

**Identification of reaction networks for bioprocesses:
determination of an incompletely known
pseudo-stoichiometric matrix**

Olivier Bernard and Georges Bastin

Bioprocess and Biosystems Engineering,
Vol. 27(5), 2005, pp. 293 - 301.

Identification of reaction networks for bioprocesses: determination of an incompletely known pseudo-stoichiometric matrix

Olivier Bernard¹, Georges Bastin²

¹INRIA-COMORE, BP93, 06902 Sophia-Antipolis Cedex, France, e-mail: olivier.bernard@inria.fr

²UCL-CESAME, av. G. Lemaitre 4-6, 1348 Louvain-La-Neuve, Belgium, e-mail: bastin@auto.ucl.ac.be

The date of receipt and acceptance will be inserted by the editor

Key words Modeling, Nonlinear systems, Bioreactors, Validation

Abstract In this paper we propose a methodology to determine the structure of the pseudo-stoichiometric coefficient matrix K in a mass balance based model and to identify its coefficients from a set of available data. The first stage 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 K . Then we propose a method to directly determine the structure of the matrix (*i.e.* mainly its zeros and the signs of its coefficients). These methods are illustrated with simulations of a process of lipase production from olive oil by *Candida rugosa*.

1 Introduction and motivation

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

$$\frac{d\xi(t)}{dt} = K r(t) + v(t), \quad (1)$$

In this model, the vector $\xi = (\xi_1, \xi_2, \dots, \xi_n)^T$ is made-up of the concentrations of the various species inside the liquid medium. The term $v(t)$ represents the net balance between inflows, outflows and dilution effects. The term $K r(t)$ represents the biological and biochemical conversions in the reactor (per unit of time) according to some underlying reaction network. The $(n \times p)$ matrix K is a constant (pseudo-)stoichiometric matrix. $r(t) = (r_1(t), r_2(t), \dots, r_p(t))^T$ is a vector of reaction rates (or conversion rates). It is supposed to depend on the state ξ and on external environmental factors such as temperature, light or pressure, etc

The pseudo-stoichiometric (PS) matrix K plays a key role in the mass balance modelling. Each column of

the matrix corresponds to a chemical or biological reaction of the underlying reaction network. The coefficients k_{ij} $j = 1, \dots, p$ are associated with the j^{th} reaction. A positive k_{ij} means that the i^{th} species ξ_i is a product of the j^{th} reaction, while a negative k_{ij} means that ξ_i is a substrate of the j^{th} reaction. If $k_{ij} = 0$ the species ξ_i is not involved in the j^{th} reaction.

In this paper, we are concerned with modelling situations where the on-line concentrations ξ_i of the involved species are measured but the structure of the reaction network is *a priori* questionable and therefore the matrix K is partially unknown. The objective, as in [6], is to provide guidelines to the user for the identification of the structure of the reaction network and the determination of the PS matrix K from the available data.

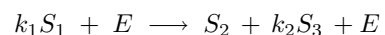
The usual approach dedicated to the determination of reaction networks relies on the linearization of the dynamics around a reference solution [12, 8] and identification of the local jacobian matrix. Here, in the spirit of [7, 3], we exploit the structure of the bioprocesses (equation (1)) and our arguments are not based on any linearizing.

The problem is illustrated with an example.

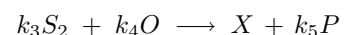
Example: Let us consider the example of a competitive growth on two substrates [13] which could represent, for instance, the production of lipase from olive oil by *Candida rugosa*. Here the microorganism is supposed to grow on two substrates that are produced by the hydrolysis of a primary complex organic substrate.

The following 3-step reaction network has been assumed in the literature:

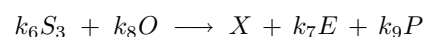
– Hydrolysis:



– Growth on S_2 :



– Growth on S_3 :



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 carbon dioxide.

