CAN WE IDENTIFY BIOTECHNOLOGICAL PROCESSES?

G. Bastin, L. Chen and V. Chotteau

Centre for Systems Engineering and Applied Mechanics, Université Catholique de Louvain,
Bâtiment Maxwell, Place du Levant, 3 B-1348 Louvain-la-Neuve, Belgium

Abstract

The identification problem of biotechnological processes is threefold : (i) determination of the number of biological reactions, (ii) identification of the underlying reaction network, (iii) identification of the kinetics. In most practical cases, these three parts of the identification problem can be completely decoupled from one another.

1 Introduction

The identification of mathematical models for biological processes in stirred tank reactors is known to be one of the major bottlenecks for the application of modern control science in biotechnology.

Biological models usually involve two kinds of parameters : the yield coefficients which rely on the *structure of the underlying reaction network* and the kinetic coefficients which rely on the *structure of the kinetics*.

In most modelling studies, however, the difference of nature between these two kinds of parameters is not taken into account and the set of both yield and kinetic coefficients is treated as a whole in the identification exercise, often leading to intricate identifiability difficulties.

It has been recently shown by several authors ([3], [2], [5], [6]) how to implement a *two-step* procedure for identifying separately the reaction structure and the kinetic structure for a general class of dynamical models of bioprocesses. In the present paper, we intend to give a tutorial presentation of this two-step approach and to analyse some of its main features.

The suggested identification procedure is actually based on a state transformation which allows to reformulate the dynamical model into separate submodels.

The first submodel only depends on the reaction structure and is independent of the kinetics. It can be linearly reparametrized and used for the identification of the yield coefficients by means of linear regression. The identifiability properties of this submodel are analyzed in [3] and [6].

Once the reaction structure and the yield coefficients are known, the second submodel is used for the identification of the kinetic structure. This submodel is in a form which enables to decouple completely the kinetic functions from one another. This means that each biological reaction occuring in the reactor can be treated separately as if it was the only one, although all the involved reactions obviously take place simultaneously.

We begin with a preliminary example.

2 A preliminary example.

Consider the growth of a population of microorganisms on a single limiting substrate in a stirred tank in batch mode. The growth of the biomass is accompanied by the formation of a product of interst. This growth reaction is described by the following scheme :

$$k_1 S \longrightarrow X + k_2 P \tag{1}$$

where S, X, P denote respectively the limiting substrate, the biomass and the product, while k_1 and k_2 are two *yield coefficients* which are usually expressed as units of mass per units of mass in biological systems. This reaction thus means that the consumption of k_1 mass units of substrate is necessary to form 1 unit of biomass and k_2 units of product. The name "yield" adopted here instead of "stoichiometric" emphasizes the fact that in biological systems, these coefficients are in general expressed in units of mass instead of number of moles for purely chemical systems. To simplify the notations, we assume that X, S, P also denote the *concentrations* of the related species in the reactor. According to the scheme (1) and since we are in here a batch reactor, the *conversion rate* of substrate into biomass and product can then be expressed as :

$$\frac{dS}{dt} = -k_1 \frac{dX}{dt} \tag{2}$$

$$\frac{dP}{dt} = k_2 \frac{dX}{dt} \tag{3}$$

From these equations, k_1 and k_2 can be expressd as :

$$k_1 = \frac{S_0 - S_t}{X_t - X_0} \tag{4}$$

$$k_2 = \frac{P_t - p_0}{X_t - X_0} \tag{5}$$

for all $t > 0$, where the subscripts $_0$ and $_t$ denote, respectively, the initial value and the value at time t of each species. In consequence, if measurements of X, S and P are available, k_1 and k_2 can be identified by linear regression using relationships (4) and (5).

Assume now that the conversion rate is modelled by *Michaëlis Menten kinetics* of the form :

$$r(S, X) \triangleq \frac{\mu_{\max} S X}{K_m + S} \tag{6}$$

where $\mu_{\max}$ (the maximum growth rate) and K_m (the Michaëlis constant) are two *kinetic coefficients*.

It follows that, the process dynamics can now be written in the form :

$$\frac{dX}{dt} = r(S, X) \tag{7}$$

$$\frac{dS}{dt} = -k_1 r(S, X) \tag{8}$$

$$\frac{dP}{dt} = k_2 r(S,X) \tag{9}$$

with $r(S,X)$ given by (6). This is called a *state-space model* in the terminology of system theory. Then, by integrating (8), we obtain the following linear equation in $\mu_{\max}$ and K_m :

$$S_0 - S_t = K_m ln\frac{S_t}{S_0} + k_1\mu_{\max}\int_0^t X(t)dt$$

it is clear that the two coefficients $\mu_{\max}$ and K_m can also be identified by linear regression from the above equation as long as measurements of S and X are available.

Comment 2.1. With this preliminary example, we see that, whenever measurements of X, S, P are available (i.e. *full-state* measurement), we can perform the parameter identification in two separate and successive steps : first the identification of the yield coefficients k_1 and k_2 (which rely on the reaction scheme (1)) independently of any knowledge or assumption regarding the kinetics; second the identification of the kinetic coefficients $\mu_{\max}$ and K_m (which rely on the structure of the reaction rate model (6)).

