix K:

$$K = \begin{pmatrix} 1 & -k_3 & 0 \\ k_2 & 0 & -k_6 \\ 0 & 0 & k_7 \\ 0 & 1 & 1 \\ 0 & -k_4 & -k_8 \\ 0 & k_5 & k_9 \end{pmatrix}, \quad \xi = \begin{pmatrix} S_2 \\ S_3 \\ E \\ X \\ O \\ P \end{pmatrix}, \quad \bar{\xi} = \begin{pmatrix} S_1 \end{pmatrix}$$

Now the following vector belongs to the kernel of matrix K^T:

$$\lambda^1 = \begin{pmatrix} 0 \\ 0 \\ \frac{k_5 - k_9}{k_7} \\ -k_5 \\ 0 \\ 1 \end{pmatrix}$$

We have $d(\lambda^1) = 3$ and $d_u(\lambda^1) = 2$. It is associated with the rank-2 submatrix $\tilde{K}_1$ and to the vector $\tilde{\lambda}^1$:

$$\tilde{K}_1 = \begin{pmatrix} 0 & 0 & k_7 \\ 0 & 1 & 1 \\ 0 & k_5 & k_9 \end{pmatrix}, \quad \tilde{\lambda}_1 = \begin{pmatrix} \frac{k_5 - k_9}{k_7} \\ -k_5 \\ 1 \end{pmatrix}$$

which is sound since the 3 possible 2×3 submatrices are of full rank.

Thus u_4, u_5 and u_7 are associated with λ^1, and related by the following relation:

$$u_7(t) = k_5 u_5(t) + \frac{k_9 - k_5}{k_7} u_4(t) \tag{7}$$

Now the kernel of matrix K^T is spanned by the 2 other sound vectors:

$$\lambda^2 = \begin{pmatrix} 0 \\ 0 \\ \frac{k_8 - k_4}{k_7} \\ k_4 \\ 1 \\ 0 \end{pmatrix}, \quad \lambda^3 = \begin{pmatrix} -k_2 \\ 1 \\ \frac{k_3 k_2 + k_6}{k_7} \\ -k_3 k_2 \\ 0 \\ 0 \end{pmatrix}$$

Obviously, we have $d(\lambda^2) = 3$, $d_u(\lambda^2) = 2$, $d(\lambda^3) = 4$, $d_u(\lambda^3) = 3$,

3.4 Regressions associated with a set of u_i

Property 2 *A vector λ associated with a set $\{u_{i_1} \dots u_{i_k}\}$ is sound if and only if the u_i are not related by any other linear relation.*

Proof: Indeed, it is clear that it is not possible to have another relation between $n_r + 1$ different u_i, otherwise this relation would be associated with a second kernel vector λ', meaning then that the kernel of $\tilde{K}^T$ is at least of dimension 2, and thus λ would not be sound.

Property 3 *Let us consider a set $\{u_{i_1} \dots u_{i_k}\}$ associated with a left kernel vector λ, and to a matrix $\tilde{K}$: $\sum_{j=1}^{k} \lambda_{i_j} u_{i_j}(t) = 0$. If λ is not sound, then any submatrix $\tilde{K}'$ obtained from $\tilde{K}$ by removing the l^{th} raw is associated with a subset of $\{u_{i_1} \dots u_{i_k},\ i_j \neq l\}$, i.e.: $\sum_{\substack{j=1 \\ j \neq l}}^{k} \lambda'_{i_j} u_{i_j}(t) = 0$ (for $l \in \{i_1, \dots, i_k\}$).*

Proof: if λ is not sound, it means that $\dim(\mathcal{K}er\tilde{K}^T) > 1$. Therefore there exists at least 2 different vectors λ^1 and λ^2 such that $\sum_{j=1}^{k} \lambda^q_{i_j} u_{i_j}(t) = 0$ for $q \in \{1, 2\}$. If the l^{th} component of λ^1 or λ^2 contains a zero, then we have the result. Otherwise, for $\lambda^1_l \lambda^2_l \neq 0$, we have

$$\sum_{j=1}^{k} \lambda^1_{i_j} u_{i_j}(t) - \frac{\lambda_{i_l}}{\lambda_{i_2}} \sum_{j=1}^{k} \lambda^2_{i_j} u_{i_j}(t) = 0$$

showing that the vector $\tilde{\lambda}'$ whose components are $\lambda^1_{i_j} - \frac{\lambda_{i_l}}{\lambda_{i_2}} \lambda^2_{i_j}$ for $i_j \in \{i_1, \dots, i_k\ i_j \neq l\}$ is associated with the matrix $\tilde{K}'$ obtained from $\tilde{K}$ by removing the l^{th} raw.