The associated PS matrix is:

$$K = \begin{pmatrix} -k_1 & 0 & 0 \\ 1 & -k_3 & 0 \\ k_2 & 0 & -k_6 \\ 0 & 0 & k_7 \\ 0 & 1 & 1 \\ 0 & -k_4 & -k_8 \\ 0 & k_5 & k_9 \end{pmatrix}$$

△

with $k_i > 0$, $i = 1, \dots, 9$.

We shall assume that this reaction network is unknown to the user and has to be discovered from data of the species concentrations. Here the data will be simulated by a model but of course in practice the data are obtained from experiments.

Generally, the choice of a reaction network and its associated PS matrix K results from modelling assumptions. Sometimes however, a complete description of the reaction network is *a priori* not available. This can be a consequence of a lack of phenomenological knowledge on some of the involved mechanisms, letting a part of the reaction network questionable. The problem can also arise when it is desired to reduce a complicated given reaction network to a much simpler model. This situation especially occurs for models describing wastewater treatment processes which involve a very large amount of bacterial species and of different molecules to be degraded (see e.g. [10]).

We first propose a method to determine the size of the matrix K i.e. the number of independent reactions that are distinguishable from the available data. Then we show how the structure of the matrix K can be estimated, using the *a priori* available knowledge on the process. By structure we mean the sign and the location of the non-zero entries of the matrix K . In addition, the method can also provide an estimate of the parameters k_{ij} if the available knowledge is sufficient.

2 Determination of the number of reactions

2.1 Introduction

In this section, we intend to determine the minimum number of reactions which are needed in order to explain the observed behaviour of the process, without any prior knowledge on the underlying reaction network. We assume that the vectors $\xi(t)$ of species concentrations and $v(t)$ of inflow/outflow balances are measured during some time interval and exhibit significant variations with time. We assume also that the number of measured variables is larger than the number of reactions: $n > p$.

The PS matrix K and the vector of reaction/conversion rates $r(t)$ are unknown.

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

The model equation (1) can be viewed as a linear dynamical system with state ξ and inputs $r(t)$ and $v(t)$ (although we know obviously that r and v may be state dependent). If we take the Laplace transform of this equation, we get:

$$s\Xi(s) = KR(s) + V(s) \quad (2)$$

where $\Xi(s)$, $R(s)$ and $V(s)$ are the Laplace transforms of $\xi(t)$, $r(t)$ and $v(t)$ respectively. A linear filter or smoother with transfer function $G(s)$ can then be used in order to clean the data (noise reduction, decrease of autocorrelations etc ...):

$$U(s) = KW(s) \text{ with } U(s) = G(s)[s\Xi(s) - V(s)]$$

and $W(s) = G(s)R(s)$. Or, in the time domain:

$$u(t) = Kw(t) \quad (3)$$

with $u(t)$ and $w(t)$ the inverse Laplace transforms of $U(s)$ and $W(s)$ respectively. They can be computed directly from the data by appropriate filtering/smoothing techniques possibly involving delay operators.

For example, the moving average is a very simple filter that can be applied to (1), and provides an expression of the form (3) with (T denotes the considered moving average window):

$$u(t) = \frac{1}{T} \left[\xi(t) - \xi(t-T) - \int_t^{t-T} v(\tau) d\tau \right] \quad (4)$$

and

$$w(t) = \frac{1}{T} \left[\int_{t-T}^t r(\tau) d\tau \right]$$

Now the question of the dimension of the matrix K can be formulated as follows: what is the dimension of the image of K , in other words, what is 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 behaviour could be obtained with a matrix K of lower dimension, by defining other appropriate reaction rates. The determination of the dimension of the $u(t)$ space is a classical problem in statistical analysis. It corresponds to the principal component analysis (see e.g. [11]) that determines the dimension of the vectorial space spanned by the vectors k_i which are the rows of K . To reach this objective, we consider the $n \times N$ matrix U obtained from a set of N estimates of $u(t)$:

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

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

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

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). We consider more time instants t_i than state variables: $N > n$.

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

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

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

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

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

Moreover, the eigenvectors associated with the σ_i generate an orthonormal basis of $\text{Im}K$.