The generalization of this two-step identification method to complex biotechnological systems is not straightforward. The difficulty is that, whenever more than one reaction is involved in the system, there is no longer equivalence between the yield coefficients and the yield of the process. Morever, when the system is not in batch mode nor in steady state, the transport dynamics also have to be taken into account.

3 Reaction Networks

A bioprocess in a stirred tank reactor is described by a set of m coupled microbiological and biochemical reactions which take place simultaneously in the reactor and which involve a set of n biological species such as microorganisms, substrates, metabolites, enzymes... The n components are denoted : $X_1, X_2, \ldots, X_n$

The *reaction network* is then a set of biological reactions of the form :

$$\begin{aligned} k_{1j}^- X_1 + k_{2j}^- X_2 + \ldots + k_{nj}^- X_n \longrightarrow k_{1j}^+ X_1 \\ + k_{2j}^+ X_2 + \ldots + k_{nj}^+ X_n \quad j = 1,m \end{aligned} \tag{10}$$

where k_{ij}^- and k_{ij}^+ are the *yield coefficients* of the i^{th} component in the j^{th} reaction.

A *substrate* or *reactant* is a component which appears in the left-hand side of a reaction with a non-zero yield coefficient k_{ij}^-.

A *product* is a component which appears in the right-hand side of a reaction with a non-zero yield coefficient k_{ij}^+.

A *catalyst* (usually an enzyme in biotechnology) is a component which appears on both sides of the same reaction with identical non-zero yield coefficients $k_{ij}^- = k_{ij}^+$.

Obviously the same component can be a product of a reaction and a substrate of another one (see the examples below). It is clear also that, without loss of generality, the yield coefficients can always be *normalised* in such a way that one of the non-zero yield coefficients is exactly equal to "1" in each reaction. The corresponding component is then called the *normalising* component of the related reaction.

Definition 3.1. The *characteristic matrix* $\boldsymbol{K} \triangleq [k_{ij}]$ of a biological process described by a reaction network of the form (10) is the $(n \times m)$-dimensional matrix with entries $k_{ij} \triangleq k_{ij}^+ - k_{ij}^-$ $(i = 1,n; j = 1,m)$ ■

Example 3.2. Competitive growth of a single population on two limiting substrates. We consider the example of a process in which a population of microorganisms can grow on two different secondary substrates produced by the hydrolysis of a primary complex organic substrate. The process is described by the following reaction network with $m = 3$ reaction and $n = 7$ species (components):

$$\text{Hydrolysis :} \quad k_1 S_1 + E \longrightarrow S_2 + k_2 S_3 + E \tag{11}$$

$$\text{Growth 1 :} \quad k_3 S_2 + k_4 O \longrightarrow X + k_5 P \tag{12}$$

$$\text{Growth 2 :} \quad k_6 S_3 + k_8 O \longrightarrow X + k_7 E + k_9 P \tag{13}$$

with S_1 representing a primary substrate, S_2 and S_3 the secondary substrates, E enzyme, O dissolved oxygen, X biomass and P carbon dioxyde. Assuming that the components are ordered as follows : $S_1, S_2, S_3, E, X, O, P$, the characteristic matrix is written :

$$\boldsymbol{K} = \begin{pmatrix} -k_1 & 0 & 0 \\ 1 & -k_3 & 0 \\ k_2 & 0 & -k_6 \\ 0 & 0 & k_7 \\ 0 & 1 & 1 \\ 0 & -k_4 & -k_8 \\ 0 & k_5 & k_9 \end{pmatrix} \tag{14}$$

The normalizing components are S_2 for the first reaction and X for the second and the third ones. This normalization induces the "structural" presence of "1" in each column of the matrix $\boldsymbol{K}$. ■

4 Reaction Kinetics

Let us now assume that a biological process, described by a reaction network of the form (10), takes place in a stirred tank bioreactor. The *reaction kinetics* (or *reaction rate* or *conversion rates*, see Section 2 for a simple example) are simply the rates at which the reactions proceed, that is the rates of substrate consumption and product formation in the tank. We thus assume that reaction kinetics denoted $r_j(t)$ $(j = 1,m;$ units of mass/unit of time) are associated with the reactions of the network.

It is a well established fact that these reaction kinetics are affected by the concentrations of the biological components in the bioreactor. Therefore they are generally represented by non-negative rational functions of the concentrations as :

$$r_j(t) \triangleq r_j(X_1(t), X_2(t), \ldots, X_n(t)) \;\; j = 1,m$$

where (by a slight abuse of language) $X_1, X_2, \ldots, X_n$ now denote the *concentrations* of the components in the liquid phase of the reactor. As usual, the argument "t" (time) will be omitted in the sequel.