Property 4 *Let us consider a sound kernel vector λ of K^T, associated with $\tilde{\lambda}$ and to a set $\{u_{i_j},\ i_j \in \{i_1..i_{d(\lambda)}\}\}$. Moreover, let us denote by $\mathcal{S}$ the set of indices j such that $\tilde{\lambda}_j$ is known. Then the following cost criterion:*

$$J(\alpha) = \sum_{t=t_1}^{t_N} (\sum_{j \in \mathcal{S}} \tilde{\lambda}_j u_{i_j}(t) - \sum_{j \notin \mathcal{S}} \alpha_j u_{i_j}(t))^2 \tag{8}$$

admits a unique minimum, of zero value, obtained for $\alpha_j = \tilde{\lambda}_j$ (for any $j \notin \mathcal{S}$).

It is worth noting that minimizing $J(\alpha)$ is exactly a linear regression problem.

3.5 Validation of the kernel of K^T with the available data

Now the validation will consist in verifying that $J(\alpha)$ (Equation 8) can be correctly minimized, or in other words, that the regression between $v = \sum_{j \in \mathcal{S}} \tilde{\lambda}_j u_{i_j}$ and the u_{i_j} $(j \notin \mathcal{S})$ is significant.

This analysis must be performed on all the sound kernel vectors λ^i of K^T. In order to maximize the quality of the regression, the u_{i_j} associated with λ^i $(j \notin \mathcal{S})$ and v should in practice span a space of dimension $d_u(\lambda^i)$. So we perform a principal component analysis

for the matrix

$$U_i = \begin{pmatrix} v(t_1) & \dots & v(t_N) \\ u_{j_1}(t_1) & \dots & u_{j_1}(t_N) \\ & \vdots & \\ u_{j_{d_u(\lambda^i)}}(t_1) & \dots & u_{j_{d_u(\lambda^i)}}(t_N) \end{pmatrix}$$

where the index j_i correspond to the unknown elements of $\tilde{\lambda}^i$. The eigenvalues of $U_i U_i^T$ represent the total variance σ_{ij} associated with the jth principal axis. We then sort the singular values so that $\sigma_1 \geq \dots \geq \sigma_{d_u(\lambda^i)} \geq \sigma_{d_u(\lambda^i)+1}$. Let us recall that, in principle, $\sigma_{d_u(\lambda^i)} > 0$ and $\sigma_{d_u(\lambda^i)+1} = 0$.

We consider the following criterion (reminiscent to the conditioning number) which assesses the balance of the variance along the axis:

$$\mathcal{B}(\lambda^i) = \frac{\sigma_1(\lambda^i)}{\sigma_{d_u}(\lambda^i)}$$

With this criterion, we can now order the kernel vectors as follows:

- We first sort the kernel vectors λ^i by sets of constant regression dimension $d_u(\lambda^i)$.
- Within the sets of constant regression dimension $d_u(\lambda^i)$, we sort the λ^i by increasing index of associated variance balance $\mathcal{B}(\lambda^i)$.

Definition 3 *The basis made of the first $n_\xi - n_r$ independent vectors λ^i is called the sound kernel basis.*

3.6 Identifiability of the pseudo-stoichiometric coefficients

The question that we want to discuss in this section is to determine whether it is possible to determine the set of pseudo-stoichiometric coefficients k_i from the values of λ_i identified from the set of regressions given by equations (6). This identifiability property when the reaction rates $r(\zeta, \psi)$ are unknown is referred to as C-identifiability in [5].