Indeed, it is a direct application of the singular decomposition theorem [9] since $\text{rank}(M) = \text{rank}(KW) = \text{rank}(K) = \text{rank}(\Sigma) = p$.

Now from a theoretical point of view, it is clear that the number of reactions can be determined by counting the number of non zero singular values of UU^T .

2.3 Practical implementation

In practice, the ideal case presented above 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 “true” matrix K is probably much larger. The reactions that are fast or of low magnitude can be considered as perturbations of a dominant low dimensional reaction network that we are actually trying 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 not available.
- In order to compute $u(t)$ we need a numerical implementation of the filter $G(s)$. Moreover an interpolation is often required to estimate the values of $\xi(t_i)$ and $v(t_i)$ at the same time instants t_i . These processes generate additional perturbations.

2.3.1 Data normalisation In order to avoid conditioning problems and to give the same weighting to all the variables, the data vectors $u(t_i)$ are normalised 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.

2.3.2 Practical determination of the number of reactions

In practice, for the reasons we have mentioned above, it is well known that 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 produce a reasonable approximation of the data $u(t)$. To answer that question, let us remark that the eigenvalues σ_i of M correspond to the variance associated with the corresponding eigenvector (inertia axis) [11].

The method then consists in selecting the p first principal axis which represent a total variance larger than a fixed confidence threshold.

For instance, in the next example, we will consider a threshold (depending on the information available on noise measurements) at 95% of the variance. This leads to the selection of 3 axis, and therefore $p = 3$.

Remark: if $\text{rank}(M) = n$ it means that $\text{rank}(W) \geq n$. In such a case we cannot estimate p and measurements of additional variables are requested in order to apply the method presented here.

Parameter	Values	Units
c_0	0.5	g/l.day ⁻¹
c_1	3	day ⁻¹
c_2	1	g/l
c_3	0.2	g/l
c_4	20	g.day ⁻¹ l ⁻¹
c_5	1	g/l
c_6	0.2	g/l
c_7	2	g ² /l ²
c_8	2	g/l
c_9	0.2	g/l
c_{10}	5	day ⁻¹
c_{11}	15	g/l
c_{12}	0.5	day ⁻¹
c_{13}	0.5	g/l

Table 1 Parameter values.

2.3.3 Example: competitive growth on two substrates We come back to the example which has been introduced above. For the simulation purpose, we assume that the kinetics of the three reactions are given by the following expressions :

$$\phi_1(S_1, E) = c_0 \frac{S_1}{S_1 + c_8} \frac{E}{E + c_9}$$

$$\phi_2(S_2, O, X) = c_1 \frac{S_2}{S_2 + c_2} \frac{O}{O + c_3} X$$

$$\phi_3(S_2, O, X) = c_4 \frac{S_3}{(S_3 + c_5)(S_2 + c_6)} \frac{O^2}{O^2 + c_7};$$

The transfer between liquid and gaseous phase is represented by the classical Henry's law:

$$q_{CO_2} = c_{10}(P - c_{11}) \text{ and } q_{O_2} = c_{12}(O - c_{13})$$

The values of the coefficients c_i can be found in Table 1. The matrix K is chosen as follows:

$$K = \begin{pmatrix} -3 & 0 & 0 \\ 1 & -5 & 0 \\ 0.3 & 0 & -0.5 \\ 0 & 0 & 0.2 \\ 0 & 1 & 1 \\ 0 & -2 & -1 \\ 0 & 0.3 & 1.5 \end{pmatrix}$$

A 30 day run of the model has been performed. The collected data have been corrupted with a white noise of high magnitude (30% of the standard deviation of each component) and sampled. Finally 380 data points are available.

The data (before sampling) are presented on Figure 1. The state variables S_2 , S_3 , E , X , P , O and of the gaseous flow rates q_{O_2} and q_{CO_2} have been measured. We assume here that the state variable S_1 was not recorded in order to illustrate the fact that our approach is applicable even if the full set of state variables is not available for measurement. Moreover the dilution rate and the substrate inflow rate have been selected in order to guarantee that the system is enough excited and therefore that the recorded signals will have a sufficiently informative content to expect good identification results. The vectors $u(t_i)$ are then computed from these data and subsequently normalised as explained before. Finally, the eigenvectors of UU^T are computed.