In a vast majority of applications, the kinetic model take a multiplicative form :

$$r_j(X_1, X_2, \ldots, X_m) = \mu_j^* \Pi_{i=1}^n \rho_{ij}(X_i) \tag{15}$$

where μ_j^* is a nominal (or specific) *rate constant* while the terms $\rho_{ij}(X_i)$ separately represent the effect of each component X_i on the rate r_j. The form given to the functions $\rho_{ij}(X_i)$ depends on the way the j^{th} reaction is affected by the component concentration X_i. One has basically three situations :

1. The rate r_j is *not* affected by X_i. This is trivially represented by : $\rho_{ij}(X_i) = 1$.
2. The rate r_j is *positively* affected by X_i in the sense that an increase of X_i enhances the reaction rate as shown in fig. 1. In such a case, X_i is called an *activator* of the reaction. This is mathematically represented by selecting a monotonic positive increasing function for $\rho_{ij}(X_i)$:

$$\rho_{ij}(X_i) \geq 0 \quad \frac{\partial \rho_{ij}(X_i)}{\partial X_i} > 0 \quad \frac{\partial^2 \rho_{ij}(X_i)}{\partial X_i^2} < 0$$

3. The rate r_j is *negatively* affected by X_i in the sense that an increase of X_i slackens the reaction rate as shown in fig.2. In such a case, X_i is called an *inhibitor* of the reaction. This is mathematically represented by selecting a monotonic positive decreasing function for $\rho_{ij}(X_i)$:

$$\rho_{ij}(X_i) \geq 0 \quad \frac{\partial \rho_{ij}(X_i)}{\partial X_i} < 0 \quad \frac{\partial^2 \rho_{ij}(X_i)}{\partial X_i^2} > 0$$

As a matter of fact, these three situations can be described by the same *elementary kinetic function* taking the form of the following first order rational fraction:

$$\rho_{ij}(X_i) \triangleq \frac{\alpha_{ij} + \beta_{ij} X_i}{\gamma_{ij} + \delta_{ij} X_i} \qquad (16)$$

This function involves four positive constant coefficients $\alpha_{ij}, \beta_{ij}, \gamma_{ij}, \delta_{ij}$ that we call *kinetic coefficients*. It is easily verified (see fig. 1 and 2) that :

$$\begin{aligned} \rho_{ij}(0) &= \frac{\alpha_{ij}}{\gamma_{ij}} \\ \rho_{ij}(\infty) &= \frac{\beta_{ij}}{\delta_{ij}} \end{aligned}$$

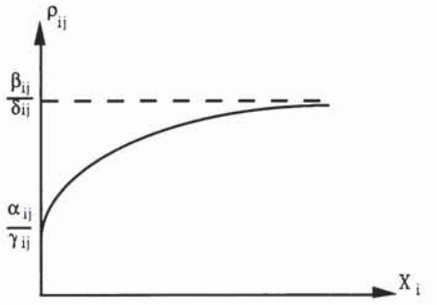

Fig.1. Elementary rate function for an activator

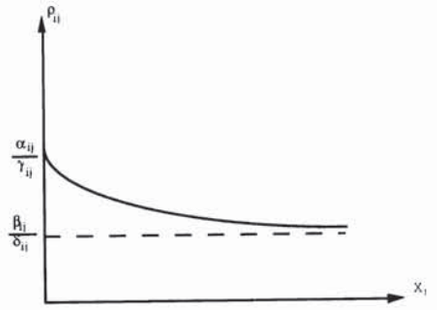

Fig.2. Elementary rate function for an inhibitor

It appears clearly that this model (16) represents the kinetics of an activator when $\frac{\beta_{ij}}{\delta_{ij}} > \frac{\alpha_{ij}}{\gamma_{ij}}$. and of an inhibitor when $\frac{\beta_{ij}}{\delta_{ij}} < \frac{\alpha_{ij}}{\gamma_{ij}}$.

As for the yield coefficients, the kinetic coefficients can also be normalized in such a way that either α_{ij} or β_{ij} and either γ_{ij} or δ_{ij} is taken equal to "1" in each elementary kinetic function. Moreover, by setting some of the other (unnormalised) coefficients to zero, we can obtain various interesting special cases which are commonly used in modelling of bioprocesses, as shown in the following table.

Model	α_{ij}	β_{ij}	γ_{ij}	δ_{ij}
Linear X_i	0	1	1	0
Michaëlis Menten $\frac{X_i}{K_M+X_i}$	0	1	K_M	1
Hyperbolic inhibition $\frac{K_P}{K_P+X_i}$	1	0	1	K_P^{-1}

Finally, we note that when a particular component has to be considered as an activator at low concentrations and as an inhibitor at high concentrations, this is easily achieved by combining two normalized elementary kinetic function of the form (16) as follows (see Fig. 3) :

$$\begin{aligned} \rho_{ij}(X_i) &= \left(\frac{\alpha_{ij} + X_i}{\gamma_{ij} + X_i}\right)\left(\frac{1 + \beta_{ij} X_i}{1 + \delta_{ij} X_i}\right) \text{ with } \frac{\alpha_{ij}}{\gamma_{ij}} < 1 \\ &\text{and } \frac{\beta_{ij}}{\delta_{ij}} < 1 \qquad (17) \end{aligned}$$

The particular case of Haldane kinetics correspond to the special case : $\alpha_{ij} = 0, \beta_{ij} = 0, \gamma_{ij} > 0, \delta_{ij} > 0$

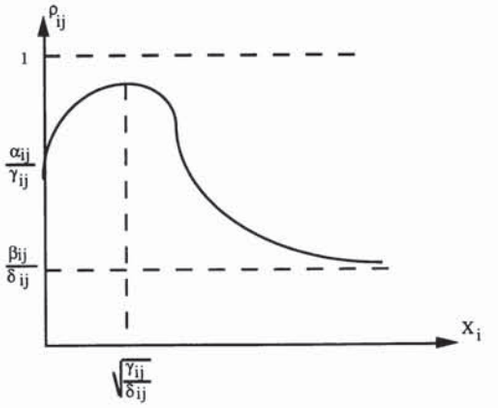

Fig.3 Rate function with activation at low concentrations and inhibition at high concentrations.