The answer to the C-identifiability question can be found in [5]. A version of this Theorem is recalled here in the considered framework of full rank matrices K:

Theorem 1 (Chen & Bastin 1995) *Let K be an $n_\xi \times n_r$ full rank matrix with $n_\xi > n_r$. The unknown elements of the j^{th} column of K are said to be C-identifiable if and only if there exists a nonsingular partition (K_a, K_b), where K_a is a full rank submatrix $n_r \times n_r$ which does not contain any unknown element in its j^{th} column.*

We propose here a broader sufficient condition for the C-identifiability:

Theorem 2 *Let K be an $n_\xi \times n_r$ full rank matrix with $n_\xi > n_r$. The unknown element k_{ij} of K is identifiable if there exists a $k \times n_r$ full rank submatrix K_a, with $k \leq n_r$, which does not contain any unknown element on its i^{th} column such that the $(k+1) \times n_r$ submatrix of K:*

$$\Xi = \begin{pmatrix} K_a \\ K_{bi} \end{pmatrix}$$

verifies rank $(\Xi) < k+1$, *where K_{bi} is the i^{th} line of K.*

Proof: If $k+1 >$ rank (Ξ), then $\dim(\mathcal{K}er\Xi^T) > 0$, there exists a kernel vector $\lambda = \begin{pmatrix} \lambda_a \\ \lambda_{bi} \end{pmatrix}$ such that $\lambda^T\Xi = 0$.

We have therefore $\lambda_a^T K_a + \lambda_{bi} K_{bi} = 0$.

Since K_a is a $k \times n_r$ full rank matrix with $k \leq n_r$, then $\dim \mathcal{K}er K_a^T = 0$, and thus λ_{bi} cannot be zero.

If λ is not sound *i.e.* $\dim \mathcal{K}er\Xi^T > 1$. We must then consider the sound vector $\tilde{\lambda}$ associated with the submatrix $\begin{pmatrix} \tilde{K}_a \\ K_{bi} \end{pmatrix}$, where $\tilde{K}_a$ is extracted from K_a according to Property 3.

The sound vector $\tilde{\lambda}$, verifies: $\tilde{\lambda}_a^T \tilde{K}_a + \lambda_{bi} K_{bi} = 0$

Let us remark that it is a matrix equality, and let us consider the j^{th} column of this matrix equation:

$$\tilde{\lambda}_a^T \tilde{K}_{ai} + \lambda_{bi} k_{ij} = 0$$

where $\tilde{K}_{ai}$ is the i^{th} column of $\tilde{K}_a$.

As we saw in Section 3.5, the coefficients of the sound kernel vector $\tilde{\lambda}$ can be identified from a linear regression. Therefore, k_{ij} can be computed as follows:

$$k_{ij} = -\frac{\tilde{\lambda}_a^T \tilde{K}_{ai}}{\lambda_{bi}}$$

Remark: This criterion, although it is more complicated than the one proposed in [5], allows to check the C-identifiability for each element of K separately and not only for the columns.

Let us consider the following matrix K:

$$K = \begin{pmatrix} k_{11} & 1 \\ 1 & 0 \\ k_{31} & 0 \end{pmatrix} \tag{9}$$

Theorem 1 states that the first column of K is not C-identifiable, since it is not possible to find a 2×2 submatrix K_a which do not contain any unknown element in its first column. Now if we apply Theorem 2, we can use the following submatrices:

$$K_a = \begin{pmatrix} 1 & 0 \end{pmatrix}, \; K_b = \begin{pmatrix} k_{31} & 0 \end{pmatrix}$$

Then Ξ is of rank 1, and verifies the condition $k+1 = 2 >$ rank (Ξ), it follows that k_{31} is C-identifiable. It is now clear that k_{11} is not C-identifiable, otherwise the first column of K would be C-identifiable.

Remark that the analysis of the kernel of matrix K^T also provides a criterion to test the identifiability of the k_{ij}. Even if this criterion is less convenient, it will give some hints on the practical identifiability, as we will see in the next Property.

Property 5 *The pseudo-stoichiometric coefficient k_{ij} is C-identifiable if and only if it can be computed from any combination of sound kernel vectors.*

In the previous example of equation (9), the sound kernel basis of K^T was

$$\tilde{\lambda} = (0,\ -k_{31},\ 1)^T$$

It follows that k_{31} is C-identifiable and that k_{11} is not C-identifiable.