Figure 2 represents the cumulated variance associated with the number of considered inertia axis. For instance, we can see that two reactions are sufficient to explain 82% of the observed variance. Since three reactions explain 95% of the total variance, it seems reasonable in this example to use 3 reactions for the model.

The reader can refer to [4] for an application to real data, for growth and vanillin production by cultures of the fungus *Pycnoporus cinnabarinus* in bioreactors.

3 Estimation of the pseudo-stoichiometric matrix K

Now that we have a value for the number of involved reactions, we are in a position to start the estimation of the (totally or partially) unknown matrix K .

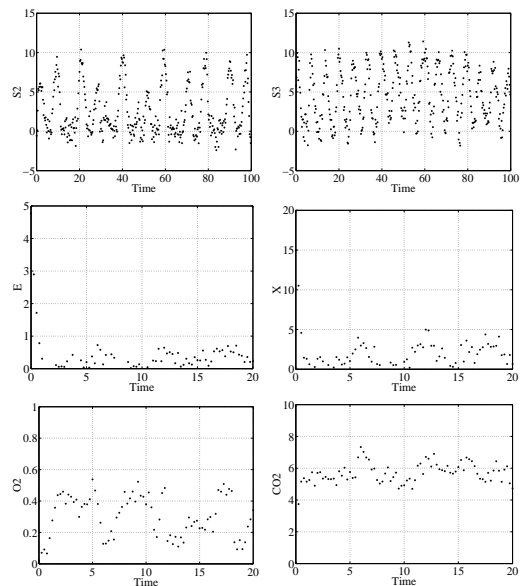


Fig. 1 Experiment simulated from the kinetic modelling corrupted with an additive white noise.

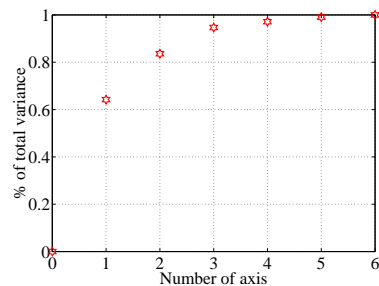


Fig. 2 Total variance explained with respect to the number of reactions for the production of lipase from olive oil by *Candida rugosa*.

3.1 Determination of ImK

Let us use the property 1 which states that ImK is spanned by the eigenvectors ρ_i associated with the non zero eigenvalues of UU^T . Now, from the experimental data collected through the matrix UU^T we get p eigenvectors ρ_i that span K . It means that each column k_i of K is a linear combination of the ρ_i . In other terms, there exists a $p \times p$ matrix G such that

$$K = \rho G$$

where the columns of matrix ρ are the eigenvectors ρ_j . In other words, the family of possible PS matrices K is parameterised by G .

Remark: In general, since the reaction rates are unknown, the matrix G (and therefore the matrix K) is not identifiable: this can be easily understood on a very simple example. If $r_1(\xi)$ and $r_2(\xi)$ are two reaction rates,

the term $Kr(\xi)$ can be written:

$$\begin{aligned} Kr(\xi) &= k_1 r_1(\xi) + k_2 r_2(\xi) \\ &= \frac{k_1+k_2}{2} (r_1(\xi) + r_2(\xi)) + \frac{k_1-k_2}{2} (r_1(\xi) - r_2(\xi)) \end{aligned}$$

And therefore both matrices $K = [k_1 \ k_2]$ and $\bar{K} = [\frac{k_1+k_2}{2} \ \frac{k_1-k_2}{2}]$ can produce the same result. The reaction rates associated with the second matrix are then: $\bar{r}_1(\xi) = r_1(\xi) + r_2(\xi)$ and $\bar{r}_2(\xi) = r_1(\xi) - r_2(\xi)$.