Example 4.1. Consider again the preliminary example of Section 2 where the growth is modelled by Michaëlis-Menten kinetics (6). With the formalism just introduced, we have the following equivalence :

$$r(S,X) = \frac{\mu_{\max} S X}{K_M + S} = \mu^* \rho_1(S) \rho_2(X)$$

with :

$$\begin{aligned} \mu^* &= \mu_{\max} \\ \rho_1(S) &= \frac{\alpha_1 + \beta_1 S}{\gamma_1 + \delta_1 S} \quad \alpha_1 = 0 \ \beta_1 = 1 \ \gamma_1 = K_M \ \delta_1 = 1 \\ \rho_2(X) &= \frac{\alpha_2 + \beta_2 X}{\gamma_2 + \delta_2 X} \quad \alpha_2 = 0 \ \beta_2 = 1 \ \gamma_2 = 1 \ \delta_2 = 0 \end{aligned}$$

■

The overall structure of the kinetics of a bioprocess can be summarized in a *table of influences*. This table is an array of dimensions $n \times m$ where the entry (i, j) is marked by one of the following symbols :

- "0" if the reaction rate r_j is not affected by X_i
- "+" if X_i is an activator of the j^{th} reaction
- "−" if X_i is an inhibitor of the j^{th} reaction
- "±" if X_i is an activator at low concentrations and an inhi at high concentrations.

Example 4.2. Consider a bioprocess made up of two simple reactions :

$$\begin{aligned} &\text{Growth :} && k_1 S \longrightarrow X \\ &\text{Enzymatic production :} && k_2 S \longrightarrow P \end{aligned}$$

Obviously, in accordance with the principle of mass action, we have to assume that the substrate S is an activator of both reactions and that the biomass X is an activator of its own production. Furthermore, we suppose that the substrate S becomes inhibitor of the enzymatic production at high concentrations while the product P inhibits the growth. On the basis of these modelling assumptions, we have the following table of influences.

	r_1	r_2
S	+	±
X	+	+
P	0	−

5 The general dynamical model of bioprocesses in stirred tank reactors

In this section, we briefly recall the general dynamical model of biological reactors as described in [1]. The reactions which occur in the reactor are supposed to be encoded into a reaction network of the form (10). The model expresses the mass balance of the various components inside the reactor. The vector of the component concentrations is called the *composition* of the reactor and is the state of the model. It is denoted :

$$\boldsymbol{\xi} = (X_1, X_2, \ldots, X_n)^T$$

The vector of the reaction kinetics is denoted :

$$\boldsymbol{r}(\boldsymbol{\xi}) = (r_1(\boldsymbol{\xi}), r_2(\boldsymbol{\xi}), \ldots, r_m(\boldsymbol{\xi}))^T$$

With these notations, the mass balance dynamics of the process components in a stirred tank reactor are described by the following nonlinear state space model (see [1], Chapter 1, for further details) :

$$\frac{d\boldsymbol{\xi}}{dt} = \boldsymbol{K}\boldsymbol{r}(\boldsymbol{\xi}) + \boldsymbol{u} \tag{18}$$

The first term $\boldsymbol{K}\boldsymbol{r}(\boldsymbol{\xi})$ represents the biological and biochemical conversions in the reactor (per unit of volume) according to the underlying reaction network ($\boldsymbol{K}$ is the characteristic matrix of the network as introduced by Definition 3.1.). The second term $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)^T$ represents the net balance between the supply feedrates, the withdrawal rates and the dilution of the components per unit of volume.

A simple example has been given in Section 2, equations (7) - (9). Here is another example.

Example 5.1. Competitive growth of a single population on two limiting substrates (continued). The reaction network of the system is given by (11)-(13). Assume now that the process takes place in a *fed-batch* aerated bioreactor supplied with the primary substrates S_1. The general dynamical model for this example is written as:

$$\frac{d}{dt}\begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ E \\ X \\ O \\ P \end{pmatrix} = \begin{pmatrix} -k_1 & 0 & 0 \\ 1 & -k_3 & 0 \\ k_2 & 0 & -k_6 \\ 0 & 0 & k_7 \\ 0 & 1 & 1 \\ 0 & -k_4 & -k_8 \\ 0 & k_5 & k_9 \end{pmatrix} \begin{pmatrix} r_1(\xi) \\ r_2(\xi) \\ r_3(\xi) \end{pmatrix} + \begin{pmatrix} -DS_1 + F_{in}V^{-1} \\ -DS_2 \\ -DS_3 \\ -DE \\ -DX \\ -DO + Q_0V^{-1} \\ -DP + Q_PV^{-1} \end{pmatrix}$$

where D represents the dilution rate, V the volume of the liquid phase in the reactor, F_{in} the primary substrate mass feeding rate, Q_0 the oxygen transfer rate, Q_P the gazeous CO_2 outflow rate.