3.7 Identification of the pseudo-stoichiometric coefficients and final validation

Now, once we know that the pseudo-stoichiometric coefficients are identifiable, we can estimate their value from experimental data using Property 5. For this, we will use the regression associated with the sound kernel basis of K^T given by equation (6). The statistical significance of the correlation will allow to test from the data whether the vectors $\tilde{\lambda}^i$ are in the kernel of K^T or not.

The final validation will consist in checking that the pseudo-stoichiometric coefficients are all positive. This test must be performed with regards to the uncertainty obtained from the linear regression (6). Indeed, because of the uncertainty obtained on the estimates for the λ_i the k_i may have a negative value, but with a confidence interval intersecting the positive domain.

3.8 Improving the method

A fermentation is often composed of several phases. In each of this phase, some reactions are not triggered. Therefore it is generally possible to find time intervals $]T_k, T_{k+1}[$ for which $r_j = 0$ for some j. In the same way, the concentration of some components may remain constant during certain periods of time.

This is for example the case in a reaction where the primary (associated with growth) and the secondary metabolisms are successively activated. During the first stage only growth takes place: no biotransformation appears since no precursor is added. During the secondary metabolism phase, the growth is inhibited and the microorganism concentrates on the bioproduction of a metabolite.

During these periods of time $]T_k, T_{k+1}[$, the system is then characterized by an index j_0 such that $r_{j_0} = 0$. System (1) is then equivalent to the following system:

$$\frac{d\xi}{dt} = \bar{K}\,\bar{r}(\xi,\psi) + D(\xi_{in} - \xi) - Q(\xi,\psi), \tag{10}$$

where the matrix $\bar{K}$ is extracted from K by removing the columns of K corresponding to the index j_0.

Finally on these time intervals, the study of system (1) can be simplified by studying (10).

3.9 Example 2 (continued)

Let us consider the example of the competitive growth on 2 substrates. Let us assume that substrates S_2 and S_3 can directly be obtained from another bioreactor where the enzyme has been purified and directly added to S_1 without the biomass. We will then consider such an experiment where the secondary substrates S_2 and S_3 are directly added. Therefore, for all these experiments we will have $r_1 = 0$. The problem reduces thus to find the kernel of the submatrix $\bar{K}$ obtained after removing the first column of K:

$$\bar{K} = \begin{pmatrix} -k_3 & 0 \\ 0 & -k_6 \\ 0 & k_7 \\ 1 & 1 \\ -k_4 & -k_8 \\ k_5 & k_9 \end{pmatrix}$$

The kernel of $\bar{K}^T$ is spanned by the following sound vectors:

$$\bar{\lambda}^1 = \begin{pmatrix} 0 \\ \frac{k_7}{k_6} \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \ \bar{\lambda}^2 = \begin{pmatrix} \frac{k_8 - k_4}{k_3} \\ 0 \\ 0 \\ k_8 \\ 1 \\ 0 \end{pmatrix}, \ \bar{\lambda}^3 = \begin{pmatrix} -\frac{k_7}{k_3} \\ 0 \\ 1 \\ -k_7 \\ 0 \\ 0 \end{pmatrix}, \ \bar{\lambda}^4 = \begin{pmatrix} \frac{k_5 - k_9}{k_3} \\ 0 \\ 0 \\ -k_9 \\ 0 \\ 1 \end{pmatrix}$$

The regression dimension are $d(\bar{\lambda}^1) = 2$, $d_u(\bar{\lambda}^1) = 1$ and $d(\bar{\lambda}^i) = 3$, $d_u(\bar{\lambda}^i) = 2$ for $i > 1$. Note that $\bar{K}$ is associated with regressions of lower dimension than K implying less unknown coefficients λ_i^j. It will therefore provide more reliable results (with the same amount of data), which will be easier to validate.

4 Comparison between several reaction networks

4.1 Statement of the problem

Once the number of reactions n_r to be taken into account has been identified, the next step consists in selecting the set of reactions which are supposed to represent the main mass transfer in the fermenter. In general, several hypotheses can be stated with respect to the available knowledge.