3.2 Additional hypotheses

In order to make the matrix G (and K) uniquely identifiable, we need to introduce additional structural constraints. At this stage, all the *a priori* knowledge on the reaction network should be considered to improve the estimation process.

3.2.1 Normalisation First, we shall impose (without loss of generality) that each reaction rate is normalised with respect to one species, and therefore that each column of the matrix K contains one +1 or one -1. This induces obviously additional constraints on the possible matrices G . Note that sometimes we may not know the sign of the element: the two possible cases must then be considered.

3.2.2 Physical assumptions One can impose the conservation of elementary mass balances, or at least only allow for a leak of mass in the system.

3.2.3 Biological and biochemical assumptions When additional constraints are still necessary, we use biochemical assumptions.

When only a subset of the components are in presence at the initial time, the production or not of new components with consumption or not of substrates is an indicator of the variables that are necessary for the reaction. It is clear for example that the first reaction will involve only the substrates which were present at the beginning of the fermentation.

We can for example deduce from this that a specific component is not involved in one of the p reactions (meaning that there is a zero in K).

3.2.4 Other assumptions One can also try to find a matrix K involving the minimum number of components in each reaction (*i.e.* containing the maximum number of zeros). If these hypotheses are not sufficient, several matrices K can then be identified, parameterised by some parameter, and their biochemical meaning must then be assessed.

3.3 Validation

The main result provided by the previous analysis is the determination of the variables which are substrates or products in the reactions or, in other words, the obtained signs of the entries of K .

Another expected result can be the determination of the variables which are not involved in a reaction, corresponding to zero elements in the matrix K . However it is actually very unlikely that the analysis will provide estimates of the elements of K which are exactly zero. The idea consist then in replacing the very small elements by zeros, and to validate the corresponding reaction network using the techniques presented in [3, 5]. These methods are based on the 5 following steps [5]:

1. Determination of the vectors which are in the left Kernel of K , *i.e.* the vectors $\lambda \in \mathbf{R}^n$ such that:

$$\lambda^t K = 0 \quad (5)$$

2. Determination of the associated regression, which is directly deduced from (3) and (5):

$$\sum \lambda_i u_i(t) = 0 \quad (6)$$

3. Verification that the $u_i(t)$ involved in (6) are not related by any other linear relation (soundness of λ)
4. Test of the significance of regression (6) with the available data
5. Test of the positivity of the PS coefficients identified in the previous step.

3.4 Example (continued)

3.4.1 Statement of the problem and considered data We shall now illustrate the proposed approach with the simulation study of lipase production from olive oil. From the previous study of the number of reactions, we know that 3 reactions should be considered.

We assume here that the first reaction is known, and therefore we only focus on the two other reactions. We are thus in the process of estimating the submatrix \bar{K} extracted from K by removing the first line and the first column.

\bar{K}	identified \bar{K}	identified \bar{K} after validation
$\begin{pmatrix} -5 & 0 \\ 0 & -0.5 \\ 0 & 0.2 \\ 1 & 1 \\ -2 & -1 \\ 0.3 & 1.5 \end{pmatrix}$	$\begin{pmatrix} -3.54 & 0 \\ 0 & -0.51 \\ 0.01 & 0.22 \\ 1 & 1 \\ -1.34 & -0.87 \\ 0.18 & 1.51 \end{pmatrix}$	$\begin{pmatrix} -4.54 & 0 \\ 0 & -0.50 \\ 0 & 0.19 \\ 1 & 1 \\ -1.33 & -0.72 \\ 0.34 & 1.24 \end{pmatrix}$

Table 2 True coefficients of matrix \bar{K} and identified values.

A set of noisy data of the state variables S_2 , S_3 , E , X , P , O and of the gaseous flow rates q_{O_2} and q_{CO_2} is produced by simulation as described in Section 2. The goal is to determine the 6×2 matrix \bar{K} from this data set. More specifically, a question that we want to address is to determine, from the data, which of the two reactions produces the enzyme E .