6 Statement of the identification problem

We assume that :

- A1 The components involved in the system are known and measurable.
- A2 An experiment where all the state variables ξ_i and all the "exogeneous" signals u_i have been measured (possibly through off-line laboratory analyses) with a reasonable sampling frequency is available.

The identification problem is then threefold :

1. Determination of the number of reactions involved in the process
2. Identification of the reaction network : determination of the structure of the reaction network and estimation of the unnormalized nonzero coefficients of the characteristic matrix $\boldsymbol{K}$.
3. Identification of the kinetics : determination of the structure of the kinetics and estimation of the unnormalized nonzero kinetic coefficients.

These determinations have to be performed from the data collected during the available experiment (Assumption A2)

As a matter of fact, in most practical cases, these three parts of the identification problem can be completely decoupled from one another, as we shall now show in the sequel of the paper.

7 Determination of the number of reactions

It is an evidence that the general dynamical model (18) for biological processes will be identifiable only if the reactions that are supposed to occur in the bioreactor are *distinguishable* from the data. A lack of distinguishability may appear for two structurally different reasons : either because the characteristic matrix $\boldsymbol{K}$ has not full rank or because the reaction kinetics are not independent. We illustrate the issue with two simple examples.

Example 7.1 As in Example 4.2, we consider the case of a simple microbial growth process with an associated formation of a product of interest. We assume that the product formation is decoupled from the growth, as represented by the following scheme with *two* reactions :

$$S \longrightarrow k_1 X \tag{19}$$
$$S \longrightarrow k_2 P \tag{20}$$

The dynamical model is written as :

$$\frac{d}{dt}\begin{pmatrix} X \\ S \\ P \end{pmatrix} = \begin{pmatrix} k_1 & 0 \\ -1 & -1 \\ 0 & k_2 \end{pmatrix} \begin{pmatrix} r_1(\boldsymbol{\xi}) \\ r_2(\boldsymbol{\xi}) \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \tag{21}$$

Let us now suppose that the decoupling between growth and production is in fact an erroneous assumption and that the reaction rates $r_1(\boldsymbol{\xi})$ and $r_2(\boldsymbol{\xi})$ are *not* independent but proportional to one another:

$$r_1(\boldsymbol{\xi}) = \alpha r_2(\boldsymbol{\xi})$$

Then, by introducing another reaction rate $\bar{r}(\boldsymbol{\xi})$ defined as :

$$\bar{r}(\boldsymbol{\xi}) \triangleq (1+\alpha) r_1(\boldsymbol{\xi})$$

and after a few calculations, it is easily shown that the model (21) is equivalent to :

$$\frac{d}{dt}\begin{pmatrix} X \\ S \\ P \end{pmatrix} = \begin{pmatrix} k_1' \\ -1 \\ k_2' \end{pmatrix} \bar{r}(\boldsymbol{\xi}) + \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$$

with $k_1' \triangleq k_1/(1+\alpha)$ and $k_2' \triangleq k_2\alpha/(1+\alpha)$. This model itself is associated with the following scheme :

$$S \longrightarrow k_1' X + k_2' P \tag{22}$$

It is thus clear that the two reactions (19) and (20) are *not* distinguishable (in the sense that the proportionnality constant α is not identifiable). ■

Example 7.2 Let us now consider the following scheme involving 3 reactions with 4 components :

$$S_1 \longrightarrow S_2 \tag{23}$$
$$S_1 \longrightarrow X + P \tag{24}$$
$$S_2 \longrightarrow X + P \tag{25}$$

The corresponding dynamical model is written :

$$\frac{d}{dt}\begin{pmatrix} S_1 \\ S_2 \\ X \\ P \end{pmatrix} = \begin{pmatrix} -1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} r_1(\xi) \\ r_2(\xi) \\ r_3(\xi) \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix} \quad (26)$$

We observe that the characteristic matrix $\boldsymbol{K}$ has not full rank : rank $(\boldsymbol{K}) = 2$. Therefore, by introducing the following new reactions rates :

$$\begin{aligned} \bar{r}_1(\xi) &= r_1(\xi) + r_2(\xi) \\ \bar{r}_2(\xi) &= r_2(\xi) + r_3(\xi) \end{aligned}$$

it is easily shown that the model (26) is equivalent to :

$$\frac{d}{dt}\begin{pmatrix} S_1 \\ S_2 \\ X \\ P \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \bar{r}_1(\xi) \\ \bar{r}_2(\xi) \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix}$$

which corresponds to the following reaction scheme :

$$\begin{aligned} S_1 &\longrightarrow S_2 \\ S_2 &\longrightarrow X + P \end{aligned}$$

Again, it is clear that the three reactions (23) (24) (25) are not distinguishable. ■

These examples lead to the following definitions.