We assume therefore that a set of q plausible reaction networks with q associated pseudo-stoichiometric matrices K_i are postulated by the user. It may *e.g.* be the result of automatic determination procedures, like those presented in [17, 18]. The aim of this section is to determine how to select among these q hypotheses those who provide a pseudo-stoichiometric matrix in agreement with the available data. Remark however that, in most cases, q is a small number since there are only a few possible reaction networks.

The method consists therefore in testing each matrix K_i by using the methodology exposed in Section (3.7) and then to select the models which pass the validation tests.

The proposed methodology will be presented through a real life case study: the modeling of the growth of the filamentous fungus *Pycnoporus cinnabarinus*

4.2 A real case study

We focus here on experimental phases were only aerobic growth of the fungus *Pycnoporus cinnabarinus* takes place. From a preliminary analysis of the available measurements, it turns out that 2 reactions are necessary to explain the observed data (see Figure 3).

The aerobic growth of the fungal biomass (X) from a carbon source (glucose G and maltose M) and a nitrogen source (N) can *a priori* be reasonably represented by the 3 following reactions networks:

- Network 1:

  The fungus is growing on maltose, glucose and nitrogen, and it can transform maltose into glucose in a first step:

$$\begin{aligned} M &\xrightarrow{r_1(.)} 2\,G \\ k_1\,N \;+\; k_2\,G \,+\, k_3\,M &\xrightarrow{r_2(.)} X \end{aligned} \tag{11}$$

- Network 2:

  The fungus is growing only on glucose and nitrogen, and it transforms maltose into glucose in a first step:

$$\begin{aligned} M &\xrightarrow{r_1(.)} 2\,G \\ k_1\,N \;+\; k_2\,G &\xrightarrow{r_2(.)} X \end{aligned} \tag{12}$$

- Network 3:

  The fungus can grow either on glucose and nitrogen or on maltose and nitrogen. In this second case glucose is produced.

$$\begin{aligned} k_4\,N \;+\; k_5\,G &\xrightarrow{r_1(.)} X \\ k_1\,N \;+\; k_2\,M &\xrightarrow{r_2(.)} X \,+\, k_3\,G \end{aligned} \tag{13}$$

The pseudo-stoichiometric matrices associated with (11), (12) and (13) are then respectively:

$$K_1 = \begin{pmatrix} 0 & -k_1 \\ -1 & -k_3 \\ 2 & -k_2 \\ 0 & 1 \end{pmatrix} \tag{14}$$

$$K_2 = \begin{pmatrix} 0 & -k_1 \\ -1 & 0 \\ 2 & -k_2 \\ 0 & 1 \end{pmatrix} \tag{15}$$

$$K_3 = \begin{pmatrix} -k_1 & -k_4 \\ -k_2 & 0 \\ k_3 & -k_5 \\ 1 & 1 \end{pmatrix} \tag{16}$$

Using the method presented in section (3.7) we give in Table 1 the sound kernel vectors and the corresponding regressions which are associated with these three pseudo-stoichiometric matrices.

PS Matrix	Sound kernel basis of K^T	Regressions	$\mathcal{B}(\lambda^i)$
K_1	$\lambda_1^1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ k_1 \end{pmatrix}$, $\lambda_2^1 = \begin{pmatrix} 0 \\ -2 \\ 1 \\ -2k_3 + k_2 \end{pmatrix}$	$u_1 = -c_1^{1+} u_4$ $2u_2 + u_3 = c_3^1 u_4$	$\mathcal{B}(\lambda_1^1) = 1$ $\mathcal{B}(\lambda_2^1) = 1$
K_2	$\lambda_1^2 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ k_1 \end{pmatrix}$, $\lambda_2^2 = \begin{pmatrix} 0 \\ 2 \\ 1 \\ k_2 \end{pmatrix}$	$u_1 = -c_1^{2+} u_4$ $2u_2 + u_3 = c_3^{2+} u_4$	$\mathcal{B}(\lambda_1^2) = 1$ $\mathcal{B}(\lambda_2^2) = 1$
K_3	$\lambda_1^3 = \begin{pmatrix} 0 \\ \frac{k_5+k_3}{k_2} \\ 1 \\ k_5 \end{pmatrix}$, $\lambda_2^3 = \begin{pmatrix} 1 \\ \frac{k_4-k_1}{k_2} \\ 0 \\ k_4 \end{pmatrix}$	$u_3 = -c_1^{3+} u_2 - c_2^{3+} u_4$ $u_1 = c_3^3 u_2 - c_4^{3+} u_4$	$\mathcal{B}(\lambda_1^3) = 4.48$ $\mathcal{B}(\lambda_2^3) = 6.29$