3.4.2 Estimation of \bar{K} Now we can compute the quantities U_i associated with the 6 state variables using a moving average. Next we compute the matrix $M = U^T U$. The eigenvectors ρ_i associated with the two largest eigenvalues are then the basis of $\mathcal{Im}K$. Since G is a 2×2 matrix, the columns k_1 and k_2 of \bar{K} can be written:

$$k_1 = \alpha_{11}\rho_1 + \alpha_{12}\rho_2 \text{ and } k_2 = \alpha_{21}\rho_1 + \alpha_{22}\rho_2 \quad (7)$$

Now we proceed in two successive steps:

i. Normalisation.

The PS coefficients associated with the biomass growth are normalised : $k_{14} = 1$ and $k_{24} = 1$. We get then:

$$\begin{aligned} k_{14} = 1 &= \alpha_{11}\rho_{14} + \alpha_{12}\rho_{24} \\ k_{24} = 1 &= \alpha_{21}\rho_{14} + \alpha_{22}\rho_{24} \end{aligned} \quad (8)$$

Using equations 7 and 8 with the obtained values of ρ_1 and ρ_2 , we can now write matrix \bar{K} parametrised by α_{11} and α_{22} as follows:

$$\bar{K} = \begin{pmatrix} -1.42\alpha_{11} - 2.65 & -1.2\alpha_{22} + 1.12 \\ 0.2\alpha_{11} - 0.13 & 0.17\alpha_{22} - 0.67 \\ -0.08\alpha_{11} + 0.062 & -0.071\alpha_{22} + 0.28 \\ 1 & 1 \\ -0.19\alpha_{11} - 1.2 & -0.16\alpha_{22} - 0.72 \\ -0.53\alpha_{11} + 0.51 & -0.45\alpha_{22} + 1.93 \end{pmatrix}$$

ii. Biological hypotheses.

Now to determine uniquely matrix \bar{K} two additional assumptions must be introduced.

Hypothesis: *A reaction still takes place when only S_2 [resp. S_3] is present at the initial time, and no S_3 [resp. S_2] is produced.*

In other words this means that S_2 is the only substrate of one reaction and that S_3 is the only substrate of the other one. Thus we will impose $k_{12} = 0$ and $k_{21} = 0$.

These additional constraints allows us to compute α_{11} (0.621) and α_{22} (0.93).

Finally we end up with an estimate of the matrix \bar{K} (see Table 2). It is worth noting that the identified matrix \bar{K} is close to the true one. The value of the (theoretically zero) coefficient k_{13} is 0.01 which can be neglected with respect to the other coefficients of \bar{K} . Hence, the unknown part of the structure of the matrix \bar{K} has been recognised. Moreover the estimates of the non-zero entries of the matrix \bar{K} are quite accurate.

3.4.3 Validation Here we will validate the identified structure for K with respect to the available data. As it was shown in the previous step, the following structure for matrix K has been identified:

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

Now the kernel of K^T is spanned by the following 4 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 associated regressions are the following:

$$\begin{aligned} \mathcal{R}_1 : \frac{k_7}{k_6} u_3(t) + u_4(t) &= 0 \\ \mathcal{R}_2 : \frac{k_8 - k_4}{k_3} u_2(t) + k_8 u_5(t) + u_6(t) &= 0 \\ \mathcal{R}_3 : -\frac{k_7}{k_3} u_2(t) + u_4(t) - k_7 u_5(t) &= 0 \\ \mathcal{R}_4 : \frac{k_5 - k_9}{k_3} u_2(t) - k_9 u_5(t) + u_7(t) &= 0 \end{aligned} \quad (9)$$

Note that these regressions are sound [5] in the sense that they do not involve a set of components that are related together by another linear relationship.

The numerical results obtained from the considered regressions are presented in Table 3. It results that all the regressions (9) are highly significant, showing that the estimated reaction network is validated.