Definition 7.3. Independent kinetics. The kinetics are independent if and only if the vector space generated by the reaction rate functions $r_1(\xi), r_2(\xi), \ldots, r_m(\xi)$ has dimension m.

Definition 7.4. Independent reactions. The reactions are independent if and only if the characteristic matrix $\boldsymbol{K}$ has full rank and the kinetics are independent. ■

The above examples have clearly shown that only sets of independent reactions are identifiable. The first step in any identification should therefore be the determination of the number of independent reaction compatible with the available data. This determination can be performed as follows (see [4]). By integrating the model equation (18) over an arbitrary time interval, we obtain:

$$\zeta(t) \triangleq \xi(t) - \xi(0) - \int_0^t \boldsymbol{u}(\tau)d\tau = \boldsymbol{K}\int_0^t \boldsymbol{r}(\xi(\tau))d\tau$$

We note that this vector $\zeta(t)$ is calculable at the sampling instants from the available data (obviously with an appropriate numerical approximation of the integral). The following matrix is constructed :

$$A \triangleq \begin{pmatrix} \zeta_1(t_0) & \zeta_2(t_0) & \ldots & \zeta_n(t_0) \\ \zeta_1(t_1) & \zeta_2(t_1) & \ldots & \zeta_n(t_1) \\ \vdots & & & \vdots \\ \zeta_1(t_N) & \zeta_2(t_N) & \ldots & \zeta_n(t_N) \end{pmatrix}$$

where $t_0, t_1, \ldots, t_j, \ldots, t_N$ denote the sampling instants and $\zeta_i(t_j)$ individual components of ζ. Then the number of independent reactions is simply given by the rank of the matrix A which is determined, in practice, by the number of singular values of A that are significantly different from zero. A very illustrative application of this method in the case of a culture of *B. subtilis* can be found in [5]

8 Identification of the reaction network

Let us now assume that the number of independent identifiable reactions is given. This means that the dimensions $n \times m$ of the characteristic matrix $\boldsymbol{K}$ are also fixed. We introduce the following additional assumption (satisfied in most practical applications) :

- A3 The number of components n is larger than the number of reactions m.

This assumption means that the number of rows n of $\boldsymbol{K}$ is larger than the number of columns m of $\boldsymbol{K}$.

We define the *structure* of the reaction network as follows.

Definition 8.1. Structure of the reaction network. The structure of the reaction network is defined by the location of the non-zero entries in the characteristic matrix $\boldsymbol{K}$. ■

Indeed, this location of non zero entries designates in fact the species that are participating in each reaction.

Under Assumption A3, for each plausible structure, a linear state transformation of the initial dynamical model (18) can be done to obtain an auxiliary model which is independent of the kinetics. Let p be the rank of the $n \times m$ characteristic matrix $\boldsymbol{K}$. We define a partition $(\boldsymbol{K}_a, \boldsymbol{K}_b)$ such that

$$\begin{pmatrix} \boldsymbol{K}_a \\ \boldsymbol{K}_b \end{pmatrix} = \boldsymbol{E}\boldsymbol{K}$$

where $\boldsymbol{E}$ is a row permutation elementary matrix and $\boldsymbol{K}_a$ is a full row rank submatrix of $\boldsymbol{K}$, i.e. the dimension of $\boldsymbol{K}_a$ is $p \times m$ and rank $(\boldsymbol{K}_a) = p$. Then by the same permutation, the dynamical model (18) can be partitioned into two parts :

$$\frac{d}{dt}\begin{pmatrix} \xi_a \\ \xi_b \end{pmatrix} = \begin{pmatrix} \boldsymbol{K}_a \\ \boldsymbol{K}_b \end{pmatrix} \boldsymbol{r}(\xi) + \begin{pmatrix} \boldsymbol{u}_a \\ \boldsymbol{u}_b \end{pmatrix}$$

with : $\begin{pmatrix} \xi_a \\ \xi_b \end{pmatrix} \triangleq \boldsymbol{E}\xi$ and $\begin{pmatrix} \boldsymbol{u}_a \\ \boldsymbol{u}_b \end{pmatrix} \triangleq \boldsymbol{E}\boldsymbol{u}$

We then consider the following linear state transformation of the initial dynamical model (18) :

$$\xi_a = \xi_a \quad (27)$$

$$z = \boldsymbol{C}\xi_a + \xi_b \quad (28)$$

where the $(n-p) \times m$ matrix $\boldsymbol{C}$ is the solution of the matrix equation

$$\boldsymbol{C}\boldsymbol{K}_a + \boldsymbol{K}_b = 0 \quad (29)$$

Since $\boldsymbol{K}_a$ has full row rank, $\boldsymbol{C}$ is uniquely defined by $-\boldsymbol{K}_b\boldsymbol{K}_a^+$ with $\boldsymbol{K}_a^+$ being an arbitrary right inverse of $\boldsymbol{K}_a$ such that $\boldsymbol{K}_a\boldsymbol{K}_a^+ = \boldsymbol{I}_p$, and $\boldsymbol{I}_p$ being an identiy matrix of dimension p. The dynamical model (18) is transformed into