Table 1: Kernel vectors and regressions associated with the pseudo-stoichiometric matrices for each of the considered reaction network for the growth of *Pycnoporus cinnabarinus* on ammonium, maltose and glucose. The real c_i^{j+} are positive, the c_i^j can be of any sign.

The regression coefficients computed from 70 data points coming from 9 different experiments are presented in Table 2. The confidence intervals for the parameters have been

estimated using a Student distribution with a 5% threshold and the significance of the regression has been tested.

RN	Parameter	min	max	Positivity	Significance	Conclusion
1	c_1^{1+}	0.41	0.79	YES	YES	$k_1 \in [0.41,\ 0.79]$
	c_3^1	1.4	1.77	/	YES	$k_2 - 2k_3 \in [1.4,\ 1.77]$
2	c_1^{2+}	0.41	0.79	YES	YES	$k_1 \in [0.41,\ 0.79]$
	c_3^{2+}	1.4	1.78	YES	YES	$k_2 \in [1.4,\ 1.78]$
3	c_1^{3+}	0.72	1.1	YES	YES	$\frac{k_5+k_3}{k_2} \in [0.72,\ 1.1]$
	c_2^{3+}	1.40	1.78	YES	YES	$k_5 \in [1.40,\ 1.78]$
	c_3^3	0.93	1.28	/	NO	$\frac{k_4-k_1}{k_2} \in [0.93,\ 1.28]$
	c_4^{3+}	-0.45	-0.11	NO	NO	$k_4 \in [-0.45,\ -0.11]$

Table 2: Estimation of intervals for parameter values and significance of the regressions (threshold 5%) associated with each of the reaction networks (RN).

From Table 2 we conclude immediately that network 3 is invalidated by the data. The coefficients associated with networks 1 and 2 have the correct signs, and therefore only these two networks are in agreement with the data and will be kept. Note that it is not possible to distinguish between network 1 and network 2. However the parameters k_2 and k_3 in network 1 are not identifiable, and thus network 2 would be preferred. Nevertheless,, if network 1 should be kept for some reasons, the value (of at least one) of the (unidentifiable) parameters k_2 and k_3 should be selected, in such a way that $k_2 - 2k_3$ belongs to the confidence interval from Table 2.

5 Conclusion

Modeling of bioprocesses is known to be a difficult issue since there does not exist universal validated laws on which the model can rely as in other fields like mechanics (fundamental equations of mechanics), electronics (ohm law), etc.

The mass balance approach presents the advantage to uncouple the part of the model related to the reaction network through matrix $\mathcal{K}$ from the part of the model related to the microbial kinetics ($r(\xi)$). Some algorithm can be based only on the mass balance part [1] which limits the uncertainty associated with variability of the biological processes. However the resulting algorithms turn out to be very sensitive to the pseudo-stoichiometric matrix. Validation of this matrix and improvement of its identification is therefore a key issue for biotechnological processes.

Acknowledgment: The authors are grateful to Benoit Chachuat for his comments and corrections. This work has been carried out with the support provided by the European commission, Information Society Technologies programme, Key action I Systems & Services for the Citizen, contract

TELEMAC number IST-2000-28256. It also presents research results of the Belgian Programme on Inter-University Poles of Attraction initiated by the Belgian State, Prime Minister's office for Science, Technology and Culture. The scientific responsibility rests with its authors.

References

[1] G. Bastin and D. Dochain, *On-line estimation and adaptive control of bioreactors.* Elsevier, 1990.

[2] G. Bastin and J. VanImpe, "Nonlinear and adaptive control in biotechnology: a tutorial," *European Journal of Control*, vol. 1, no. 1, pp. 1–37, 1995.

[3] A. F. M. Eiswirth and and J. Ross, *Mechanistic classification of chemical oscillators and the role of species*, vol. 80 of *Advances in Chemical Physics*, ch. 1, pp. 127–199. New-York: Wiley, 1991.