Moreover, the following quantities are estimated in Table 3:

$$\left\{ \frac{k_7}{k_6}, \frac{k_8 - k_4}{k_3}, k_8, \frac{k_7}{k_3}, k_7, \frac{k_5 - k_9}{k_3}, k_9 \right\}$$

It is easy to compute the values of k_3 to k_9 from this set, it leads to the estimate of matrix \bar{K} proposed in Table 2. The validation step consists in verifying that the estimates of the PS coefficients are all positive. This concludes the validation procedure.

4 Conclusion

Determining a reaction network for a bioprocesses is a difficult issue mainly because of the complexity inherent to biological systems. This problem is fundamentally ill

	Significance	Unknown	Value	Interval
\mathcal{R}_1	YES	$\frac{k_7}{k_6}$	2.36	[2.06 2.66]
\mathcal{R}_2	YES	$\frac{k_8-k_4}{k_3}$ k_8	-0.14 0.72	[-0.18 0.09] [0.66 0.77]
\mathcal{R}_3	YES	$\frac{k_7}{k_3}$ $-k_7$	-0.04 -0.2	[-0.05 -0.03] [-0.21 -0.18]
\mathcal{R}_4	YES	$\frac{k_5-k_9}{k_3}$ $-k_9$	-0.35 -1.25	[-0.42 -0.29] [-1.33 -1.16]

Table 3 Significance (threshold 5%) of the regressions (9) and parameter values.

stated since the Pseudo-stoichiometric matrix K is generally not identifiable from a data set. We show in this paper how to identify the space generated by the columns of K and how to add constraints in order to determine a unique (or a set of) matrix K .

Through the studied example we have demonstrated that the proposed method can accurately estimate the values of the PS coefficients in spite of noises due to measurements and low sampling frequency.

It is worth noting that this approach does not necessarily need the complete measurement of the state variables ξ . Of course, if the measurement of the i^{th} biochemical component ξ_i is not available the i^{th} line of matrix K won't be determined by the method.

Acknowledgement: 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 TELEMAT 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.F. VanImpe, *Nonlinear and adaptive control in biotechnology: a tutorial*, European Journal of Control **1** (1995), no. 1, 1-37.
3. O. Bernard and G. Bastin, *Structural identification of nonlinear mathematical models for bioprocesses*, Proceedings of the Nonlinear Control Systems Symposium 98, Enschede, July 1-3, 1998, 1998, pp. 449-454.
4. 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 (1999), 558-571.
5. O. Bernard and G. Bastin, *First step in mass balance modeling of bioprocesses: estimation and validation of the stoichiometric matrix*, Math. Biosciences (submitted).
6. P. Bogaerts and A. Vande Wouwer, *Systematic generation of identifiable macroscopic reaction schemes*, Pro-

ceedings of the 8th IFAC Conference on Computer Applications in Biotechnology (CAB8), Montral, Canada, 2001.

7. L. Chen and G. Bastin, *Structural identifiability of the yield coefficients in bioprocess models when the reaction rates are unknown*, Math. Biosciences **132** (1996), 35-67.
8. T. Chevalier, I. Schreiber, and J. Ross, *Toward a systematic determination of complex reaction mechanisms*, J. Phys. Chem **97** (1993), 6776 - 6787.
9. R. A. Horn and C. R. Johnson, *Matrix analysis*, Cambridge University Press, Cambridge MA, 1993.
10. IWA Task Group for Mathematical Modelling of Anaerobic Digestion Processes, *Anaerobic digestion model no. 1 (adm1)*, IWA Publishing, London, 2002.
11. R. A. Johnson and D. W. Wichern, *Applied multivariate statistical analysis*, Prentice Hall, 1992.
12. A. Freund M. Eiswirth and J. Ross, *Mechanistic classification of chemical oscillators and the role of species*, Advances in Chemical Physics, vol. 80, ch. 1, pp. 127-199, Wiley, New-York, 1991.
13. P. Serra, J.L. del Rio, J. Robust, M. Poch, C. Sola, and A. Cheruy, *A model for lipase production by *Candida rugosa**, Bioprocess Engineering **8** (1992), 145-150.