$$\dot{\xi}_a = \boldsymbol{K}\boldsymbol{r}(\xi_a, z - \boldsymbol{C}\xi_a) + \boldsymbol{u}_a \quad (30)$$

$$\dot{z} = \boldsymbol{C}\boldsymbol{u}_a + \boldsymbol{u}_b \quad (31)$$

The second part (31) of this transformed model does not involve explicitly the reaction rates and can thus be used to estimate the yield coefficient without modelling the reaction kinetics. More precisely, the identification of the yield coefficients is performed as follows. First the elements of the matrix $\boldsymbol{C}$ are identified from the following *linear* auxiliary model derived from (31) and (28):

$$\dot{z} = \boldsymbol{C}\boldsymbol{u}_a + \boldsymbol{u}_b \quad (32)$$

$$\xi_b = z - \boldsymbol{C}\xi_a \quad (33)$$

This identification is carried out in the following way. For each tentative value of $\boldsymbol{C}$, equations (32) and (33) are solved to give an estimate of ξ_b denoted $\hat{\xi}_b(\boldsymbol{C})$. Then, according to a least squares principle, the quadratic criterion :

$$J(\boldsymbol{C}) = \sum_{t=t0}^{t_N} [\xi_b - \hat{\xi}_b(\boldsymbol{C})]^2$$

is minimized with respect to $\boldsymbol{C}$.

Afterwards, the recovering of the yield coefficients for the given structure of $\boldsymbol{K}$ is performed by solving equation (29):

$$\boldsymbol{C}\boldsymbol{K}_a + \boldsymbol{K}_b = 0$$

with respect to the $k_i's$. The conditions under which the yield coefficients k_i are identifiable with this method are explicited in [6] and [3].

Example 8.2. Competitive growth of a single population on two limiting substrates(continued). In this example, there exist several partitions that can provide different auxiliary models. For instance, one can take rows 2, 5 and 4 for $\boldsymbol{K}_a$ and the remaining rows for $\boldsymbol{K}_b$:

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

The value of $\boldsymbol{C}$ is structurally defined by

$$\boldsymbol{C} = -\boldsymbol{K}_b \boldsymbol{K}_a^{-1} = \begin{pmatrix} k_1 & k_1k_3 & -k_1k_3k_7^{-1} \\ -k_2 & -k_2k_3 & (k_2k_3 - k_6)k_7^{-1} \\ 0 & k_4 & (k_8 - k_4)k_7^{-1} \\ 0 & -k_5 & (k_5 - k_9)k_7^{-1} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ 0 & c_{32} & c_{33} \\ 0 & c_{42} & c_{43} \end{pmatrix}$$

The induced partitions are

$$\begin{aligned} \boldsymbol{\xi}_a &= (S_2, X, E)^T, \ \boldsymbol{\xi}_b = (S_1, S_3, O, P)^T \\ \boldsymbol{u}_a &= (-DS_2, -DX_2, -DE) \\ \boldsymbol{u}_b &= (-DS_1 + DS_{1in}, -DS_3, -DO + Q_0V^{-1} \\ &\quad -DP + Q_PV^{-1})^T \end{aligned}$$

Define

$$\begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{pmatrix} = \boldsymbol{C} \begin{pmatrix} S_2 \\ X \\ E \end{pmatrix} + \begin{pmatrix} S_1 \\ S_3 \\ O \\ P \end{pmatrix}$$

The auxiliary model (32)-(33) then takes the following specialized form :

$$\frac{d}{dt} \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{pmatrix} = -D \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{pmatrix} + \begin{pmatrix} F_{in}V^{-1} \\ 0 \\ Q_0V^{-1} \\ Q_PV^{-1} \end{pmatrix}$$

$$\begin{pmatrix} S_1 \\ S_3 \\ O \\ P \end{pmatrix} = \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{pmatrix} - \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ 0 & c_{32} & c_{33} \\ 0 & c_{42} & c_{43} \end{pmatrix} \begin{pmatrix} S_2 \\ X \\ E \end{pmatrix}$$

which is used to identify the elements of $\boldsymbol{C}$. Finally, the recovering of the yield coefficients is obtained from the structure of $\boldsymbol{C}$ as follows :

$$k_1 = c_{11}, \ k_2 = c_{23}, \ k_3 = \frac{c_{12}}{c_{11}}, \ k_4 = c_{32}, \ k_5 = -c_{42}$$
$$k_6 = -c_{22} + \frac{c_{12}c_{23}}{c_{13}}, \ k_7 = -\frac{c_{12}}{c_{13}}, \ k_8 = c_{32} - \frac{c_{12}c_{23}}{c_{13}},$$
$$k_9 = -c_{42} + \frac{c_{12}c_{43}}{c_{13}}$$

■

This method is then applied to each candidate structure of the reaction network. The choice between the candidates is then finally performed by using the classical validation techniques of the System Identification theory.

9 Identification of the kinetics

Once the yield coefficients are known, the identification of the reaction kinetics can be considered. For each plausible structure of the reaction kinetics (see Section 4) the identification consists of estimating the involved kinetic coefficients. We assume :

- A4 The $n \times m$ characteristic matrix $\boldsymbol{K}$ has full column rank, i.e. $\text{rank}(\boldsymbol{K}) = m$.