[4] T. Chevalier, I. Schreiber, and J. Ross, "Toward a systematic determination of complex reaction mechanisms source : journal of physical chemistry," *J. Phys. Chem*, vol. 97, pp. 6776 – 6787, 1993.

[5] L. Chen and G. Bastin, "Structural identifiability of the yield coefficients in bioprocess models when the reaction rates are unknown," *Math. Biosciences*, vol. 132, pp. 35–67, 1996.

[6] O. Bernard and G. Bastin, "Structural identification of nonlinear mathematical models for bioprocesses," in *Proceedings of the Nonlinear Control Systems Symposium 98*, pp. 449–454, Enschede, July 1-3, 1998, 1998.

[7] R. A. Johnson and D. W. Wichern, *Applied multivariate statistical analysis.* Prentice Hall, 1992.

[8] R. A. Horn and C. R. Johnson, *Matrix analysis.* Cambridge University Press, Cambridge MA, 1993.

[9] C. D. Boor, "Applied mathematical sciences," in *A Practical guide to splines* (Springer, ed.), p. 392, New York: Wiley, 1978.

[10] R. Johnson and D. Wichern, *Applied multivariate statistical analysis, 4th ed.* Prentice-Hall, 1998.

[11] B. Falconnier, C. Lapierre, L. Lesage-Meessen, G. Yonnet, P. Brunerie, B. Colonna-Ceccaldi, G. Corrieu, and M. Asther, "Vanillin as a product of ferulic acid biotransformation by the white-rot fungus *Pycnoporus cinnabarinus* I-937: identification of metabolic pathways," *J. Biotechnol*, vol. 37, pp. 123–132, 1994.

[12] L. Lesage-Meessen, M. Delattre, M. Haon, J. Thibault, B. Colonna-Ceccaldi, P. Brunerie, and M. Asther, "A two-step bioconversion process for vanillin production from ferulic acid combining *Aspergillus niger* and *Pycnoporus cinnabarinus*," *J. Biotechnol*, vol. 50, pp. 107–113, 1996.

[13] O. Bernard, G. Bastin, C. Stentelaire, L. Lesage-Meessen, and M. Asther, "Mass balance modelling of vanillin production from vanillic acid by cultures of the fungus *Pycnoporus cinnabarinus* in bioreactors," *Biotech. Bioeng*, pp. 558–571, 1999.

[14] M. Fjeld, "On a pitfall in stability analysis of chemical reactions," *Chem. Engin. Science*, vol. 23, pp. 565–573, 1968.

[15] O. Asbjornsen and M. Fjeld, "Response modes of continuous stirred tank reactors," *Chem. Engin. Science*, vol. 25, pp. 1627–1636, 1970.

[16] P. Serra, J. del Rio, J. Robusté, M. Poch, C. Sola, and A.Cheruy, "A model for lipase production by *Candida rugosa*," *Bioprocess Engineering*, vol. 8, pp. 145–150, 1992.

[17] P. Bogaerts and A. Vande Wouwer, "Sytematic generation of identifiable macroscopic reaction schemes," in *Proceedings of the 8th IFAC Conference on Computer Applications in Biotechnology (CAB8)*, Montréal, Canada, 2001, 2001.

[18] O. Bernard and G. Bastin, "Identification of reaction schemes for bioprocesses: determination of an incompletely known yield matrix," in *Proceedings of ECC03*, Cambridge, UK, 2003.

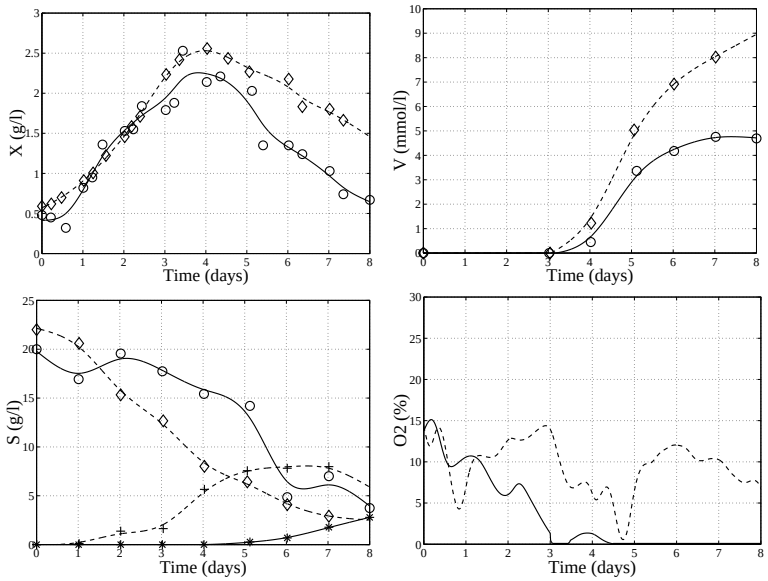

Figure 1: Measurements of biomass (X), vanillin (V), maltose and glucose (M and G) and oxygen (O) for experiments C ($\diamond$) and D (o). The continuous lines are the smoothing splines that will be used in the sequel (C: - -, D: –).

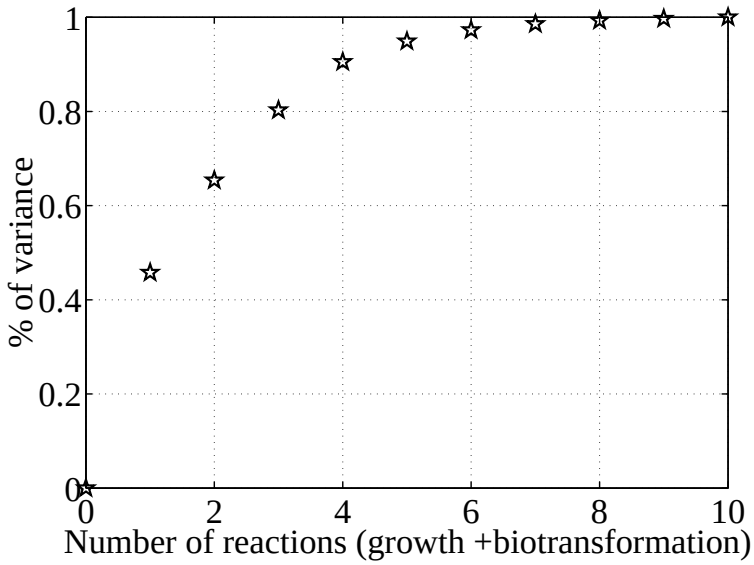

Figure 2: Total variance explained with respect to the number of reactions for the growth and bioproduction of the filamentous fungi *Pycnoporus cinnabarinus*.

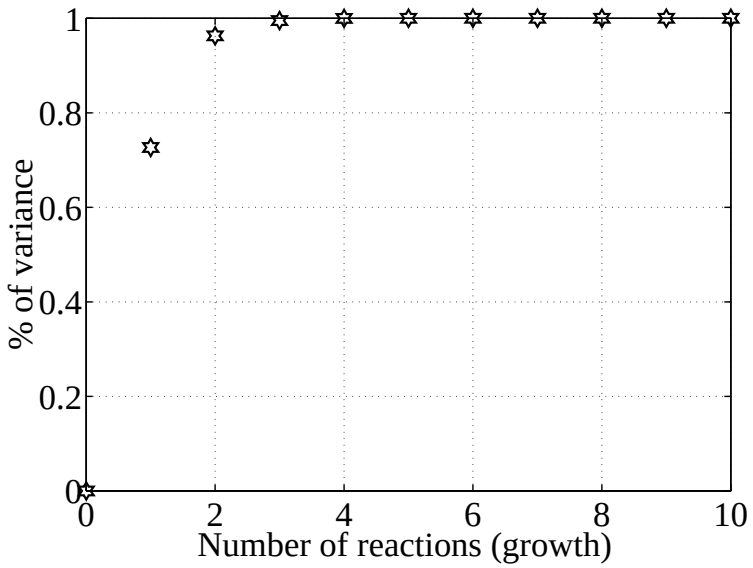

Figure 3: Total variance explained with respect to the number of reactions for the growth of the filamentous fungi *Pycnoporus cinnabarinus*.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399