Under this assumption, we know that there exists a feft inverse $\boldsymbol{K}^+$ of $\boldsymbol{K}$ such that $\boldsymbol{K}^+ \boldsymbol{K} = \boldsymbol{I}_m$. We define $\boldsymbol{y} = \boldsymbol{K}^+ \boldsymbol{\xi}$. The dynamics of $\boldsymbol{y}$ are derived from the initial dynamical model (18) as follows :

$$\frac{d}{dt}\boldsymbol{y} = \boldsymbol{r}(\boldsymbol{\xi}) + \boldsymbol{K}^+\boldsymbol{u}$$

or, element by element :

$$\frac{d}{dt}y_j = r_j(\boldsymbol{\xi}) + (\boldsymbol{K}^+\boldsymbol{u})_j, \quad j = 1, \ldots, m$$

In this way, it appears that each reaction rate is completely decoupled from the others : reaction rate r_j only intervenes in the dynamics of y_j. It is then possible to identifiy the model of r_j *as if it were the only reaction taking place in the system !* The identification is thus much easier to perform than in the case where all the reaction rates are to be identified together.

Example 9.1. Competitive growth of a single population on two limiting substrates (continued). Suppose that the three reaction rates of this system are be represented by the following equations :

$$\begin{aligned} r_1 &= \mu_1^* E S_1 \qquad r_3 = \frac{\mu_3^* S_3 X}{K_m + S_3} \\ r_2 &= \mu_2^* S_2 X \end{aligned}$$

where $\mu_1^*, \mu_2^*, \mu_3^*$ and K_m are unknown kinetic coefficients. The first two reaction rates are derived from the mass action principle while the third one is a Michaëlis-Menten type model in S_3. We can choose

$$\boldsymbol{K}^+ = \begin{pmatrix} -\frac{1}{2}k_1^{-1} & \frac{1}{2} & 0 & -\frac{1}{2}k_3k_7^{-1} & \frac{1}{2}k_3 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{2}k_7^{-1} & \frac{1}{2} & -\frac{1}{2}k_1k_9\Delta^{-1} & -\frac{1}{2}k_1k_8\Delta^{-1} \\ 0 & 0 & 0 & \frac{1}{2}k_7^{-1} & 0 & \frac{1}{2}k_1k_5\Delta^{-1} & \frac{1}{2}k_1k_4\Delta^{-1} \end{pmatrix}$$

with $\Delta = k_1(k_4k_9 - k_5k_8)$

We have :

$$\boldsymbol{K}^+\boldsymbol{u} = \begin{pmatrix} -DZ_1 - \frac{1}{2}k_1^{-1}DS_{1in} \\ -DZ_2 - \frac{1}{2}k_1\Delta^{-1}(k_9Q_0V^{-1} + k_8Q_PV^{-1}) \\ -DZ_3 - \frac{1}{2}k_1\Delta^{-1}(k_5Q_0V^{-1} + k_4Q_PV^{-1}) \end{pmatrix}$$

The dynamics of $\boldsymbol{y} = \boldsymbol{K}^+\boldsymbol{\xi}$ are then given by

$$\begin{aligned} \frac{d}{dt}y_1 &= \mu_1^* E S_1 - DZ_1 - \frac{1}{2}k_1^{-1}DS_{1in} \\ \frac{d}{dt}y_2 &= \mu_2^* S_2 X - DZ_2 - \frac{1}{2}k_1\Delta^{-1}(k_9Q_0V^{-1} + k_8Q_PV^{-1}) \\ \frac{d}{dt}y_3 &= \frac{\mu_3^* S_3 X}{K_m + S_3} - DZ_3 - \frac{1}{2}k_1\Delta^{-1}(k_5Q_0V^{-1} + k_4Q_PV^{-1}) \end{aligned}$$

It appears clearly that the three reaction rates are decoupled. The first two equations are linearly parametrized by μ_1^* and μ_2^*, respectively. The third equation has a similar form as equation (8) of our preliminary example (see Section 2).

References

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

[2] Chen L. and G. Bastin, "On the model identifiability of stirred tank bioreactors", *Proc. 1st European Control Conference*, Grenoble, France, pp. 243-247, July 1991.

[3] Chen L., G. Bastin and D. Dochain, "Structural identifiability of the yield parameters in nonlinear compartmental models for bioprocesses", *29th IEEE Conf. on Decision and Control*, Honolulu, pp. 1074-1079, December 1990.

[4] Bonvin D. and D.W.T. Rippin, "Target factor analysis for the identification of stoichiometric models", *Chemical Engineering Sciences*, Vol.45, pp. 3417-3426, 1990.

[5] Saner U.M., "Modelling and on-line estimation in a batch culture of *Bacillus subtilis*", PhD Thesis, Swiss Federal Institute of Technology, Zurich, Switzerland, Diss. ETH n 9481, 1991.

[6] Chen L., "Modelling, identifiability and control of complex biotechnological systems", *PhD Thesis*, CESAME, Louvain University, Belgium, February 1992.